UNIVERSITÉ DE GENÈVE

Archive ouverte UNIGE
https://archive-ouverte.unige.ch

Thèse 2017

Open Access

This version of the publication is provided by the author(s) and made available in accordance with the copyright holder(s).

Diversity of Foraminifera and applications of protist metabarcoding in bioindication: focus on freshwater environment

Apotheloz-Perret-Gentil, Laure

How to cite

APOTHELOZ-PERRET-GENTIL, Laure. Diversity of Foraminifera and applications of protist metabarcoding in bioindication: focus on freshwater environment. Doctoral Thesis, 2017. doi: 10.13097/archive-ouverte/unige:95588

This publication URL: https://archive-ouverte.unige.ch/unige:95588
Publication DOI: 10.13097/archive-ouverte/unige:95588

© This document is protected by copyright. Please refer to copyright holder(s) for terms of use.

UNIVERSITÉ DE GENÈVE — FACULTÉ DES SCIENCES

Département de Génétique et Evolution — Professeur Jan Pawlowski

Diversity of Foraminifera and applications of protist metabarcoding in bioindication: focus on freshwater environment

THÈSE

présentée à la Faculté des sciences de l'Université de Genève
pour obtenir le grade de Docteur ès sciences, mention biologie

par

Laure APOTHÉLOZ-PERRET-GENTIL

de

Le Locle (NE)

Thèse n° 5087

GENÈVE

REPROMAIL

2017

UNIVERSITÉ DE GENÈVE

FACULTÉ DES SCIENCES

DOCTORAT ÈS SCIENCES, MENTION BIOLOGIE

Thèse de Madame Laure APOTHELOZ-PERRET-GENTIL

intitulée :

«Diversity of Foraminifera and Applications of Protist Metabarcoding in Bioindication: Focus on Freshwater Environment»

La Faculté des sciences, sur le préavis de

Monsieur J.PAWLOWSKI, professeur associé(e) et directeur de thèse
Département de génétique et évolution

Monsieur F. ALTERMATT, professeur
Department of Evolutionary Biology and Environmental Studies,
University of Zurich, Switzerland

Monsieur T. STOECK, professeur
Department of Ecology, University of Kaiserslautern, Germany

Madame A. BOUCHEZ, docteure
Centre Alpin de Recherche sur les Réseaux Trophiques des Ecosystèmes Limniques, Institut National de la Recherche Agronomique,Thonon-les-Bains, France

Monsieur A. GOODAY, professeur
Ocean Biogeochemistry and Ecosystems group, National Oceanography Centre, Southampton, United Kingdom

autorise l'impression de la présente thèse, sans exprimer d'opinion sur les propositions qui y sont énoncées.

Genève, le 20 juin 2017

Thèse - 5087 -

Le Doyen

N.B. - La thèse doit porter la déclaration précédente et remplir les conditions énumérées dans les "Informations relatives aux thèses de doctorat à l'Université de Genève".

REMERCIEMENTS

Je tiens à remercier tous les gens qui de près ou de loin m'ont aidée tout au long de ces années de thèse, aussi bien scientifiquement que pour la gestion de mon travail en harmonie avec mon nouveau rôle de mère.

Evidemment tout d'abord un grand merci à Jan. Pour m'avoir donné l'opportunité d'accomplir ce travail et pour tout ce que tu m'as appris pendant plus de 6 ans ! J'apprécie beaucoup la confiance que tu nous portes pour la gestion de nos projets et j'espère que nous te le rendons bien. Un merci particulier pour le soutien et la compréhension que tu as eu par rapport à mon choix de faire non pas un mais deux enfants pendant mon doctorat.

Je tiens à remercier aussi tous les membres du labo que j'ai eu la chance de croiser pendant toutes ces années : Maria, Manu, Sofia, Loïc, Roberto, Léo, Philippe, Tristan, Franck, Joana, Joanna, Amine, Alexandra, José, Jackie, Sasha, Ivan et tout ceux que j'ai pu oublié

Un merci particulier à Maria, pour les longues discussions diverses et variées, les séances de yoga et surtout pour m'avoir supportée dans ton bureau pendant autant d'années !

Manu, la liste des remerciements est bien trop longue (et surtout trop privée !) pour figurer ici mais je tiens quand même à te dire à quel point ta présence a été importante pour moi. D'un point de vue purement professionnel, merci pour les heures de discussion a propos de l'optimisation des méthodes au labo. Et sinon humainement, merci pour tout le reste !

Sofia, un grand merci pour ton soutien dans les périodes difficiles, de doutes ou de très grande fatigue. Ton savoir et ton expérience en tant que scientifique et en tant que mère m'ont beaucoup apportés !

Loïc et Roberto, merci pour les rires, les bières, les anecdotes et aussi pour les discussions scientifiques quand même…

Merci à tous les gens avec qui j'ai été amenée à collaborer, toujours avec le sourire et la bonne humeur, je pense notamment à Arielle, François, Jennifer, Ilham, Slim, Agathe, Ferry, Sigal, Gily, Yamama, Regine…

Juan je tiens à te remercier de ta présence toujours agréable et de ta disponibilité surtout en cette fin de thèse qui s'est avérée quelque peu compliquée…

Je remercie également l'Université de Genève et le département de génétique et évolution avec un merci tout particulier à Valérie et Corinne pour votre serviabilité et efficacité dans la gestion des problèmes administratifs. Merci aussi au Fond National Suisse pour la Recherche ainsi qu'au canton de Genève pour leur soutien financier.

My sincerest thanks to Agnès Bouchez, Andrew Gooday, Florian Altermatt and Thorsten Stoeck for accepting to review my thesis !

Merci à tous mes amis ainsi qu'à ma famille pour leur soutien et leur compréhension face à mon manque évident de temps.

Un merci spécial à toi maman pour m'avoir aidée et grandement soulagée avec mon organisation familiale. C'est une véritable chance de t'avoir à mes côtés au quotidien.

Merci papa ! Des étoiles que tu as dans les yeux quand on fait référence à mon parcours, des discussions qui se rapprochent finalement plus souvent de la philosophie que de la science... Merci de croire en moi et d'en être fier !

Et pour finir le plus grand des merci à Sébastien, Léanne et Maxine : les amours de ma vie. Merci d'être là au quotidien près de moi et de me rappeler qu'il n'y a pas que le travail dans la vie… Je vous aime de tout mon cœur !

RÉSUMÉ

La première partie de cette thèse se concentre sur l'étude de la diversité génétique des foraminifères, avec une attention particulière sur les milieux d'eau douce. Quatre nouvelles espèces de foraminifères monothalames ont été décrites et une espèce décrite précédemment a été caractérisée moléculairement. Deux de ces foraminifères, avec un test organique à une seule loge, proviennent du golfe de Eilat dans la Mer Rouge (**Chapitre 2** et **3**). Ces deux espèces sont morphologiquement similaires, cependant les analyses de la petite sous-unité de l'ARN ribosomique (le gène 18S rRNA) montrent qu'elles sont phylogénétiquement éloignées. Une de ces deux espèces, *Arnoldiellina fluorescens*, branche à l'intérieur d'un clade déjà connu de foraminifères monothalames et affiche la particularité d'émettre de l'autofluorescence verte. La deuxième espèce, *Leannia veloxifera*, se place proche d'une séquence environnementale, loin des autres clades connus de monothalames. En plus de cela, trois foraminifères d'eau douce ont été caractérisé génétiquement, deux d'entre eux (*Lacrogromia cassipara* et *Limnogromia sinensi)* ont été nouvellement décrits de manière morphologique (**Chapitre 4**). Grâce à ces descriptions morphologiques et moléculaires, trois des quatre clades majeurs, jusque-là connus uniquement grâce aux séquences ADN environnementales, ont pu être définis morphologiquement.

Afin de mieux définir la diversité des foraminifères d'eau douce, nous avons analysé 98 échantillons de sédiment provenant du bassin genevois (**Chapitre 5**). C'est dans cette même région que les premiers foraminifères présentant une morphologie similaire à *Lacrogromia* et *Limnogromia* ont été observés il y a de cela un siècle. Cette étude, qui s'est étendue sur presque 4 ans d'échantillonnage, a démontré génétiquement que les foraminifières étaient présents dans presque toutes les rivières et plans d'eau genevois. Les 48 phylotypes identifiés durant cette étude groupent dans les quatre clades de foraminifères d'eau douce auparavant décrits ainsi que dans un nouveau clade. Les analyses phylogénétiques suggèrent que la colonisation des milieux d'eau douce par les foraminifères s'est produite plusieurs fois au cours de leur évolution. La grande diversité génétique des foraminifères d'eau douce a été confirmée au cours d'une étude sur 68 échantillons de sédiment et 43 de biofilm en utilisant une approche de métabarcoding de l'ADN environnemental à l'aide de la technologie du séquençage haut-débit.

La grande diversité des organismes révélée par les études de séquençage haut-débit est souvent largement supérieure à la diversité des espèces morphologiques. La variation intragénomique peut partiellement expliquer ce phénomène. Nous avons donc étudié le niveau de polymorphisme du gène 18S de l'ARN ribosomique chez 130 espèces de foraminifères (**Chapitre 6**). Notre étude, basée sur le séquençage à haut débit d'un seul spécimen à la fois, a confirmé de précédents résultats réalisés avec la technique du séquençage en Sanger lesquelles indiquaient que le polymorphisme était largement répandu chez les foraminifères. Nous avons mis en évidence différents haplotypes constitués soit de substitutions individuelles de nucléotides soit de plus larges segments d'expansion. Ces différents haplotypes ont toujours été trouvés dans les régions hypervariables du 18S mais le taux de divergence diffère parmi les espèces étudiées. Des variations plus importantes ont été observées chez les espèces d'eau peu profonde comparées aux espèces abyssales, cependant cette hypothèse doit être confirmée par des études complémentaires.

Dans la deuxième partie de cette thèse, nous avons mis l'accent sur l'application du métabarcoding utilisant les technologies du séquençage à haut débit afin d'évaluer la qualité écologique des cours d'eau. Nous avons tout d'abord calculé le statut écologique de 27 rivières genevoises en utilisant les données ADN et ARN provenant des communautés diatomiques des biofilms (**Chapitre 7**). Pour cela nous avons comparé la valeur donnée par l'indice diatomique suisse (DI-CH) basée sur l'approche morphologique traditionnelle à celle que nous avons obtenue en assignant les séquences à des espèces connues. Dans une deuxième étude (**Chapitre 8)**, nous avons ajouté 60 échantillons et comparé une nouvelle fois la valeur d'indice générée par l'approche traditionnelle (DI-CH) et par les données ADN. De plus, nous avons développé un nouvel indice moléculaire uniquement basé sur les données génétiques sans faire de références aux espèces morphologiques. Ce nouvel indice a montré des résultats très encourageants et il a donc ensuite été appliqué à des données génétiques représentant d'autres groupes taxonomiques (foraminifères, ciliés et autres protistes et métazoaires). Ces données ont été obtenues à partir de 78 échantillons de biofilm provenant des rivières genevoises (**Chapitre 9**). Comme attendu, la meilleure corrélation entre la morphologie et la moléculaire a été trouvé avec les diatomées. Cependant, nous avons aussi trouvé que d'autres algues (Chlorophyceae et Chrysophyceae) ont donné des résultats

comparables aux diatomées, suggérant leur utilisation potentielle comme bioindicateur dans l'évaluation des cours d'eau.

Finalement nous discutons les différents enjeux générés par les études de métabarcoding appliquées en autre à la biosurveillance et à l'évaluation de la qualité des cours d'eau (**Chapitre 10**).

ABSTRACT

My thesis is divided into two parts. The first part focused on the genetic diversity of foraminifera, with an emphasis on freshwater environments. Four new species of morphologically simple, single-chambered (monothalamous) foraminifera were described and one morphospecies was genetically characterised. Two organic-walled monothalamous foraminifera were described from the Gulf of Eilat in the Red Sea (**Chapter 2** and **3**). Both species were morphologically similar but their phylogeny based on the 18S rRNA gene sequences showed that they are only distantly related. One of the two species, *Arnoldiellina fluorescens*, showed the particularity to emit green autofluorescence. The second species, *Leannia veloxifera*, branched close to an unknown environmental marine phylotype but separately from other known clades of monothalamids. We also characterised genetically three freshwater monothalamous species, two of which (*Lacogromia cassipara* and *Limnogromia sinensis*) were newly described (**Chapter 4**). Thanks to this genetic work and associated morphological descriptions, three of the four major phylogenetic clades of freshwater foraminifera known only through their environmental DNA sequences could be characterized morphologically.

To better characterize the environmental diversity of freshwater foraminifera we investigate 98 sediment samples from the Geneva basin, where morphotypes similar to *Lacogromia* and *Limnogromia* have been reported a century ago (**Chapter 5**). This almost four years long metabarcoding study revealed that foraminifera are genetically present in almost all samples from Geneva rivers and standing waters. In result of this study 48 new phylotypes branching within the four known freshwater foraminifera clades and one newly described clade were described. Phylogenetic analyses suggested that the colonization of freshwater habitats by foraminifera occurred several times during their evolution. This was confirmed by high genetic diversity of freshwater foraminifera detected in a survey of 68 sediment and 43 biofilm samples analysed using high-throughput sequencing (HTS) metabarcoding.

At the end of this first part, the results of the single-cell HTS study of intragenomic polymorphism in foraminifera have been presented. This study addressed a common observation that the high sequence diversity revealed by HTS metabarcoding studies is largely superior to the number of morphospecies from the same environment. Intragenomic variations can account for this diversity and we therefore

investigated the level of polymorphism of the 18S rRNA gene in 130 specimens of foraminifera (**Chapter 6**). Our study using a single-cell HTS approach, confirms previous studies based on Sanger sequencing showing that intragenomic polymorphism is widely spread in foraminifera. We highlight different patterns of polymorphism involving single-nucleotide substitutions and larger expansion segments variations. Many different haplotypes were found within the hypervariable regions of the 18S rDNA but the rate of sequence divergence depends on the species. We observed more important levels of polymorphism in shallow water than in deep-sea species, but this has to be confirmed by further studies.

The second part of this thesis has been focused on the application of HTS metabarcoding to assess the ecological quality of watercourses. In the first study, we inferred the ecological status of 27 river sites in Geneva using the information provided by DNA and RNA from the diatom community of biofilm samples (**Chapter 7**). We compared the value of the Swiss Diatom Index (DI-CH) inferred by traditional morphological approach to the molecular data generated by HTS that were assigned to morphospecies. In the second study (**Chapter 8)** we completed our dataset with additional 60 biofilm samples and we compared the DI-CH inferred by morphology to a new molecular index based only on the genetic dataset without referring to morphospecies. This new taxonomy-free approach showed very promising results and was applied to other taxonomic groups (foraminifera, ciliates and whole eukaryotic diversity) on 78 biofilm samples from the Geneva basin (**Chapter 9**). As expected, the best correlation between molecular and morphological DI-CH index was found using diatoms. However, we have also found that the other algae (Chlorophyceae and Chrysophyceae) give similar results as diatoms, suggesting that they could be used in watercourses biomonitoring as complementary bioindicators.

At the end of this second part, we discuss various challenges raised by the HTS metabarcoding surveys applied to biomonitoring and the assessment of water quality (**Chapter 10**).

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

1.1. DNA barcoding and metabarcoding

1.1.1. DNA Barcoding

The DNA barcoding concept is based on the idea that we can identify species using a short DNA sequence (Hebert *et al.* 2003). For this purpose, suitable gene markers have to be chosen carefully. Indeed those genes have to be sufficiently conserved among the group to allow the design of amplification primers but also possess enough variability inside the fragment to distinguish between different species (Hebert *et al.* 2003). Moreover, a high copy number of the gene is necessary to ensure DNA amplification from a single specimen, used as voucher for DNA barcodes reference database.

An international organization, the International Barcode of Life (http://www.ibol.org) coordinates and promotes the international barcoding efforts on the different taxonomic groups. The IBOL manages the work performed by the different countries in order to generate a extensive publicly available database containing barcodes sequences for various taxa, promordially of metazoans: the Barcode of Life Data (BOLD) system. In many countries, the IBOL objectives are supported by local organisations that coordinate the barcoding activities at national scale (SwissBOL in Switzerland, NORBOL in Norway, GBOL in Germany, ABOL in Austria, FINBOL in Finland).

In order to standardize the DNA barcodes, several genes have been selected as the most suitable for DNA-based identification, for example the mitochondrial cytochrome c oxidase 1 (COI) for most of metazoan species (Hebert *et al.* 2003; Hajibabaei *et al.* 2006), the two locus ribulose-bisphosphate carboxylase (*rbc*L) and maturaseK (*matK*) with the addition of the *trn*H-*psb*A non-coding plastid region for plants (Hollingsworth *et al.* 2009) and the nuclear ribosomal internal transcribed spacer (ITS) region for fungi (Schoch *et al.* 2012). In the case of protists, the variable region V4 of the nuclear SSU rRNA gene has been considered as the universal DNA barcode, while specific DNA barcodes have been proposed for different taxonomic groups of eukaryotes (Pawlowski *et al.* 2012).

1.1.2. Applications of DNA barcoding

DNA barcoding has been used in wide range of applications (Figure 1.1). One of the most common applications of DNA barcoding is the identification of unknown or cryptic species (Hebert *et al.* 2004). Indeed, DNA barcoding allows to distinguish related species that could not be distinguish morphologically (Blaxter 2004; Smith *et al.* 2008; Pauls *et al.* 2010; Huemer *et al.* 2014; Blanco-Bercial *et al.* 2014). It can also be used to identify immature life-stages, such as eggs or larvae of invertebrates, which are often very difficult to assign to an adult form (Jousson *et al.* 1998; van Nieukerken *et al.* 2012; Meiklejohn *et al.* 2013). This is also applicable to plants, for example in the authentication of medicinal plants (see Techen *et al.* 2014). Another application more connected to society is the food safety field; for example the DNA barcoding has been successfully applied to detect frauds or mislabelling of seafood products available on the market that have been already processed and therefore difficult to recognize (Hanner *et al.* 2011; Huxley-Jones *et al.* 2012; Vartak *et al.* 2015; Shokralla *et al.* 2015). Furthermore, the DNA barcoding of crop species allows, in addition to species identification, to trace the quality, the origin and the possible genetic modification of seeds, which are key components of the food safety and control procedures (Auer 2003; Ren *et al.* 2006; Mattia *et al.* 2008). DNA barcoding also proved to be useful in forensic science, for example for the identification of immature stages of flesh flies (Boehme *et al.* 2012; Meiklejohn *et al.* 2013).

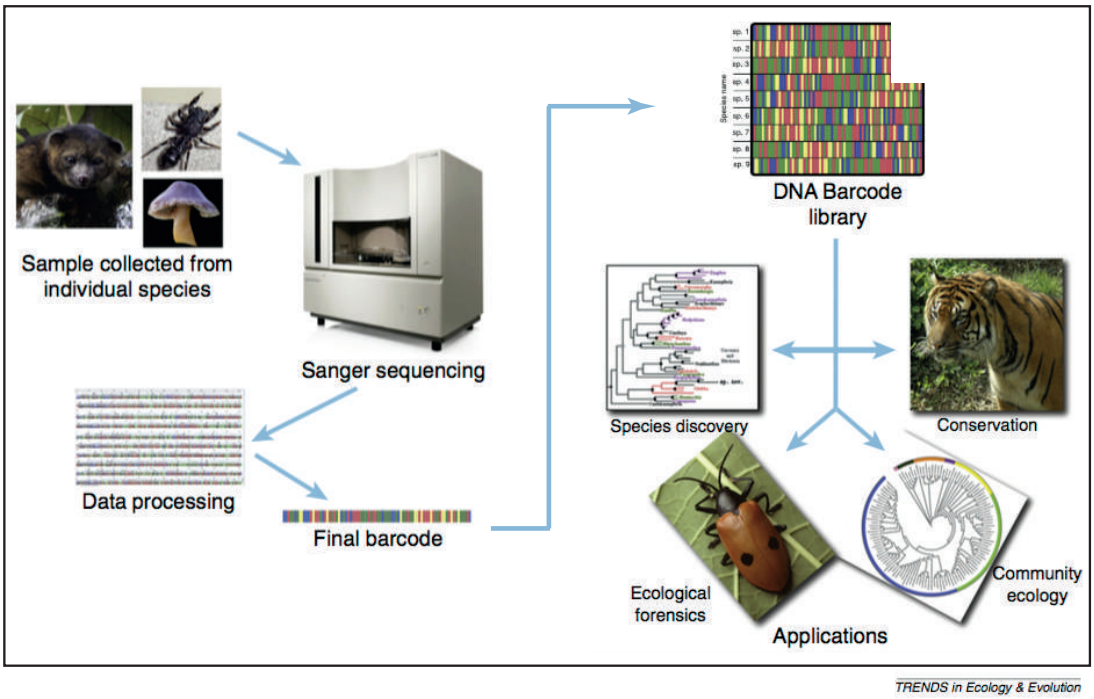


Figure 1.1 Basic workflow for generating DNA barcodes using Sanger sequencing. *From* Kress et al., 2015

The classical DNA barcoding approaches are based on the sequencing of a fragment or entire specimen but the DNA can also be used to detect particular species in environmental samples. It is known that large organisms can shed a lot of DNA in their environment, so called extracellular DNA, via for example reproductive or epithelial cells, faeces, bodily fluids and decomposition of dead bodies (see Barnes & Turner 2016). Therefore, species of interest can be targeted in an environmental DNA sample (Figure 1.2A). Ficetola *et al.* (2008), have been the first to use extracellular DNA in conservation biology; they targeted the invasive bullfrog species into French ponds by detecting their DNA in water samples. Latter, this kind of applications has been widely used to detect the presence of invasive (Goldberg *et al.* 2013; Egan *et al.* 2013; Moyer *et al.* 2014) or endangered species (Wilcox *et al.* 2014; Rees *et al.* 2014; Sigsgaard *et al.* 2015). In addition, this approach can be extrapolated to other application fields, like health or forensics sciences (e.g. analysis of gut contents of malaria vectors – Garros *et al.* 2008) as well as food safety (e.g. find predator of a crop of interest – Karp *et al.* 2014).

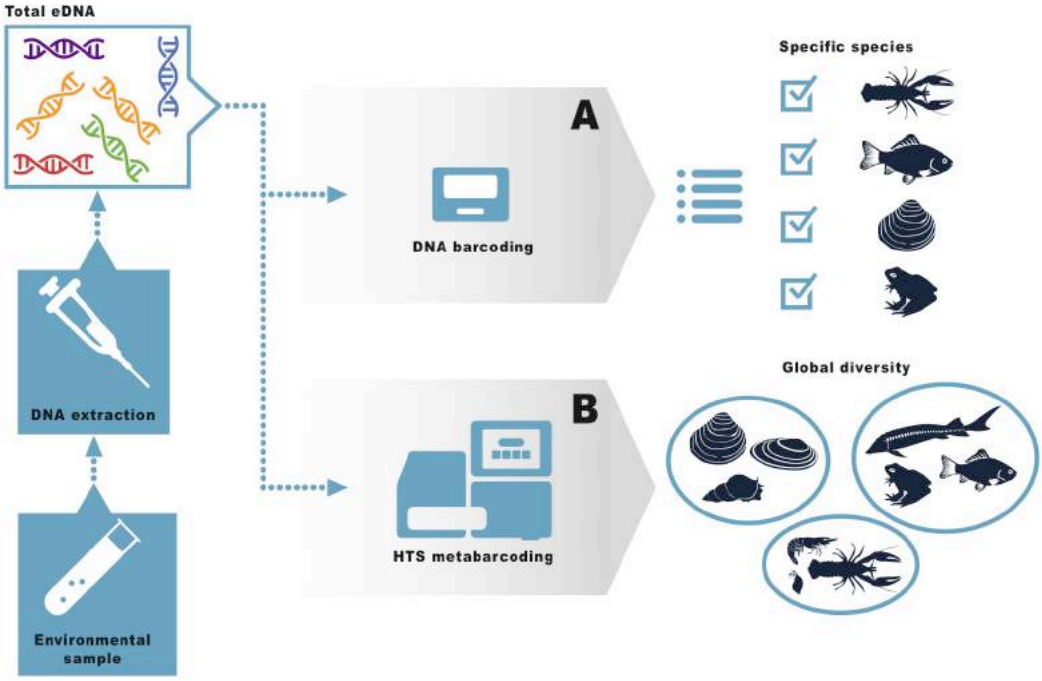


Figure 1.2 General workflow for eDNA barcoding (A) and metabarcoding (B) surveys.

1.1.3. High-throughput metabarcoding

The DNA-based identification of the whole community of species present in environmental samples is called metabarcoding or eDNA metabarcoding (Figure 1.2B, Taberlet *et al.* 2012). This approach, called sometimes environmental genomics, is widely used for identification of microbial communities composition and characterisation of microbiome in medical and environmental research (Zaets *et al.* 2016; Robinson *et al.* 2016; Beckers *et al.* 2016; Galan *et al.* 2016). Until the advent of high-throughput sequencing, the analyses of species communities were performed by cloning and Sanger sequencing. However, this method allows obtaining in the best case only few hundred of sequences per sample. Over the last decade, new high-throughput sequencing (HTS) technologies bring another perspective to this field. Indeed, several sequencing platforms (e.g. Illumina, Ion Torrent, Pacifics Biosciences, Oxford Nanopores Technologies (Reuter *et al.* 2015)) are now available on the market, allowing the sequencing of a large amount of DNA sequences (up to 5

billion for Illumina HiSeq technologies). The importance of HTS metabarcoding in ecology and conservation biology was highlighted by several authors (Bohmann *et al.* 2014; Valentini *et al.* 2015). The application of metabarcoding was extended to the same fields as traditional DNA barcoding, addressing slightly different questions. For example, the metabarcoding was used in food safety to investigate the foraging preference of honeybees to detect the botanical and entomological origin of the honey (Prosser & Hebert 2017) or for identification of soil origin in function of their eukaryotic diversity for forensic investigations (Giampaoli *et al.* 2014).

Metabarcoding has also been shown to be a very powerful tool to explore the environmental diversity of microbial eukaryotes. Several metabarcoding studies reveal a huge diversity of protists in various ecosystems, including the most inhospitable ones (Amaral-Zettler 2012; Stoeck *et al.* 2014). The metabarcoding studies conducted in the framework of large oceanic campaigns (Tara Ocean; Biomarks, etc) revealed immense diversity of marine protists, comprising many undescribed species and higher taxa (Logares *et al.* 2014; de Vargas *et al.* 2015; Massana *et al.* 2015; Le Bescot *et al.* 2016). The same has been observed in freshwater and soil ecosystems, inhabited by a myriad of eukaryotic species known only through their DNA sequences (Lentendu *et al.* 2014; Arjen de Groot *et al.* 2016; Mahé *et al.* 2017). The HTS metabarcoding also allowed investigating the seasonal variations of these taxa (Egge *et al.* 2015b; Simon *et al.* 2015), as well as their changes through time based on paleogenomic data (Coolen *et al.* 2013; Capo *et al.* 2015). Among numerous other applications, the metabarcoding was essential to identify many parasitic protists, which importance was largely underestimated (Guillou *et al.* 2008; Bass *et al.* 2015).

1.1.4. Limitations and biases of metabarcoding

Although metabarcoding is a powerful tool, it is also subject to various biases that should not be underestimated because they can lead to erroneous interpretation of generated HTS data. Biases are present through all the process from eDNA sampling to PCR amplification and high-throughput sequencing. The Figure 1.3 summarizes some of the limitations and biases generated by the HTS metabarcoding studies.

Some biases might occur already during the sampling, particularly when working with benthic protists, because of the patchiness of the environment. It is well known that sediments are not homogenous (Morrisey *et al.* 1992; Gooday & Jorissen 2012) highlighting the importance of sampling replicates. Moreover, as mentioned before, the DNA can be well preserved in the environment as free molecules or inside inactive cells (Barnes & Turner 2016), therefore the sampled environmental DNA does not always correspond to living species. On top of that, special precautions are necessary during sampling and lab work to avoid and detect the external and cross-contaminations events. These precautions can be inspired by on ancient DNA studies, where contaminants are particularly problematic (Llamas *et al.* 2017).

In the wet lab, the main sources of errors are the PCR amplification (Berney *et al.* 2004; Aird *et al.* 2011) as well as the sequencing step (Schirmer *et al.* 2015). During the PCR amplification, the main errors are resulting from the erroneous insertions of nucleotides by Taq polymerase (Eckert & Kunkel 1991; McInerney *et al.* 2014; Lee *et al.* 2016), substitutions induced by the DNA damage caused by the temperature cycling of the PCR (Potapov & Ong 2017) and the formation of chimeras (Fonseca *et al.* 2012). Chimeric PCR products are generated when aborted amplification fragments used as templates in the next amplification step do not originate from the same genome. This phenomenon is particularly problematic when several variable regions are intermingled with conserved regions, which is the case of the rRNA genes. A lot of HTS protocols are multiplexing the samples via a short barcode tag introduced during the PCR step. The shuffle of those tags, so called misstagging, results in the wrong sample assignation. Esling *et al.* (2015) proposed a filter to detect and remove the mistagging errors. The most recent utilisation of PCR-free method in the preparation of sequencing libraries highly reduces this issue, however they requires higher amount of starting DNA, which is not always possible.

Finally, the computer processing of the metabarcoding data involves several potential sources of biases related to quality filtering, paired-end assembly, chimera removal and sequences clustering into Operational Taxonomic Units (OTUs) that are managed by various available pipelines (Mysara *et al.* 2017, mothur, QIIME). Afterwards, the most common methods to assign the data to known species are using BLAST (Altschul *et al.* 1990) or sequence similarity against a curated database (Brandes *et al.* 2016; Boscaro *et al.* 2016).

Regardless of the method chosen to assign the OTUs, an extensive reference database is needed. Several public databases are available for specific markers, such as BOLD for DNA barcodes of metazoans, plants and fungi (www.boldsystems.org, Ratnasingham & Hebert 2007) and the SILVA database (www.arb-silva.de, Yilmaz *et al.* 2014), as well as Ribosomal Database Project (RDP) for rRNA genes. There are also reference database specific to taxonomic groups such as the R-syst databases (http://www.rsyst.inra.fr), which contains several databases for different groups of interest as diatoms (Rimet *et al.* 2016), or Barcoding Project for benthic foraminifera (www.forambarcoding.unige.ch). However those databases are not exhaustive and this can be an issue, particularly when taxonomic assignment at species level is needed.

The biological meaning of the number of sequences obtained by HTS metabarcoding studies is also a major concern. Indeed, the differences in the relative abundance of number of sequences generated for a specific taxa (reads) compared to the relative abundance of the related individuals are widely accepted (Nolte *et al.* 2010; Amend *et al.* 2010; Stoeck *et al.* 2014; Elbrecht & Leese 2015). However, several studies seem to indicate that the relative abundance of sequences can be used for at least some unicellular organisms but not as a direct correlation between the relative number of cells and reads (Pawlowski *et al.* 2014a; Giner *et al.* 2016).

Overall, the protocols and methods involved in the HTS metabarcoding pipeline can drastically change the interpretation of the results and therefore the comparison of different metabarcoding studies should be done with caution.

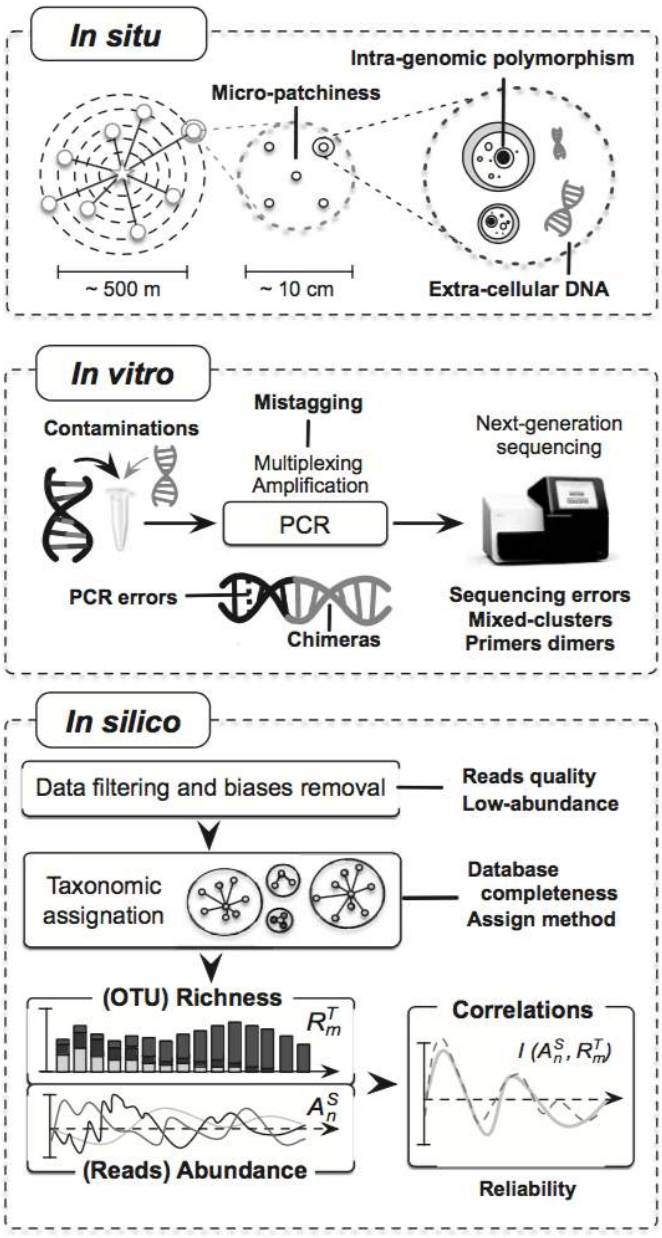


Figure 1.3 The schematic workflow of high-throughput sequencing (HTS) environmental DNA studies with various biological and technical biases indicated in bold. *From* Pawlowski *et al.* 2014b

1.2. Foraminifera

1.2.1. Diversity of foraminifera

Foraminifera are a large group of single-cell eukaryotes counting almost 9'000 living species according to the World Register of Marine Species (WoRMS, www.marinespecies.org). The foraminifera are morphologically characterized by a web of interconnected granuloreticulopods and for most of them the presence of a test that can be organic, calcareous or agglutinated, although few naked species have also been described. The foraminiferal test, when present, consists of one or several chambers, which arrangement and form are used as morphological key features in the description of different species.

Foraminifera are widely spread in marine habitats from deep-sea to coral reefs. Few species have also been described from freshwater (Penard 1902, 1905, 1907; Nauss 1949; Dellinger *et al.* 2014) and soil (Meisterfeld *et al.* 2001) environments. Based on the SSU rRNA gene sequences, Foraminifera have been classified into 3 classes: the single-chambered agglutinated or organic-walled Monothalamea that form a paraphyletic group at the base of multi-chambered Globothalamea and Tubothalamea (Figure 1.4, Pawlowski *et al.* 2013). This topology is consistent with other phylogenies based on the partial SSU rRNA (Pawlowski *et al.* 2002b, 2003; Bowser *et al.* 2006), actin (Flakowski *et al.* 2005), tubulin (Habura *et al.* 2006) and RNA polymerase (Longet & Pawlowski 2007) genes, as well as combined multigene phylogeny (Groussin *et al.* 2011). Monothalamea are divided into several clades comprising marine morphospecies (clades A-M, Pawlowski *et al.* 2002b), 8 marine environmental clades (ENFOR 1-8, Pawlowski *et al.* 2011b) and 4 freshwater clades (Lejzerowicz *et al.* 2010) represented by environmental sequences.

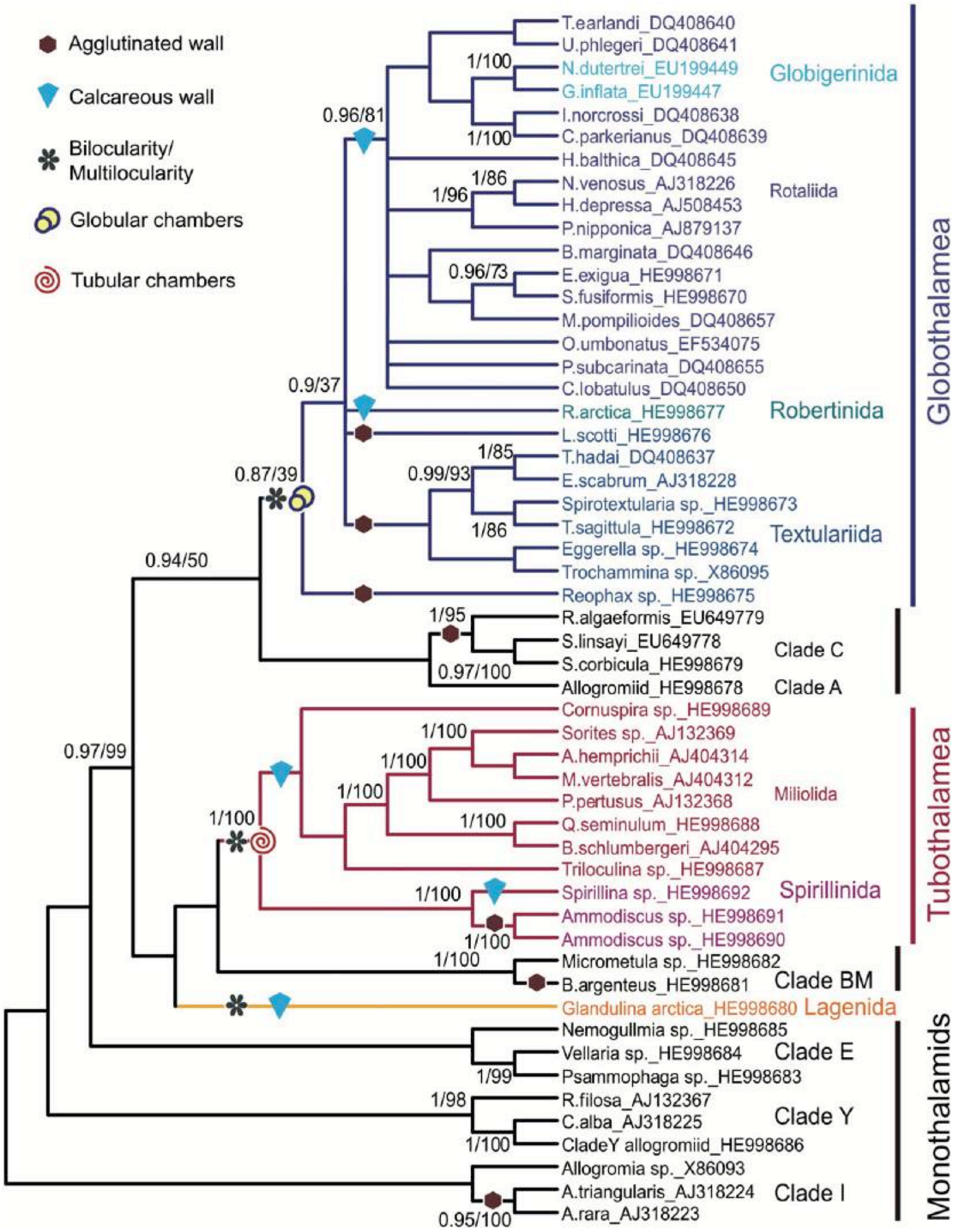


Figure 1.4 Bayesian phylogenetic tree (GTR + G model) showing the phylogeny of Foraminifera inferred from 53 complete SSU rDNA sequences. Numbers at nodes indicate (from left to right) posterior probabilities (BI) and bootstrap values (ML). The tree was rooted with *Allogromia* sp., *A. triangularis* and *A. rara*, as suggested by protein phylogenies. Color symbols indicate stem lineages of Globothalamea and Tubothalamea, as well as groups having agglutinated and calcareous wall. *From* Pawlowski *et al.* 2013

The calcareous and agglutinated tests are easily preserved as microfossils and possess a lot of morphological characteristics that have been widely used in micropaleontology. The multi-chambered hard-shelled groups are more extensively studied than the monothalamids. Indeed with classical micro-paleontological approach, single-chambered soft-walled foraminifera can be difficult to isolate as well as to preserve for a subsequent identification. Therefore, a lot of the knowledge on foraminifera is coming from fossil record and ignore the poorly preserved monothalamid species. These single-chambered species are particularly abundant in the deep-sea (Gooday *et al.* 2004, 2017; Cedhagen *et al.* 2009; Lecroq *et al.* 2011), and polar regions (Sabbatini *et al.* 2004; Sinniger *et al.* 2008; Pawlowski & Majewski 2011). However, few of them have also been described from coastal temperate and tropical areas (Cedhagen & Pawlowski 2002; Gooday *et al.* 2006, 2011).

With the advent of metabarcoding, it becomes clear that the monothalamous foraminifera are much more diverse than suggested by microscopic observations. Several environmental clades have been established with no morphological features characterizing them. In order to contribute to increase our knowledge of this overlooked group of foraminifera, we isolate two monothalamous species from our laboratory cultures and characterize them morphologically and genetically (Chapter 2 and 3)

1.2.2. Freshwater environment

Compared to the marine fauna, the freshwater foraminifera remains poorly known, even if the first observation of freshwater foraminifera went back to more than a century (Claparède & Lachmann 1859). This first described freshwater species, named *Lieberkhunia sp*, showed an organic flexible test with a very large pseudopodial network. At the turn of the XIX century, the Geneva basin has been explored by two Swiss protistologists, Henri Blanc et Eugène Penard, who described six morphospecies of foraminifera with a single-chambered agglutinated test (Blanc 1886; Penard 1905). Since then, one species of agglutinated foraminifera (Thomas 1961) and three naked species belonging to the family Reticulomyxidae (Nauss 1949; Dellinger *et al.* 2014; Wylezich *et al.* 2014) have been described. Recently, a

soil species, *Edaphoallogromia australica*, has been described from tropical forest (Meisterfeld *et al.* 2001). According to the molecular phylogenies, freshwater foraminifera grouped into four clades (Lejzerowicz *et al.* 2010) that are not shared with marine species, except for *E. australica* which branches within a marine clade (Meisterfeld *et al.* 2001). *Edaphoallogromia australica* and *Reticulomyxa filosa* are the only two freshwater species that have been characterized both by morphology and molecular data. All other freshwater foraminifera are known exclusively through their rDNA sequences (Holzmann & Pawlowski 2002; Geisen *et al.* 2015) and very little is known about their diversity, evolution and ecology.

In Chapter 4 and 5, we attempt to fill this gap by largely expanding the environmental survey of freshwater foraminiferal populations in Geneva basin and by searching for morphospecies that would correspond to these sequences.

1.2.3. DNA barcoding and intragenomic polymorphism of foraminifera

Foraminifera are identified genetically using SSU rRNA gene (Pawlowski & Holzmann 2014). A fragment of this gene located in the 3' part, was chosen because of its sufficient resolution to distinguish species as well as its suitable size for Sanger sequencing (Pawlowski & Lecroq 2010). This fragment is composed of six hypervariable regions, which form expansion segments in the folded structure of the SSU rRNA (Figure 1.5). Three of these regions (corresponding to helices 43e, 45e, 49e) are common to all eukaryotes whereas the other three (37f, 41f, 47f) are specific to foraminifera. In order to collect and share the molecular data of foraminifera, an online barcoding database has been established (www.forambarcoding.unige.ch).

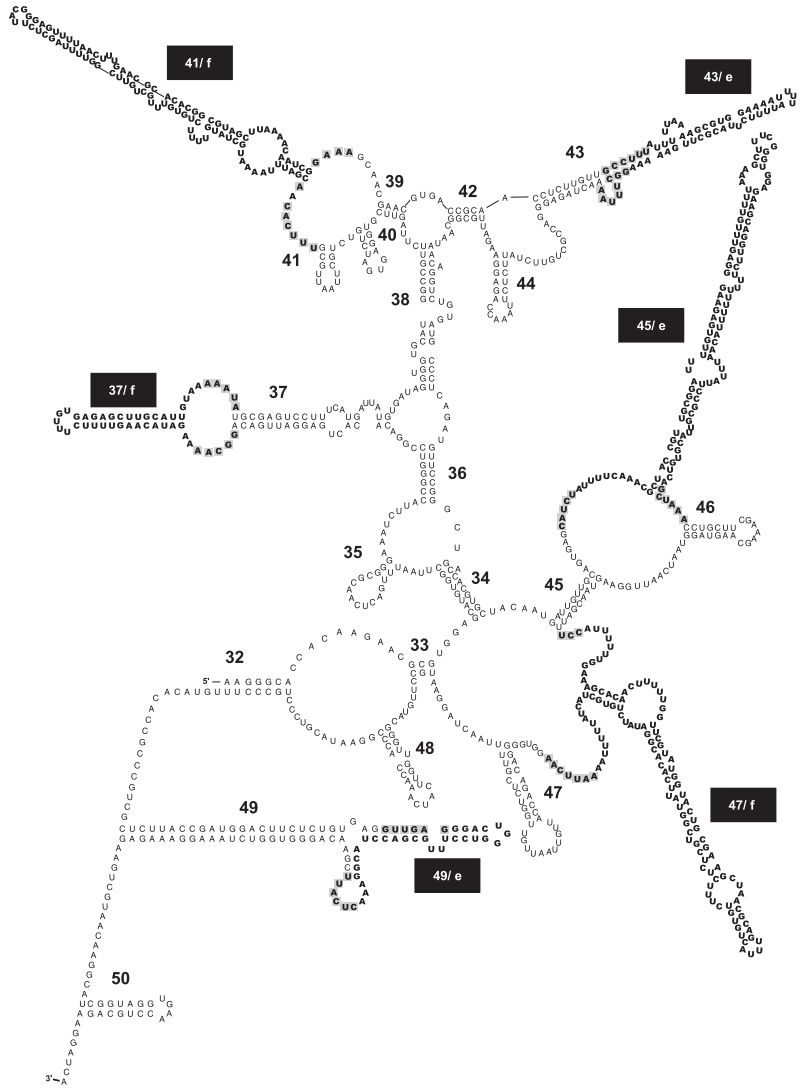


Figure 1.5 Predicted secondary structure of the SSU rRNA barcoding region of *Micrometula hyalostriata*. Helix marked with a “e” are common to all eukaryotes whereas those marked with a “f” are specific to foraminifera. *From* Pawlowski & Lecroq 2010.

One of the particularities of foraminifera is high divergence of their ribosomal genes. It is expressed by the presence of hypervariable regions specific to this group, which contribute to the extreme length of these genes that in some species may exceed 4000 nucleotides (Pawlowski 2000). Substitutions have been observed even in the most conserved regions, which often impedes using universal eukaryotic primers to amplify foraminifera in metabarcoding studies. Moreover, most interestingly, foraminifera show high level of intragenomic polymorphism, which often requires cloning PCR products even from single individuals.

The intragenomic variations of the rRNA genes are not unique to foraminifera. The phenomenon has also been observed in several other taxonomic groups, including metazoans (Escobar *et al.* 2012; Královál-Hromadová *et al.* 2012; Bik *et al.* 2013; Pereira & Baldwin 2016) and fungi (Chand Dakal *et al.* 2016; Thiéry *et al.* 2016; Wu *et al.* 2016). Several studies also reported intragenomic variation among protists such as dinoflagellates (Gribble & Anderson 2007; Miranda *et al.* 2012), ciliates (Gong *et al.* 2013), diatoms (Orsini *et al.* 2004; Alverson & Kolnick 2005), amoebozoa (Zlatogursky *et al.* 2016) and radiolarians (Decelle *et al.* 2014). The frequent presence of intragenomic variations in protists may be linked to high number of rRNA genes copies that can vary from 12'000 in dinoflagelattes and some microalgal strains (Zhu *et al.* 2005; Godhe *et al.* 2008) to 37'000 copies in diatoms (Godhe *et al.* 2008) and up to 170'000 in ciliates (Gong *et al.* 2013).

According to the recent study, Foraminifera possess about 10'000-30'000 rRNA gene copies (Weber & Pawlowski 2013). Intragenomic polymorphism has been found in many of them (Pillet *et al.* 2012; Weber & Pawlowski 2014). However, the true dimension of this particular feature was difficult to evaluate using classical cloning/Sanger sequencing approach. That is why we have further investigate this issue here using single-cell HTS approach applied to more than 130 specimens representing 23 species (Chapter 6)

1.3. Assessment of Water Quality

Effects of human activities have both impacts on marine and freshwater environment and assessing the health of aquatic ecosystems is therefore a priority for the management of the global environment. Several national and international legislations have been adopted to protect water bodies, including the Water Framework Directive in the European Union (WFD, Directive 2000/60/EC), the Clean Water Act of the Environmental Protection Agency in the USA (CWA, https://www.epa.gov/cwa-404/) or the Water Protection Ordinance in Switzerland (WPO, Swiss Federal Council 1998). One of the goals of those legislations is to evaluate the ecological status of aquatic ecosystems and maintain or restore it to achieve a "good ecological status", which corresponds to a reference status of the same type of environment not impacted by anthropogenic activities. The United Nations Convention on the Law of the Sea (UNCLOS, http://www.un.org/depts/los/convention_agreements/texts/unclos/UNCLOS-TOC.htm) is also partly dedicated to the protection and preservation of the marine environment, recommending that states have to take measures in order to prevent and reduce the pollution of chemical and biological origins. It is largely accepted that a single indicator is not sufficient to assess water quality and an integrative approach using several tools is recommended to conduct such evaluation (Boulton 1999). For example, to assess the sediment quality an integrative approach, called Sediment Quality Triad (SQT, Chapman 1990), combining chemistry, bioassays toxicity tests and analysis of benthic communities is recommended for scientists and environmental managers (McGee *et al.* 2009; Lopes *et al.* 2014; Moreira *et al.* 2017). Schematic representation of the water quality assessment is indicated in the Figure 1.6.

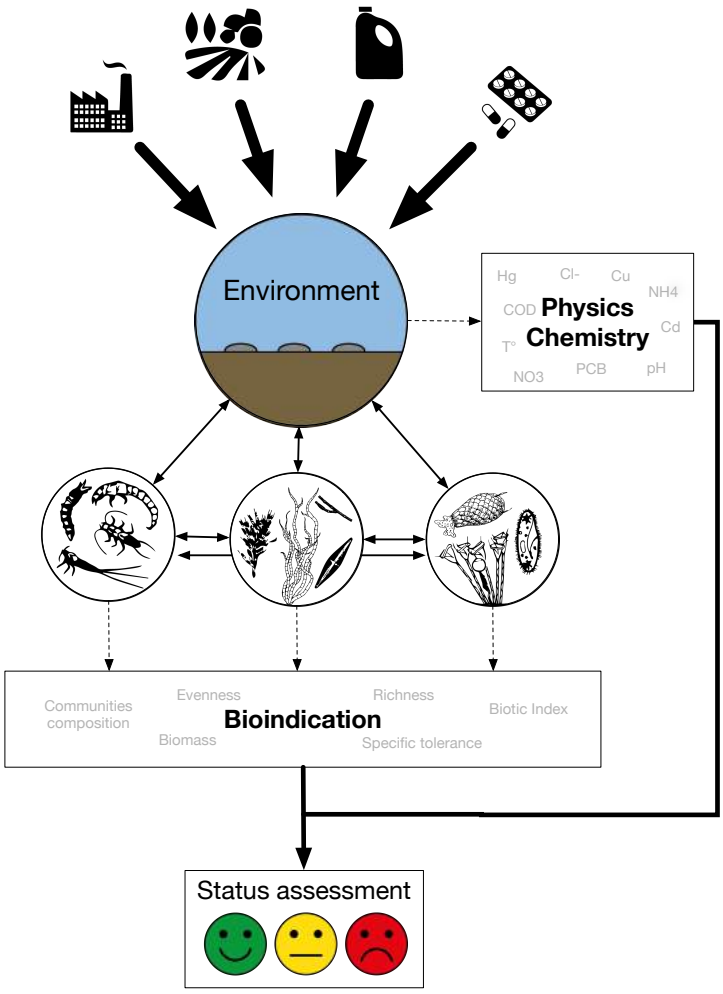


Figure 1.6 Schematic view of environmental assessment integrating physical, chemical and biological metrics.

1.3.1. Physical and chemical metrics

One category of tools used to evaluate the ecological status of aquatic ecosystems are the environmental parameters, such as the temperature, the turbidity, the flow, the land uses or the type of substrate (Phinn *et al.* 2005; Huang *et al.* 2016; Yerubandi *et al.* 2016; Cunha *et al.* 2016). Other classical environmental metrics widely used includes the physico-chemical parameters (pH, electrical conductivity), nutrients (the different chemical forms of nitrogen, the total phosphorus), organic matter (biological and chemical oxygen demand (BOD and COD)), specific dissolved ions (chloride, sulphates) (Chanudet *et al.* 2016; Boehler *et al.* 2017) but also heavy metals (Svobodová *et al.* 2017; Polidoro *et al.* 2017; Duivenvoorden *et al.* 2017). The

chlorophyll-a is often measured to estimate the algae biomass and hence evaluate the trophic condition of a waterbody (Jordaan & Bezuidenhout 2016; Pothoven *et al.* 2016). Recently, pesticides and pharmaceutical substances become widely used in water quality surveys (Teklu *et al.* 2016; Le Coadou *et al.* 2017; Munz *et al.* 2017).

1.3.2. Bioindication

Another category of tools used to evaluate the water quality are the biological parameters. The bioindication includes all biological methods based on individual organisms or communities that can be used to evaluate the status of an environment. Several parameters can be used to investigate the environmental impact: (1) the community composition and structure (abundance of the representative species, the total biomass of a group, the richness and evenness of species), (2) the tolerance of specific species to one or several specific parameters, or (3) the biological traits such as feeding or reproductive strategies, mobility, life cycle or lifespan (see O'Brien *et al.* 2016). To be qualified as bioindicator, the taxonomic group has to answer several criteria, which according to Arndt *et al.* (1987) include: (1) the wide knowledge and measurable reaction to environmental changes, (2) the ease of use for sampling, for taxa identification and for the archiving of the data (e.g. microscopic slides or ethanol preparation for a possible further reassessment) and (3) the presence of the bioindicator taxa across different environmental conditions and during all the year.

Within metazoans, several groups or species are widely used as bioindicators in water assessment. Among them the group with the largest previous records of ecological observations in both marine and freshwater environments are the benthic macro-invertebrates. They are known to be sensitive to several environmental impacts such as acidification (Johnson 2007; McFarland *et al.* 2010; Sandin *et al.* 2014), eutrophication (Ruse 2010; Böhmer *et al.* 2014; Jyväsjärvi *et al.* 2014; Bazzanti *et al.* 2017) and morphological alteration of the landscape (Gabriels *et al.* 2010; Miler *et al.* 2013; Urbanič 2014) (see Poikane *et al.* 2016). Fishes have been used to assess pollution with heavy metals and other contaminant substances (Authman & Abbas 2007; Birungi *et al.* 2007; Cerveny *et al.* 2016) or eutrophication level (Deegan *et al.* 1999; Sagouis *et al.* 2016).

Phototrophic organisms are also widely used as bioindicators of aquatic ecosystems. As primary producers, they occupy an important place in the food chain and are a major component of the eutrophication process, responding directly to the excess of nutrients (Misra *et al.* 2016). The main phototrophic groups targeted in biomonitoring are macrophytes (Orfanidis *et al.* 2014; Kennedy *et al.* 2016; Tarkowska-Kukuryk & Mieczan 2017), and periphyton, mainly composed of diatoms, (Lobo & Callegaro 2003; Bere 2016; MacDougall *et al.* 2016) or both of them (Feio *et al.* 2012; Gray 2015). The diatoms are particularly well known and information about their autecology is widely documented in the literature (Lowe 1974; Dam *et al.* 1994; see Lobo *et al.* 2016). They are known to be sensitive to water chemistry (Braak & Dame 1989; McCormick & Cairns 1994) and are widely used to assess the trophic status of rivers and streams, as well as indicators of heavy metal pollution (Galal & Farahat 2015; Lambert *et al.* 2016; Sánchez-Quiles *et al.* 2017).

Diatoms are one of the four groups of protists that are commonly used as bioindicators. According to Chen *et al.* (2016), microbial eukaryotes (protists) perform as good or even better than metazoans to assess ecological status. Compared to metazoans, protists show several advantages as bioindicators. They are widely distributed across the different aquatic habitats and usually are very abundant. Besides, they have relatively short generation time, which allows them to respond quickly to environmental changes. Finally, their small size, although sometimes hampering morphological identification, presents some advantages in term of sampling and preparation protocols (see Payne 2013).

Besides diatoms, the three groups of protists bioindicators are testate amoebae, foraminifera and ciliates. Several studies showed the impact of environmental changes on testate amoebae communities in different environments such as soil (Wanner & Dunger 2001), peatbog (Turner & Swindles 2012; Valentine *et al.* 2013) but also lakes (Patterson *et al.* 2013; Roe & Patterson 2014). The sensitivity of testate amoebae to heavy metals has also been demonstrated (Nguyen-Viet *et al.* 2007; Yang *et al.* 2011). However, compared to other groups, they use of testate amoebae as bioindicators is still very scarce.

Benthic foraminifera are together with macro-invertebrates the main group of indicators of environmental changes impacting marine environment. The forams have been used to assess the heavy metals contaminations (Cadre *et al.* 2003;

Bergin *et al.* 2006), the organic enrichment of the sediment (Debenay *et al.* 2009; Vidovic *et al.* 2009, 2014) or to measure the impact of offshore oil drilling on the seafloor (Mojtahid *et al.* 2008; Denoyelle *et al.* 2010). Calcareous and agglutinated foraminifera are convenient because their tests are preserved in the sediment, even after death of the cell, which allow historical reconstruction of the environmental changes (Alve *et al.* 2009; Gooday 2009). The hard-shells of foraminifera can also be easily identified because of large spectrum of distinctive morphological features.

Ciliates are a well-known indicator of activated sludge in wastewater treatment (Martín-Cereceda *et al.* 1996; Nicolau *et al.* 2001) but they have also been shown to respond to organic pollution (Madoni & Bassanini 1999; Jiang *et al.* 2013; Xu *et al.* 2014). An index, based among others on ciliates, has been established to access the trophic status of streams (DIN 38 410, Berger *et al.* 1997).

1.3.3. Indices

A powerful way to integrate biological metrics into a tool usable in routine assessment is the establishment of biological indices. The basic principle is to combine all the information into one index that could be assigned to an ecological status (usually several classes between very good and very bad quality). Several indices have been developed for different taxonomic groups and environments, including three main classes of indices: the multimetric index, the predictive models and the indicator index.

The multimetric indices combine several metrics into a single index. For example, the Index of Biological Integrity (IBI) has been developed in Ohio (USA) to measure the human impact on the water quality of streams. In this index, twelve metrics involving fishes are used, including species identification, percentage of tolerant species, percentage of species with various life-cycle stages, and percentage of the different trophic groups. Each metrics are assigned to a score and the sum of those scores gives a single ecological evaluation. Another multimetric index, the Vegetation Index of Biological Integrity (VIBI) is used to assess wetland water quality based on plants. The metrics used by this index include the richness, the coverage, the density or the frequency of different category of plants (tree, shrub, grass, vine or forb).

Predictive models use statistical methods to predict ecological status based on previous knowledge. One example is the AQUAFLORA (Feio *et al.* 2012), which assess the water quality of streams based on diatoms and macrophytes communities. It takes into account all available variables, including physico-chemical metrics, and the assessment is done based on the distance to the reference community. Other predictive models have been developed using either diatoms (Feio *et al.* 2009) or macrophytes only (Aguiar *et al.* 2011). RIVPACS, AUSRIVAS and BEAST are predictive models involving macro-invertebrates communities (Reynoldson *et al.* 1995; Smith *et al.* 1999; Wright *et al.* 2000). They were later adapted to include also diatoms, fishes or macrophytes (Joy & Death 2002; Mazor *et al.* 2006; Carlisle *et al.* 2008; Aguiar *et al.* 2011) (see Feio & Poquet 2011).

The last group of most widely used indices are the indicator indices or biotic indices. This kind of indices is based on the classification of species, or higher taxonomic units, into specific classes corresponding to water quality status. This classification is based on the ecological tolerance and preferences of each species to one or several parameters. The Table 1.1 summarizes some of the most used biotic indices across the world with the targeted taxonomic group. This list is far from being exhaustive and a lot of specific indices have been established in different geographical regions, especially since the WFD has been adopted. However, the number of taxa used as bioindicators is relatively limited; for example the assessment of streams is mainly conducted with macro-invertebrates and diatoms. There is still a worldwide paucity of well-established indices for biological evaluation of lakes and relatively few indices exist for the assessment of marine environments.

Acronym	Habitats	Target group	Reference
IBGN	Streams	Macroinv.	AFNOR 2004
IBCH	Streams	Macroinv.	Stucki 2010
BMWP	Streams	Macroinv.	Armitage *et al.* 1983
BI	Streams	Macroinv.	Hilsenhoff 1988
SIGNAL	Streams	Macroinv.	Chessman 1995
TDI	Streams	Diatoms	Kelly *et al.* 1995
SPI	Streams	Diatoms	CEMAGREF 1982
DI-CH	Streams	Diatoms	Hürlimann & Niederhauser 2007
IBD	Streams	Diatoms	AFNOR 2000
TIM	Streams	Macrophytes	Schneider & Melzer 2003
RMI	Streams	Macrophytes	Kuhar *et al.* 2011
LBI	Lake	Macroinv.	Verneaux *et al.* 2004
AMBI	Marine	Macroinv.	Borja *et al.* 2000
BENTIX	Marine	Macroinv.	Simboura & Zenetos 2002
BQI	Marine	Macroinv.	Rosenberg *et al.* 2004
ITI	Marine	Macroinv.	Word 1979

Table 1.1 List of some of the most commonly used biotic index in streams, lakes and marine environments.

1.3.4. Assessment of water quality in rivers and streams in Switzerland

To answer the requirements of the Swiss decree on water protection (Swiss Federal Council 1998) and to ensure a global management of watercourses the Federal Office for the Environment (FOEN), in collaboration with the Swiss Federal Institute of Aquatic Science and Technology (EAWAG), proposed a stepwise procedure composed of different complementary modules to assess the quality of surface water (http://www.modul-stufen-konzept.ch/index_EN). The evaluation is performed on several scales, the regional, the watershed and the water bodies specific sections. Not all the modules are applied to all the scales but at the end the different modules at different scales are integrated to provide the best possible assessment of the ecosystem.

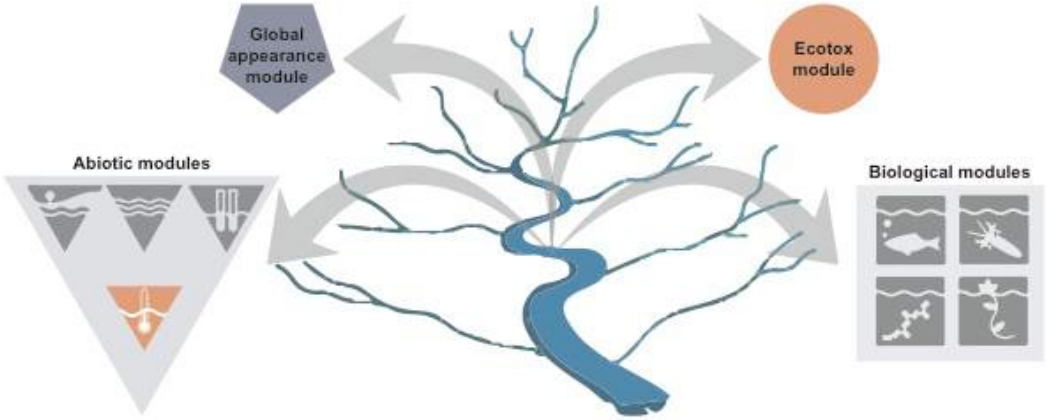


Figure 1.7 Schematic representation of the ten modules available in the stepwise procedure in Switzerland. They are grouped into four categories: the Global appearance module, the ecotoxicology module, the abiotic modules (Ecomorphology, Hydrology, Water Chemistry and Temperature) and the biological modules (fish, macroinvertebrates, macrophytes and diatoms). Modules under development are coloured in orange. *Adapted from* the FOEN website.

For running waters, ten modules (Figure 1.7) are available and each module gives a status divided into classes represented by a colour code. The Appearance module provides a global evaluation of the stream at regional scale. It is based on turbidity, coloration of the water, scum, odour, presence of iron sulphide, level of clogging, the type of vegetation and the presence and abundance of solid wastes, heterotrophic organisms and filamentous algae. Each parameter gives an independent evaluation, represented by a colour code. Four abiotic modules (ecomorphology, hydrology, temperature, and water chemistry) are based on physico-chemical parameters referring to two levels: the regional scale and the watershed scale. The Ecomorphology module comprises metrics like the width of the river bed, the substrate of the channel bed or the structure of the riparian zone. The Hydrological module, based on the regional scale, is designed to identify the different impacts on water management (e.g. hydroelectric dam) and evaluate their implications on the water flow. The module Temperature, which gives an estimate on the grade of "neutrality" of the streams, is under development. The Water Chemistry module is an important module that regroups all analyses needed to evaluate the physico-

chemical quality of the water. A list of the parameters used in the method with the threshold for each class is summarized in Table 1.2. A minimum frequency of one sample per month for each station is required. Results from chemical analysis are finally grouped into one appreciation, which reflects the global legal requirements.

Chemical [mg/L]	Criteria	Very good	Good	Average	Poor	Very poor
Total phosphorus	filtered	< 0.025	0.025 – 0.05	0.05 – 0.075	0.075 – 0.1	> 0.1
Orthophosphate		< 0.02	0.02 – 0.04	0.04 – 0.06	0.06 – 0.08	> 0.08
Nitrates		< 1.5	1.5 – 5.6	5.6 – 8.4	8.4 – 11.2	> 11.2
Nitrites	Cl^- < 10	< 0.01	0.01 – 0.02	0.02 – 0.03	0.03 – 0.04	> 0.04
Nitrites	Cl^- 10–20	< 0.02	0.02 – 0.05	0.05 – 0.075	0.075 – 0.1	> 0.1
Nitrites	Cl^- > 20	< 0.05	0.05 – 0.1	0.1 – 0.15	0.15 – 0.2	> 0.2
Ammonium	T > 10°C or pH > 9	< 0.04	0.04 – 0.2	0.2 – 0.3	0.3 – 0.4	> 0.4
Ammonium	T < 10°C	< 0.08	0.08 – 0.4	0.4 – 0.6	0.6 – 0.8	> 0.8
Dissolved organic carbon (COD)		< 2	2 – 4	4 – 6	6 – 8	> 8
Total nitrogen		< 2	2 – 7	7 – 10.5	10.5 - 14	> 14
Biochemical oxygen demand (BOD)		< 2	2 – 4	4 – 6	6 – 8	> 8

Table 1.2 Chemical parameters and threshold for each ecological class used in the module Water Chemistry from the Swiss Modular stepwise Procedure.

An Ecotoxicology module, at the junction between chemistry and biology, based on the standardisation of bioassays for routine assessment is under development.

Finally, four biological modules are based on the classical bioindicator communities: macro-invertebrates, fishes, macro-phytes and diatoms.

The Macro-invertebrates module (IBCH) is based on the French index IBGN (Indice Biologique Global Normalisé, AFNOR 2004) and is suitable for small and medium sized streams. The main difference with the French index is the sampling methodology. The index value is determined in function of the richness of the sample and the determination of the faunistic indicator group. Only one sampling is recommended per year for the regional scale of this index.

The Fish module is based on four metrics: (1) the composition and dominance of community, (2) the population structure of indicator species, (3) the density of indicative species, and finally (4) the observation of morphological deformation. The global assessment is an integration of all those metrics that classify the site into one of the five ecological classes.

The Macrophytes module is still under elaboration, the indications given by the FOEN only concern sampling in order to harmonize the database and therefore give the possibility to develop a macrophyte-based assessment method adapted to the Swiss ecosystems. Meanwhile, several indices developed in other European countries (AFNOR 2003; Schaumburg *et al.* 2004; Pall & Moser 2009) can be used with the collected data.

The Diatoms module is based on the community of diatoms present in the epilithic biofilm of streams. This method was developed to assess the water quality of small and medium streams and is representative of the chemical quality related to anthropogenic impact. The calibration of the index has been performed on more than 3'500 sites, including about 1'200, for which the chemical assessment was also available. Six chemicals parameters have been selected for the evaluation of the Diatom index (ammonium, nitrite, total nitrogen, total phosphorus, chloride and Dissolved organic carbon). For about 200 diatom species, the authors of the index inferred two ecological values representing the optimal and tolerance range related to the organic pollution on a scale of 1 to 8. The autecological values of those species are available on the web site of the modular system (http://www.modul-stufen-konzept.ch/fg/module/diatomeen/index_FR). Those data contain also for each species its geographic distribution, the associated chemical values for the six parameters selected for the calibration, the number of occurrence per site and the

distribution of the given species across the different ecological status (based on chemical and DI-CH evaluations) for all the sites used for the initial calibration of the index. The index is then calculated following the weighted average equation of Zelinka & Marvan (1961) where D corresponds to the optimal ecological conditions, G is the weighting factor corresponding to the tolerance range and H the relative abundance. D, G and H are specific for each species *i* and *n* corresponds to the number of species found in a given sample. The inferred index is then classified into the 5 ecological classes used by the stepwise procedure.

$$DICH = \frac{\sum_{i=1}^{n} D_i \, G_i \, H_i}{\sum_{i=1}^{n} G_i \, H_i}$$

All these ten modules are detailed with a lot of precision on how to proceed with the sampling, the evaluation, analysis and valorisation of data. Difficulties and limitations of each method are also mentioned in the protocols. For the establishment of the final report, the evaluations of different modules based on the five colour-codes scale are combined per site. To evaluate a site, the three abiotic modules (Chemistry, Ecomorphology and Hydrology) and the Appearance module are mandatory. In exceptional cases, the Chemistry module can be replaced by the Diatoms module. Moreover, four modules including two biological are the minimum required to allow the evaluation of the site. The cartographic synthesis for a river is presented as example in the Figure 1.8.

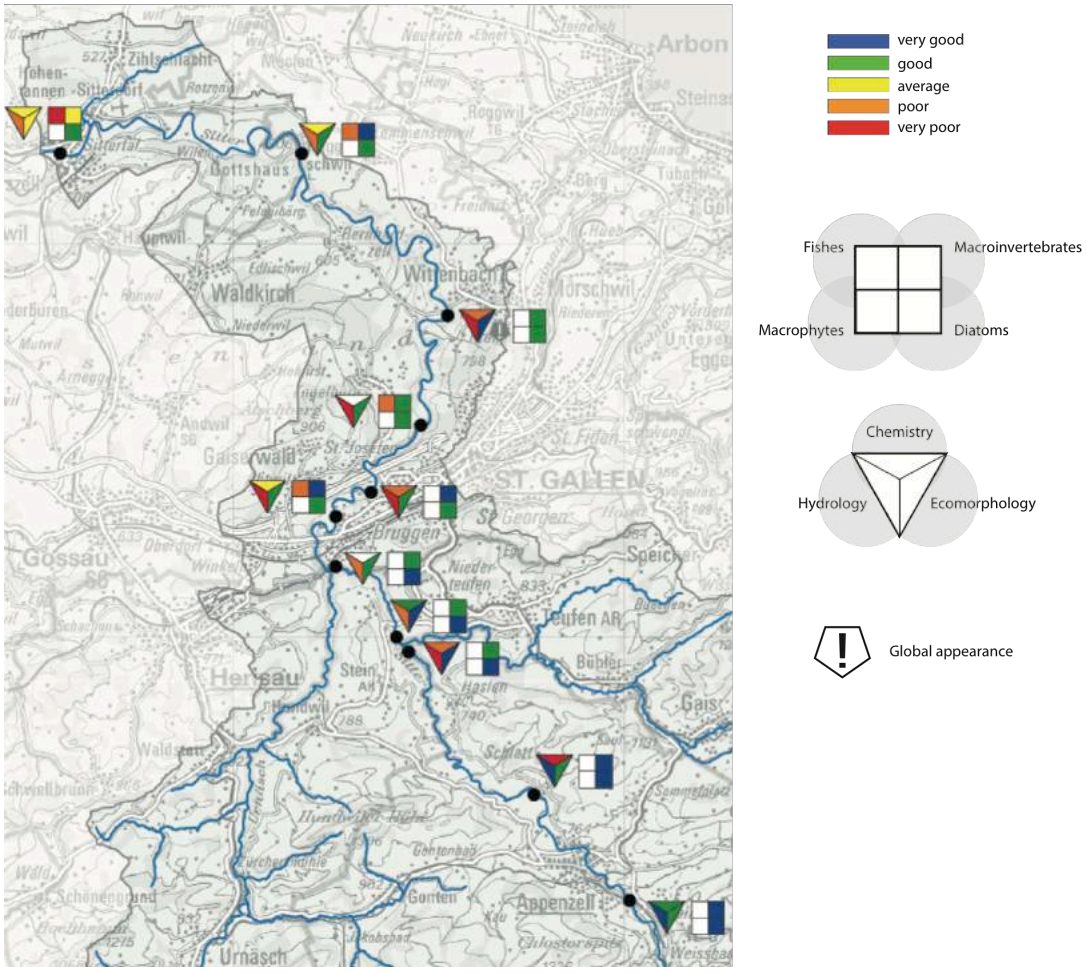


Figure 1.8 Cartographic representation of a modular system in the case of a specific stream (Sitter, CH). The appearance module is mentioned only if at least one parameter reflects of a bad state. From http://www.modul-stufen-konzept.ch/

The global interpretation of the ecosystem health is not only based on different parameters of water quality, but it also incorporates the ecological expertise of the specialists who know local biodiversity and the environment. Biotic indices represent a powerful tool to provide an overview, to target problematic sites and to communicate with different parties involved in water management.

Although the Swiss legislation includes also the importance and obligation to assess and monitor the ecological status of lakes, no tools are yet available, except a module on the ecomorphology of the shores. The FOEN highlights this absence and encourages researchers to develop new modules.

1.4. Molecular indices

1.4.1. Limitations of traditional approaches

Current traditional biotic indices are based on the morphological identification of organisms. Different taxonomic levels, from species, genus to family level, are considered as sufficient depending on the type of index and group of interest. Morphological identification requires not only high quality equipment (microscopes or binoculars) but above all highly trained and expertised taxonomists (Manoylov 2014). According to this author, a good taxonomist should possess some particular qualities such as the patience, the curiosity, and pay particular attention to details. In routine work, an experienced taxonomist may spend about 4h just for the identification and counting of a single diatom sample (Manoylov 2014). Moreover, the need to process a lot of samples to cover large-scale assessment can easily lead to misidentification of species. It is an important issue to be considered when the robustness of indices is discussed, even if some authors suggest that the misidentification errors do not affect the evaluation because of a buffering effect (Stevenson *et al.* 2010). The taxonomists often organise inter-calibration exercises to evaluate and reduce the impact of the identification errors (Kelly *et al.* 2009, 2014; Kahlert *et al.* 2009; Birk *et al.* 2013). The results of these exercises show that the variance between sample assessments is mostly due to the taxonomic errors rather than to the preparation of samples or sample replicates (Prygiel *et al.* 2002; Lavoie *et al.* 2005). The taxonomic expertise is particularly important for small organisms such as diatoms that need to be identified to species level more than for macroinvertebrates, where the identification to family level is often sufficient for large scale monitoring program (Chessman 1995; Hewlett 2000).

1.4.2. HTS metabarcoding based to biomonitoring

As mentioned in the section 1.1.3, DNA metabarcoding are widely used in several fields of applications. Chariton *et al.* (2010) performed a pioneer study that proposed to use metabarcoding for biomonitoring. During following years, several review papers highlight the importance of integrating the HTS metabarcoding approach in the monitoring surveys (Baird & Hajibabaei 2012; Taberlet *et al.* 2012; Woodward *et al.* 2013; Wood *et al.* 2013) showing the strong impact that these new molecular

technologies can have on our concept of biodiversity monitoring. DNA metabarcoding proved to be useful in community analysis to correctly recover of alpha and beta-diversity (Yu *et al.* 2012; Ji *et al.* 2013; Leray & Knowlton 2015), encouraging its further use in routine bioassessment. Moreover, several studies highlights the time and cost effectiveness of HTS-based bioassessment compared to traditional approach (Yu *et al.* 2012; Ji *et al.* 2013; Stein *et al.* 2014). The extracellular DNA approach has also been exploited for monitoring diversity, presenting the advantage of being non-invasive for the targeted organisms and allowing only one sampling strategy for all taxonomic groups (Bohmann *et al.* 2014; Hoffmann *et al.* 2016; Deiner *et al.* 2016). More recently, Bohan *et al.* (2017) have even come up with the idea of a fully automated system based on machine-learning methods to reconstruct the networks of ecological interactions from environmental DNA data.

Different taxonomic groups of bioindicators have been investigated using HTS technologies either in freshwater (Table 1.3) or marine environments (Table 1.4). In freshwater ecosystems most of studies focus on diatoms and macro-invertebrates, the two groups that are also the most commonly investigated in morphological surveys (Table 1.3). Few studies also concern freshwater bacterial communities, which are usually not included in conventional biomonitoring. Interestingly, molecular investigations of communities living in rivers and streams are much more frequent than the studies of lakes, which mainly focus on bacterial metabarcoding (Eiler *et al.* 2013; Chen *et al.* 2016). In marine environment, most of biomonitoring-related metabarcoding studies also focus on microbial organisms, including global surveys of prokaryotic and eukaryotic communities as well as the specific groups of protists such as foraminifera or dinoflagellates.

The main topic of interest of all these pilot studies is the reliability and the accuracy of the HTS metabarcoding data in order to obtain exactly the same information about ecological status as the morphological survey. Indeed, several studies focus on the direct comparison of both dataset (Zimmermann *et al.* 2014; Groendahl *et al.* 2017), but only few of them reach the step of inferring the index values from molecular dataset (Kermarrec *et al.* 2013; Aylagas *et al.* 2014; Visco *et al.* 2015 (Chapter 7); Lejzerowicz *et al.* 2015; Aylagas *et al.* 2016; Pawlowski *et al.* 2016a; Aylagas *et al.* 2017). In general, the studies comparing molecular and morphology-based indices are giving very promising results. However, their authors insist on the

incompleteness of the reference database for the species required in the different index calculations. Up to now, only two studies inferred new indices analysing HTS dataset of eukaryotic communities without reference to the morphotaxonomy, in diatoms (Apothéloz-Perret-Gentil *et al.* 2017 - Chapter 8) and foraminifera (Cordier *et al*. - submitted).

Taxon	Marker	Type of	Main issues	Reference
Bacteria	16S	Natural	Bioassays	Binh *et al.* 2014
Bacteria	16S	Natural	Lake	Chen *et al.* 2016
Bacteria/fungi	16S/ITS2	Natural	Land-water interface	Veach *et al.* 2015
Phytoplankton	16S	Natural	Lake	Eiler *et al.* 2013
Diatoms	18S, *rbc*L, COI	Mock	Taxonomic assignment	Kermarrec *et al.* 2013
Diatoms	18S, *rbc*L	Natural	SPI index	Kermarrec *et al.* 2014
Diatoms	*rbc*L	Natural	SPI Index	Vasselon *et al.* 2017
Diatoms	18S V4	Natural	NGS vs morphology	Zimmermann *et al.* 2014
Diatoms	18S V4	Natural	DI-CH index	Visco *et al.* 2015
Diatoms	18S V4	Natural	DI-CH index tax-free	Apothéloz-Perret-Gentil *et al.* 2017
Diatoms	18S V4	Mesocosm	NGS vs morphology	Groendahl *et al.* 2017
Chironomids	COI, CytB	Mock	Marker resolution	Carew *et al.* 2013
Macroinv	COI	Natural	Shotgun sequencing	Zhou *et al.* 2013
Macroinv	COI	Mock/Natural	Bulk samples	Hajibabaei *et al.* 2011
Macroinv	COI	Natural	Ethanol samples	Hajibabaei *et al.* 2012
Macroinv	COI	Natural	Gene enrichment	Dowle *et al.* 2016
Macroinv	COI	Mock	Primer bias	Elbrecht & Leese 2015
Macroinv	COI	Mock	Primers design	Elbrecht & Leese 2017
Macroinv	COI	Natural	Index	Elbrecht *et al.* 2017
Macroinv	COI	Natural	Diversity metrics	Gibson *et al.* 2015
Oligochaetes	COI	Mock	IOBS index	Vivien *et al.* 2016b
Oligochaetes	COI	Mock	Formalin fixation	Vivien *et al.* 2016a
Fish/amph.	12S/16S	Mesocosm	Quantification	Evans *et al.* 2016

Table 1.3 List of HTS metabarcoding studies focused on freshwater biomonitoring, classified according to the indicator taxa, genetic marker, type of communities and technical issues.

Taxon	Marker	Type of	Main issues	Reference
Bacteria	16S	Natural	microgAMBI index	Aylagas *et al.* 2017
Bacteria	16S	Natural	Marine aquaculture	Dowle *et al.* 2015
Bacteria	16S	Natural	Oil spill biosensors	Smith *et al.* 2015
Bacteria	16S 18S V4	Natural	Sequencing platform	Ferrera *et al.* 2016
Eukaryotes	18S	Natural	Oil spill assessment	Bik *et al.* 2012
Eukaryotes	18S	Natural	Estuary	Chariton *et al.* 2010,
Eukaryotes	18S V4 V9	Natural	Ballast Water	Lohan *et al.* 2016
Eukaryotes	COI	Natural	Ballast Water	Zaiko *et al.* 2015
Eukaryotes	18S V9	Natural	Communities variations	Brannock *et al.* 2016
Eukaryotes	18S V7	Natural	Marine canyons	Guardiola *et al.* 2015
Eukaryotes	18S	Natural	Estuaries	Lallias *et al.* 2015
Phytoplankton	23S	Natural	Cost-effective protocol	Yoon *et al.* 2016
Foraminifera	18S 37f	Natural	Marine aquaculture	Pawlowski *et al.* 2014a
Foraminifera	18S 37f	Natural	Marine aquaculture –	Pochon *et al.* 2015
Foraminifera	18S 37f	Natural	Marine aquaculture – forams index	Pawlowski *et al.* 2016a
Foraminifera	18S 37f	Natural	Oil drilling sites	Laroche *et al.* 2016
Foraminifera	18S 37f	Natural	Marine aquaculture - machine learning	Cordier *et al*. submitted
Dinoflagellates	18S V4 LSU	Mock/Natural	Marker comparison	Smith *et al.* 2017
Meiofauna	18S V9	Natural	Extraction, data analysis	Brannock & Halanych 2015
Macroinv	COI	Mock/Natural	gAMBI index	Aylagas *et al.* 2016
Macroinv	COI, 18S	Natural	gAMBI index, database	Aylagas *et al.* 2014
Marcoinv	18S V4	Natural	Marine aquaculture - ITI and AMBI Index	Lejzerowicz *et al.* 2015
Macroinv	COI, 18S	Natural	Seagrass communiity	Cowart *et al.* 2015

Table 1.4 List of HTS metabarcoding studies focused on marine biomonitoring, classified according to the indicator taxa, genetic marker, type of communities and technical issues.

CHAPTER 2

ARNOLDIELLINA FLUORESCENS GEN. ET SP. NOV. – A NEW GREEN AUTOFLUORESCENT FORAMINIFER FROM THE GULF OF EILAT (ISRAEL)

LAURE APOTHÉLOZ-PERRET-GENTIL, MARIA HOLZMANN, JAN PAWLOWSKI

Published in European Journal of Protistology, **49**, 210–216, 2013

2.1. Project description

Sediment samples from the Red Sea were maintained in enriched medium to establish cultures of foraminifera. The new allogromiid showed up quite long time after the sample was brought to the lab. Its reticulopodial network was impressive, spreading out rapidly at the bottom of the culture dish. We wanted to look at it more closely and since I was doing a lot of fluorescent microscopy at this time, I looked with UV light excitation. This is how I found out the astonishing characteristic of this species. Although the species survived only for a relatively short time in the culture, I could collect sufficient number of specimens to describe it.

2.2. Abstract

A new monothalamous (single-chambered) soft-walled foraminiferal species, *Arnoldiellina fluorescens* gen. et sp. nov., was isolated from samples collected in the Gulf of Eilat, Israel. The species is characterized by a small elongate organic wall with a single aperture of allogromiids. It is characterized by the emission of green autofluorescence (GAF) that has so far not been reported from foraminifera. Phylogenetic analysis of a fragment of the 18S rDNA indicates that the species is related to a group of monothalamous foraminiferans classified as clade I. Although the morphology of the new species is very different compared to the other members of this clade, a specific helix in 18S rRNA secondary structure strongly supports this position.

2.3. Introduction

Foraminifera are a large and diverse group of protists well known from marine environments (Murray 2006) but also found in terrestrial and freshwater habitats (Meisterfeld *et al.* 2001; Lejzerowicz *et al.* 2010). Most foraminiferans produce either 'single chambered' (monothalamous) or 'multi-chambered' (polythalamous) tests, with organic, agglutinated or calcareous walls while some of them lack a test at all (athalamids). Foraminiferal research has focused largely on polythalamous calcareous species, whose hard-walled shells are well preserved in the fossil record (Haynes 1981; Murray 2006). The diversity of soft-walled monothalamous foraminifera, also called allogromiids, remains largely unknown as they are poorly preserved. The interest in this group increased recently, due to their abundance in the deep-sea and polar regions (Gooday 2002; Gooday *et al.* 2005) and their application in genomic studies (Habura *et al.* 2005; Parfrey & Katz 2010). Many new monothalamous species have been described in the last decade (Gooday & Pawlowski 2004; Sabbatini *et al.* 2004; Gooday *et al.* 2004; Altin *et al.* 2009; Pawlowski & Majewski 2011).

The study of monothalamous foraminiferans was also prompted by the development of molecular systematics, which greatly facilitated the identification of their morphologically rather featureless tests. Molecular studies completely changed our view of their phylogenetic relationships and led to the discovery of a huge diversity in

this group (Pawlowski *et al.* 2002a; b, 2003). A new dimension of monothalamiid diversity was revealed by environmental DNA surveys of foraminiferal assemblages (Habura *et al.* 2004, 2008; Lecroq *et al.* 2011; Pawlowski *et al.* 2011b).

The new monothalamid species described here was discovered in a culture dish containing sediment and algal debris from the Gulf of Eilat (Israel). Molecular analysis of three specimens showed that they all belong to the same species that is genetically well distinguished from other monothalamids, resulting in a description of a new species and new genus.

2.4. Materials and Methods

2.4.1. Isolation and culture

Specimens were isolated from surface sediment samples collected by SCUBA diving at 5–10 m in front of the Inter-university Institute for Marine Sciences (IUI), near Eilat, Israel, on January 2011.

The sediment was distributed in two Petri dishes and cultured in Erdschreiber medium (5% soil extract, 1mM $NaNO_3$, 0.07 mM Na_2HPO_4, 10 mM Tris, pH = 8, filled up with sterile seawater) and filtered seawater. A few drops of heat killed *Dunaliella salina* (Chlorophyceae) were added for nutrition every two weeks. Specimens with extended pseudopodia were first observed in culture dishes 6 months after collection. The specimens were abundant during a period of 3 months, but later disappeared from the dish and have not been observed again.

2.4.2. Fixation and colouration

Cultured specimens were transferred by means of a pipette to a 10% formalin solution. They were fixed for 1 h at room temperature and afterwards washed briefly in PBS (Phosphate Buffered Saline, 137 mM NaCl, 2.7 mM KCl, 10 mM Na2HPO4, and 2 mM KH2PO4, adjusted pH to 7.4). A final immersion lasting 30 min was carried out in a dark room at ambient temperature using 4′,6′-diamidino-2-phénylindole (DAPI) at 5.10E−4 mg/ml to stain and subsequently identify nuclei.

2.4.3. Morphological studies

Living and fixed specimens were observed with an inverted microscope (Nikon Eclipse Ti) and a fluorescence micro- scope (Nikon Eclipse E200). Photographs were taken with Leica DFC 420C and Nikon Digital DXM 1200 cameras. Videos were made with the Imaging Source DFK 41AF02 camera. They are available in the online version of this article and at http://forambarcoding.unige.ch/movies.

2.4.4. Molecular analyses

DNA from 13 specimens was extracted in guanidine lysis buffer (Pawlowski 2000), each extraction was performed with a single specimen. PCR amplifications of a fragment of the 18S rDNA were performed using the primer pair s14F3 (5′ACG CA(AC) GTG TGA AAC TTG) and 20R (5′GAC GGG CGG TGT GTA CAA). PCR products were re- amplified using the nested primer s14F1 (5′AAG GGC ACC ACA AGA ACG C) and 20R. PCR amplifications for a shorter fragment of the 18S rDNA were performed using the primer pair s14F3 and s17 (5′CGG TCA CGT TCG TTG C). PCR products were re-amplified using the nested primer s14F1 and s17. The amplified PCR products were purified using High Pure PCR Purification Kit (Roche Diagnostics) and sequenced directly. Sequencing reactions were performed using the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) and analysed on a 3130XL Genetic Analyser (Applied Biosystems). The 3 new sequences reported in this paper were deposited in the EMBL/GenBank database (accession numbers HE775247–HE775249). The secondary structure was created using the RNAfold program from the University of Vienna (Gruber *et al.* 2008).

The obtained sequences were aligned to 57 other foraminiferans using Seaview v.4.3.3. software (Gouy *et al.* 2010). After elimination of the highly variable regions, 869 sites were left for analysis. The phylogenetic tree was constructed using maximum likelihood method based on the GTR + G model, using RAxML BlackBox (Stamatakis et al. 2008).

2.5. Results

Systematics

Supergroup RHIZARIA Cavalier-Smith, 2002 Phylum FORAMINIFERA D'Orbigny, 1826 Genus *Arnoldiellina* gen. nov.

Type species: *Arnoldiellina fluorescens* sp. nov.

Etymology: The genus was named in honour of Zach Arnold, Professor Emeritus of Palaeontology at the University of California, Berkeley who described several monothalamous foraminiferans and studied their life cycles and evolution.

Diagnosis: Test free, monothalamous, fusiform, <300 µm in length and <70 µm in width; organic wall transparent from 2 to 7 µm in width, thicker around the aperture. The single aperture is funnel-shape with a tubular internal extension. Multinucleate cytoplasm (up to 11 nuclei); granular, in constant rapid movement. Reticulopodes very active with rapidly forming reticulopodial network and fast moving granules. Specimens emit GAF, which disappeared with fixation.

Remarks: The new genus was introduced because the species is morphologically very different from previously described genera and our phylogenetic analyses do not show any close relationship with other sequenced monothalamous species.

Arnoldiellina fluorescens sp. nov.

Holotype: MHNG INVE 82002.

Type material: A specimen preserved in formalin was selected as holotype and deposited at the Museum of Natural History in Geneva (MHNG) together with 7 paratypes (MHNG INVE 82003).

Type locality: Gulf of Eilat, Israel.

Other material examined: 35 additional specimens were either extracted in guanidine (13 specimens), preserved in formalin (8 specimens), fixed for DAPI staining (4 specimens) or observed and photographed alive (10 specimens). The rapid streaming of protoplasm can be observed in the two videos. The multiple nuclei in *Arnoldiellina* are shown in Figure 2.2D.

Etymology: The species name is based on the ability of this foraminifer to emit green autofluorescence.

Diagnosis: As for genus.

Description: Measurement of length and width of 16 different specimens are shown in the Table 2.1. All specimens were fusiform; however, the ratio length/width may vary between the specimens. The length is 3–5 times the width. The most compressed specimens is the one shown in Figure 2.2(C, D); one of the paratypes shown in Figure 2.2B was the most elongated.

Description of the holotype: Test free, monothalamous, fusiform, 270 m in length and 65 m in width, organic wall transparent of 6.6 m width; the wall increases in thickness around the single terminal aperture. The aperture is funnel- shaped with a tubular extension inside the test.

Remarks: Compared to the other species assembled in clade I, *Arnoldiellina* differs in its morphology by its small size and organic wall. All other members of clade I are characterized by the presence of cell bodies and agglutinated tests surrounding them.

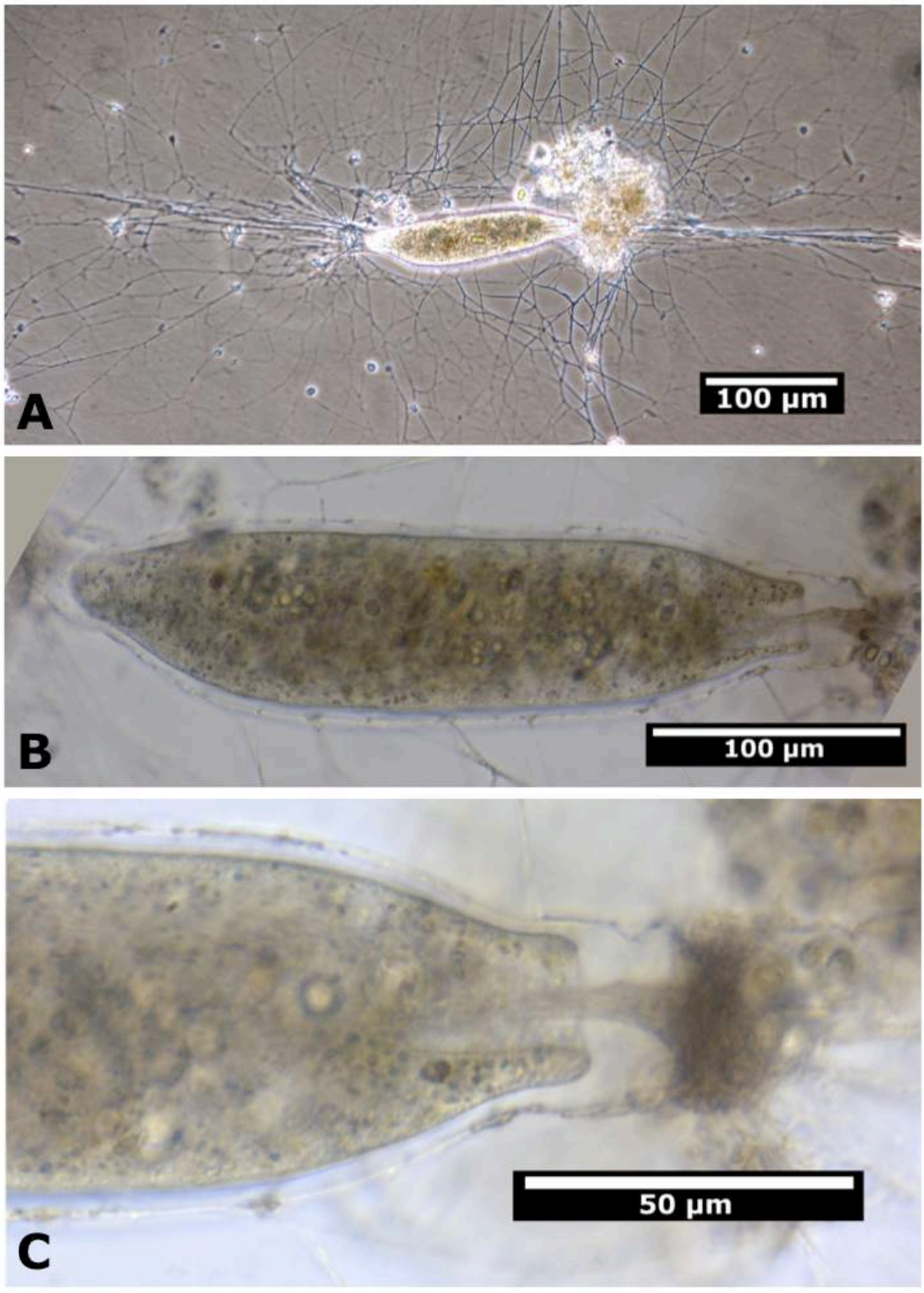


Figure 2.1 Living specimens of *Arnoldiellina fluorescens*, gen. and sp. nov. Overview of granuloreticulopodial network (A). View of a specimen (B) with close up of the terminal aperture (C). Pictures were taken with differential interference contrast.

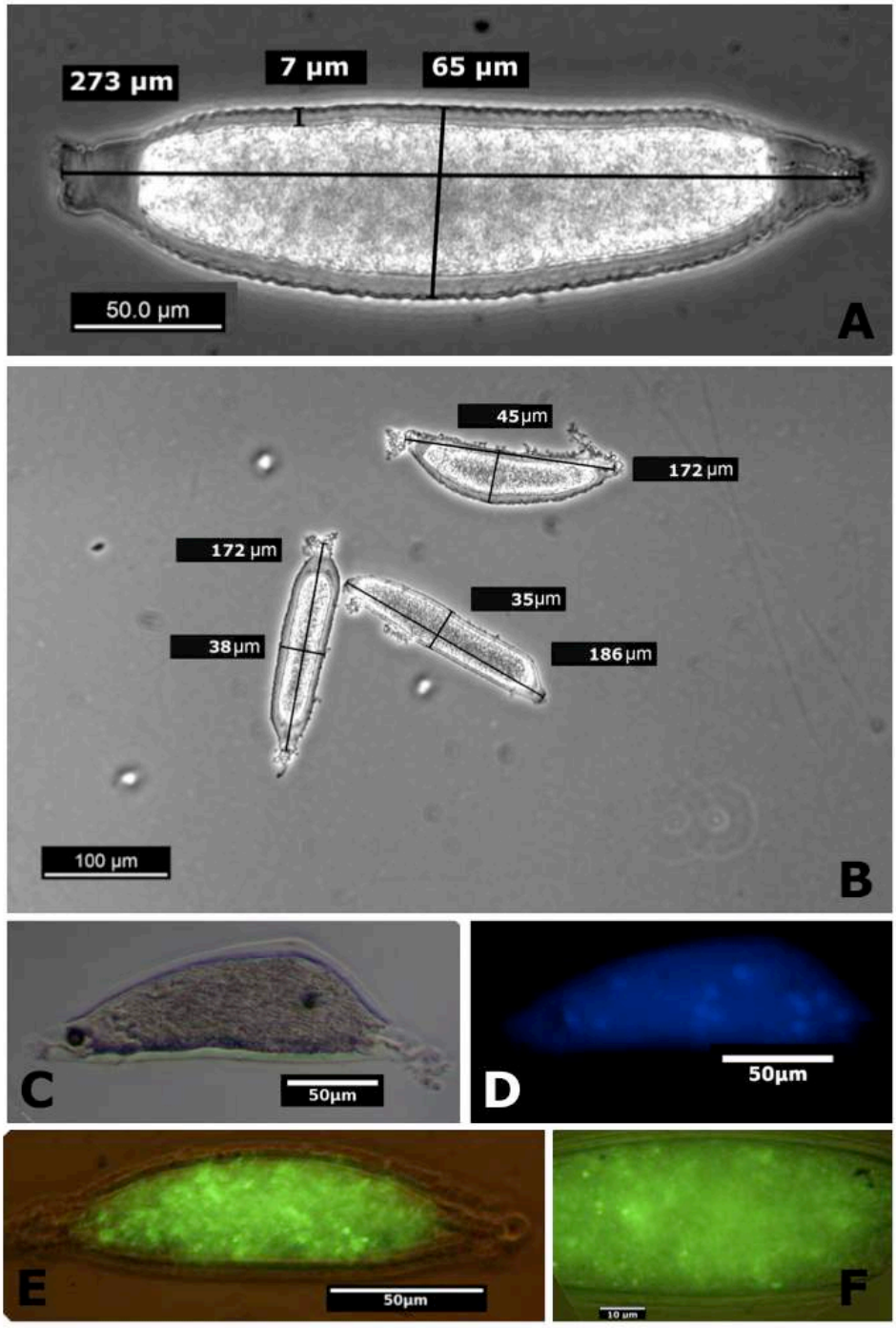


Figure 2.2 Fixed and stained specimens of *Arnoldiellina fluorescens* gen. and sp. nov. Fixed holotype (A) and paratypes (B) indicating their respective size. Specimen stained with DAPI (blue) viewed with differential interference contrast (C) or UV light excitation (D). With DAPI staining, the multiple nuclei show up as light-blue coloured rounded spots. Living specimen showing the green autofluorescence viewed with differential interference contrast (E) and UV light excitation (460–500 nm) (F).

#	Length [µm]	Width [µm]	Ratio Length/width	Remarks
1	273	65	4.2	Holotype: Fig 2A
2	172	44	3.9	Paratype: Fig 2B
3	172	37	4.6	Paratype: Fig 2B
4	186	35	5.3	Paratype: Fig 2B
5	167	54	3.1	
6	160	38	4.2	
7	173	35	4.9	
8	154	36	4.3	
9	152	34	4.5	
10	205	70	2.9	Fig 2(C,D)
11	166	47	3.5	Fig 2(E,F)
12	170	38	4.5	Fig 1A
13	300	68	4.4	Fig 1B
14	253	69	3.7	
15	244	63	3.9	
16	174	54	3.2	

Table 2.1 Measurements of 16 specimens of *Arnoldiellina fluorescens*.

Molecular characterization: A total of 13 single-cell DNA extracts were obtained. PCR amplification of the 18S rDNA fragment produced positive results for seven DNA extractions. Three sequences were obtained for the fragment s14F1-20r and four additional sequences were obtained for a shorter fragment (s14F1-s17). All obtained sequences were nearly identical. Only the longer fragments (s14F1-s20r) were used for the following analysis.

The three sequences of *A. fluorescens* were aligned to 57 sequences of monothalamous foraminifera selected from our database. We arbitrarily used

environmental clades (Pawlowski *et al.* 2011b) as an outgroup (Figure 2.3). Our analysis shows that the *Arnoldiellina* sequences branch within clade I (Pawlowski *et al.* 2002b), as sister group to *Pelosina* and *Astrammina*, but this relationship is weakly supported (59%). Higher bootstrap value (85%) was obtained for the whole clade I, including *Armorella*, *Saccodendron* and *Pelosinella* (Figure 2.3). Interestingly, an insertion of about 50 nucleotides characteristic for clade I is also present in *Arnoldiellina* (Video S1). Analysis of the secondary structure shows that this insertion forms a helix situated between helices 45 and 47, absent in other foraminiferans, except some lineages of clade C (Figure 2.3A).

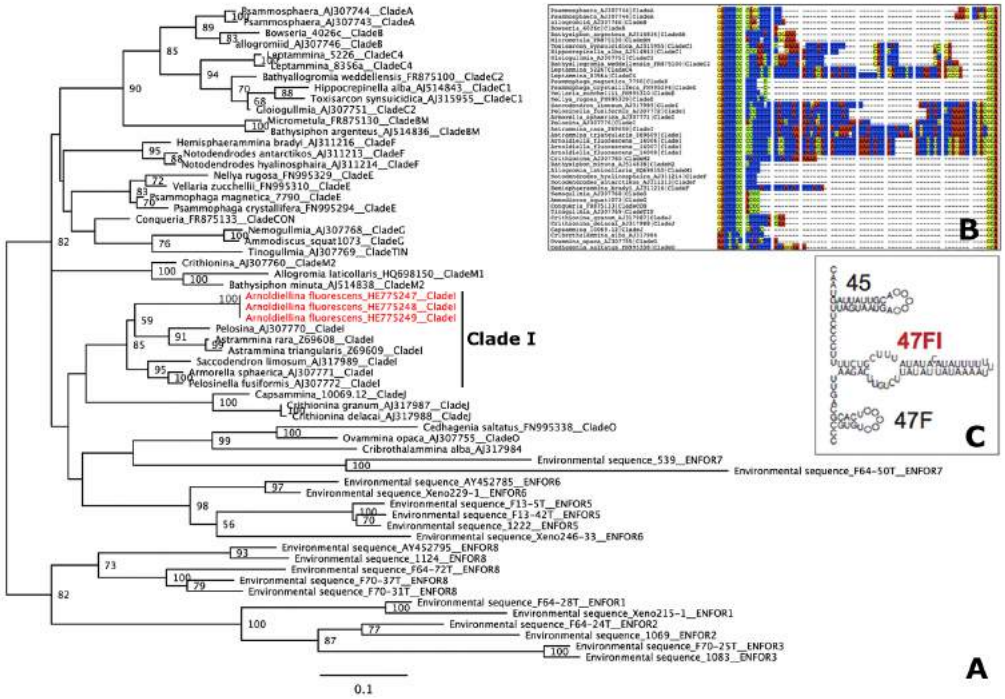


Figure 2.3 (A) Phylogenetic tree of monothalamous foraminifera based on partial 18S rDNA sequences, showing the position of *Arnoldiellina fluorescens* gen. and sp. nov. Support values are given as RaxML bootstrap; only values ≥50 are shown. (B) Alignment of the region of the 18S rDNA between the helix 45 and 47, showing the insertion specific to clade I. (C) Secondary structure of the insertion in *Arnoldiellina fluorescens*.

2.6. Discussion

Our study is the first report of GAF in foraminifera, but the phenomenon is relatively well known in protists. It seems to be a common feature in heterotrophic and autotrophic dinoflagellates (Chang & Carpenter 1991; Tang & Dobbs 2007), considered sometimes as a useful taxonomic character (Elbrächter 1994). Its presence in all life-history stages of the parasitic dinoflagellate *Amoebophrya* (Chambouvet *et al.* 2011) is commonly used to detect infection of phytoplankton (Coats & Bockstahler 1995; Park *et al.* 2004). The GAF was also found in diatoms, chlorophytes, raphidophytes, and other microalgae (Tang & Dobbs 2007). Among heterotrophic protists other than dinoflagellates, GAF was only observed in ciliates (Laval-Peuto & Rassoulzadegan 1988).

The case of *Arnoldiellina* confirms that the presence of GAF is not specifically linked to autotrophic activity. Although in many algae GAF is found in association with chloroplasts, its localisation is often very different, for example near the dinoflagellate stigma (Tang & Dobbs 2007) or in the flagellum of brown and golden algae (Coleman 1988). In *Arnoldiellina*, the GAF is evenly distributed throughout the cytoplasm, suggesting the presence of a fluorescent compound produced by the cell. The nature of this compound is unknown, but it might be similar to luciferase or the green fluorescent protein present in many organisms (Shimomura *et al.* 1962; Gould & Subramani 1988), or else the flavoprotein found in the posterium flagellum of brown algae (Fujita *et al.* 2005).

The evolutionary importance of GAF in foraminifera is questionable. *Arnoldiellina* is the first well documented case of a foraminiferan that emits green autofluorescence. However, this property might occur more often among foraminifera as assumed so far. Some unpublished observations suggest GAF activity in other foraminiferal species (Sam Bowser, Ivan Volsky, pers. commun.). In fact, until now very few foraminiferans have been examined using epifluorescence microscopy. A systematic use of this technique in foraminiferal research may reveal other cases of natural green autofluorescence in this group and possibly also in other protists.

CHAPTER 3

MOLECULAR PHYLOGENY AND MORPHOLOGY OF *LEANNIA VELOXIFERA* N. GEN. ET SP. UNVEILS A NEW LINEAGE OF MONOTHALAMOUS FORAMINIFERA

LAURE APOTHÉLOZ-PERRET-GENTIL & JAN PAWLOWSKI

Published in Journal of Eukaryotic Microbiology, **62**, 353–361, 2015

3.1. Project description

Together with Maria Holzmann and Emanuela Reo, we went sampling in the Red Sea. One of the purposes of this trip was to collect a great amount of *Amphisorus* species to try to keep them in culture in the lab. Those foraminifera are quite big and can be found abundantly on seagrass leaves, wherefrom they can be collected by hand without the use of a binocular. Back to the lab, while observing small juvenile *Amphisorus* I found, by chance, a new allogromiid. It was very small and had no particular morphological features that could assign it to any previously described species. Its sequences confirmed that it is a new species and even a new genus. Unfortunately, likewise *Arnoldiellina*, the lifetime of Leannia was very short, and its maintenance in culture was not possible.

3.2. Abstract

Monothalamous (single-chambered) foraminifera have long been considered as the "poor cousins" of multichambered species, which calcareous and agglutinated tests dominate in the fossil record. This view is currently changing with environmental DNA surveys showing that the monothalamids may be as diverse as hard-shelled foraminifera. Yet, the majority of numerous molecular lineages revealed by eDNA studies remain anonymous. Here, we describe a new monothalamous species and genus isolated from the sample of sea grass collected in Gulf of Eilat (Red Sea). This new species, named *Leannia veloxifera*, is characterized by a tiny ovoid test (about 50–100 μm) composed of thin organic wall, with two opposite apertures. The examined individuals are multinucleated and show very active reticulopodial movement. Phylogenetic analyses of SSU rDNA, actin, and beta-tubulin (ß-tubulin) show that the species represents a novel lineage branching separately from other monothalamous foraminifera. Interestingly, the SSU rDNA sequence of the new species is very similar to an environmental foraminiferal sequence from Bahamas, suggesting that the novel lineage may represent a group of shallow-water tropical allogromiids, poorly studied until now.

3.3. Introduction

Recent development of high-throughput sequencing technology tremendously speeds up the process of the discovery of new environmental lineages of protists. Several high-rank taxonomy groups composed mainly of environmental sequences have been proposed, such as MAST 1-11 (Logares *et al.* 2012; Massana *et al.* 2014). Some of these groups could not be assigned to any supergroup and have no morphologically characterized representatives, e.g. Rappemonads (Kim *et al.* 2011). The interpretation of others has changed after a microscopic examination of cultivated isolates, e.g. Picozoa, formely Picobiliphytes (Seenivasan *et al.* 2013). The integrated taxonomy of protists based on morphological and molecular study appears as a necessity (Moreira & López-García 2014). Indeed, few studies combining the single DNA-barcoding with morphological and ultrastructural data have been very successful in identifying the enigmatic environmental lineages (Rueckert *et al.* 2011). However, such studies are time-consuming and require a good taxonomic expertise.

Therefore, they are rare and can hardly fill the taxonomic gap in some poorly known groups such as monothalamous foraminifera.

Monothalamids are a heterogeneous assemblage of diverse foraminiferal lineages characterized by organic-walled or agglutinated single-chambered tests, called allogromiids or astrorhizids, respectively (Pawlowski *et al.* 2002b). Because their tests are poorly preserved in dried samples routinely studied by foram specialists, the diversity of monothalamids has never been extensively examined. It is well known that the group dominates in some marine habitats, especially in the deep-sea and high-latitude regions (Gooday 2002; Gooday *et al.* 2005), but they are also common in warm water environments (Habura *et al.* 2008) and in freshwater (Holzmann *et al.* 2003; Dellinger *et al.* 2014). Many new monothalamous species have been described in the last decade (Gooday & Pawlowski 2004; Sabbatini *et al.* 2004; Gooday *et al.* 2004, 2010; Altin *et al.* 2009; Pawlowski & Majewski 2011; Apothéloz-Perret-Gentil *et al.* 2013 - Chapter 2; Voltski *et al.* 2014). Yet, as suggested by large number of undetermined species found in monothalamids diversity surveys (Majewski 2005; Gooday *et al.* 2005; Majewski *et al.* 2007), our knowledge of the group is still very fragmentary.

The immense diversity of monothalamids was confirmed by environmental DNA (eDNA) studies. The sequences assigned to monothalamous lineages dominate in all eDNA surveys of foraminiferal communities, both those that used clonal approach (Habura *et al.* 2004, 2008; Pawlowski *et al.* 2011b; Tsuchiya *et al.* 2013; Bernhard *et al.* 2013) and those using next-generation sequencing technology (Lecroq *et al.* 2011; Pawlowski *et al.* 2011b; Lejzerowicz *et al.* 2013; Pawlowski *et al.* 2014a). In some deep-sea samples, the proportion of monothalamids reaches up to 74% and may be even higher if we consider that most of unassigned OTUs also belong to this group (Lecroq *et al.* 2011). Most of the monothalamous sequences retrieved from deep-sea samples group within eight large clades defined as ENFOR 1-8 (Pawlowski *et al.* 2011a), but many represent independent lineages comprising usually one or few sequences. Remarkably, none of these environmental lineages comprises morphologically described species, what makes them even more enigmatic.

To know more about monothalamid diversity, we started to systematically examine the morphology and obtain genetic data for all monothalamous species that appeared in our samples. Previously, we described a new fluorescent allogromiid

from Gulf of Eilat (Apothéloz-Perret-Gentil *et al.* 2013 - Chapter 2). Here, we report another new species from the same locality. Phylogenetic study of this species shows that it represents a novel lineage of monothalamids, which also comprises the environmental sequence of an uncultured foraminifer from Bahamas.

3.4. Materials and methods

3.4.1. Isolation

Samples of the Halophila leaves were collected by SCUBA diving at 15 m in front of the Interuniversity Institute for Marine Sciences (IUI), in Eilat, Israel, on December 2012. The coordinates of the sampling spot are: 29.51482 N 34.92674 E. Large benthic foraminifera of the genus *Amphisorus* were detached by hand from the sea grass and transferred to Petri dishes filled with filtered seawater to which few drops of Erdschreiber medium (5% soil extract, 1 mM $NaNO_3$, 0.07 mM Na_2HPO_4, 10 mM Tris pH = 8, filled up with sterile seawater) were added. The specimens of small-sized allogromiid foraminifera that are described in this paper appeared in the culture dishes several days after placing *Amphisorus* there. They flourish in culture dishes for few weeks, probably in result of asexual reproduction of few individuals, and then rapidly disappeared.

3.4.2. Morphology and cytology

Three living specimens were incubated 5 min at ambient temperature using 4’,6’-diamidino-2-phenylindole (DAPI) at 5.10E-4 mg/ml to stain and identify nuclei. The procedure was carried out in a dark room. Five specimens were fixed in a 10% solution of formalin. Living and fixed specimens were observed with an inverted microscope (Nikon Eclipse Ti, Nikon Instruments Europe, Amsterdam, Netherlands), a fluorescence microscope (Nikon Eclipse E200) and a stereoscopic one (Leica M205C, Leica, Hamburg, Germany). Photographs were taken with a Leica DFC 420C, an Imaging Source DFK 41AF02 camera, and a Leica DFC 450C, respectively. Movies were made on the fluorescent microscope and the inverted microscope with the same camera. They are available at: http://forambarcoding.unige.ch/movies

3.4.3. DNA/RNA extraction, amplification, cloning, and sequencing

DNA from three specimens was extracted in guanidine lysis buffer (Pawlowski 2000), each extraction was performed with a single specimen. RNA extraction was performed with seven specimens using the NucleoSpin RNA XS kit (Macherey-Nagel, Düren, Germany). Afterwards cDNA was synthetised using the iScript Select cDNA synthesis Kit (BioRad, Hercules, CA) with random primers.

PCR amplifications of the complete SSU rDNA were performed in three steps. The first fragment was amplified using the primer pair s14F3 (5'ACG CA(AC) GTG TGA AAC TTG) and B (5'TGA TCC TTC TGC AGG TTC ACC TAC). PCR products were re-amplified using the nested primer s14F1 (5'AAG GGC ACC ACA AGA ACG C). The second fragment was amplified using the primer pair 6F (5'CCG CGG TAA TAC CAG CTC) and 17 (5'CGG TCA CGT TCG TTG C). PCR products were re-amplified using the nested primer 15A (5'CTA AGA ACG GCC ATG CAC CAC C). The third fragment was amplified using the primer pair A10 (5'CTC AAA GAT TAA GCC ATG CAA GTG G) and 12R (5'G(GT)T AGT CTT (AG)(AC)(ACT) AGG GTC A). PCR products were re-amplified using the nested primer 7R (5'CTG (AG)TT TGT TCA CAG T(AG)T TG). The sequenced fragments have been assembled to retrieve the complete SSU rDNA.

PCR amplifications of a fragment of the actine gene were performed using the primer pair ActN2 (5'ACC TGG GA(CT) GA(CT) ATG GA) and 1354R (5'GGA CCA GAT TCA TCA TA(CT) TC). PCR products were re-amplified using the nested primer ActF1 (5'CNG A(AG)G C(AGT)C CAT T(AG)A A(CT)C), as described in Flakowski *et al.* (2005).

PCR amplifications of a fragment of the ß-tubulin gene were performed using the primer pair BtubF1 (5'CAA TGT GGT AAC CAA ATT GC) and BtubR1 (5'CAT CTT GTT TGT CTT GAT ATT CAG T). PCR products were re-amplified using the nested primer BtubF2 (5'AAT TGG GCA AAA GGA CAT TA), as described in Habura *et al.* (2005).

The amplified PCR products were purified using High Pure PCR Purification Kit (Roche Diagnostics, Hoffmann- La Roche AG, Basel, Switzerland) and cloned with the TOPO10 kit from Invitrogen (Thermo Fisher Scientific, Waltham, MA). Between two and four clones were sequenced per PCR. Sequencing reactions were performed using the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Thermo Fisher Scientific) and analysed on a 3130XL Genetic Analyser (Applied Biosystems). The five new sequences reported in this paper were deposited in the EMBL/GenBank database (LM994876–LM994880).

3.4.4. Sequence alignments and phylogenetic analysis

The gene coding sequences were translated into amino acid sequences using Seaview vs 4.3.3. software (Gouy *et al.* 2010). All the sequences were aligned using the same program.

The SSU rDNA sequences were aligned to 28 foraminiferan sequences and 1,943 sites of the alignment were used for the analysis using GTR+G+I model. For the short fragment, 37 environmental sequences were added to 25 foraminferan ones and the whole alignment of 2,016 sites was used with GTR+G+I as model of evolution. Actin sequences were aligned to 35 sequences of Retaria (27 foraminiferans and 8 radiolarians used as outgroup) and 274 sites were used for the analysis using WAG+G model. ß-tubulin sequences were aligned to 28 sequences of Retaria (21 foraminiferans and 7 radiolarians used as outgroup) and 262 sites were used for the analysis using the WAG+G model. In addition, we performed a concatenated analysis of the actin and ß-tubulin genes with 35 sequences of Retaria; for 19 species no ß-tubulin gene data were available.

Best models for all analyses were calculated using Mega5 (Tamura *et al.* 2011). Phylogenetic trees were constructed using maximum likelihood program RAxML Black-Box (Stamatakis *et al.* 2008). In addition, Bayesian analyses were performed for all gene trees using MrBayes 3.2.1 (Ronquist & Huelsenbeck 2003) with four chains running in parallel for 10,000,000 generations. For each analysis, a burnin of 20% was carried out to construct the best tree and calculate posterior probabilities.

3.5. Results

3.5.1. Morphologic description

The new species is a monothalamid without test. Specimens present an ovoid shape (ratio length/width between 1 and 2) between 72 and 113 µm in length and 41 and 84 µm in width. The measurement of each specimens observed is recorded in Table 1. Their organic wall is transparent and measure from 1 to 3 µm in width. They possess two opposite apertures, funnel-shaped with a tubular internal extension. Cytoplasm is multinucleate (Figure 3.1E) and granular, with rapid movement (Movie S1). However, the multinucleate nature may represent only a stage of the life cycle. Reticulopodes are very active. They rapidly form large reticulopodial network and fast moving granules inside (Movie S2).

Seventeen additional specimens were used either for DNA or RNA extraction and subsequent amplification, fixed in formalin or observed and photographed alive. Description of the used specimens is summarised in Table 3.1.

	Specimens	Length [µm]	Width µm]	Ratio length/width	Figure	Sequences		
						18S	ß-tubulin	Actin
1	DNA (17004)	100	54	1.9		LM994876	LM994880	LM994879
2	DNA (17005)	108	64	1.7		LM994877		
3	DNA (17006)	88	60	1.5	Fig. 3.1(C)	LM994878		
4	Holotype	111	64	1.7	Fig. 3.1(A,E)			
5	Paratype	100	76	1.3	Fig. 3.1(B)			
6	Paratype	106	77	1.4	Fig. 3.1(B)			
7	Paratype	94	63	1.5	Fig. 3.1(B)			
8	Paratype	102	62	1.7	Fig. 3.1(B)			
9	Paratype	95	64	1.5	Fig. 3.1(B)			
10	DAPI	74	69	1.1				
11	DAPI	83	62	1.3				
12	Formaline	87	84	1.0				
13	Formaline	82	79	1.0				
14	Formaline	83	46	1.8				
15	Formaline	66	46	1.4	Fig. 3.1(F)			
16	Formaline	68	41	1.7				
17	RNA	106	70	1.5				
18	RNA	106	67	1.6				
19	RNA	103	57	1.8				
20	RNA	72	60	1.2				
21	RNA	102	61	1.7				
22	RNA	95	61	1.6				
23	RNA	113	62	1.8	Fig. 3.1(D)			

Table 3.1 Description of 23 specimens of *Leannia veloxifera* n. gen. et sp.

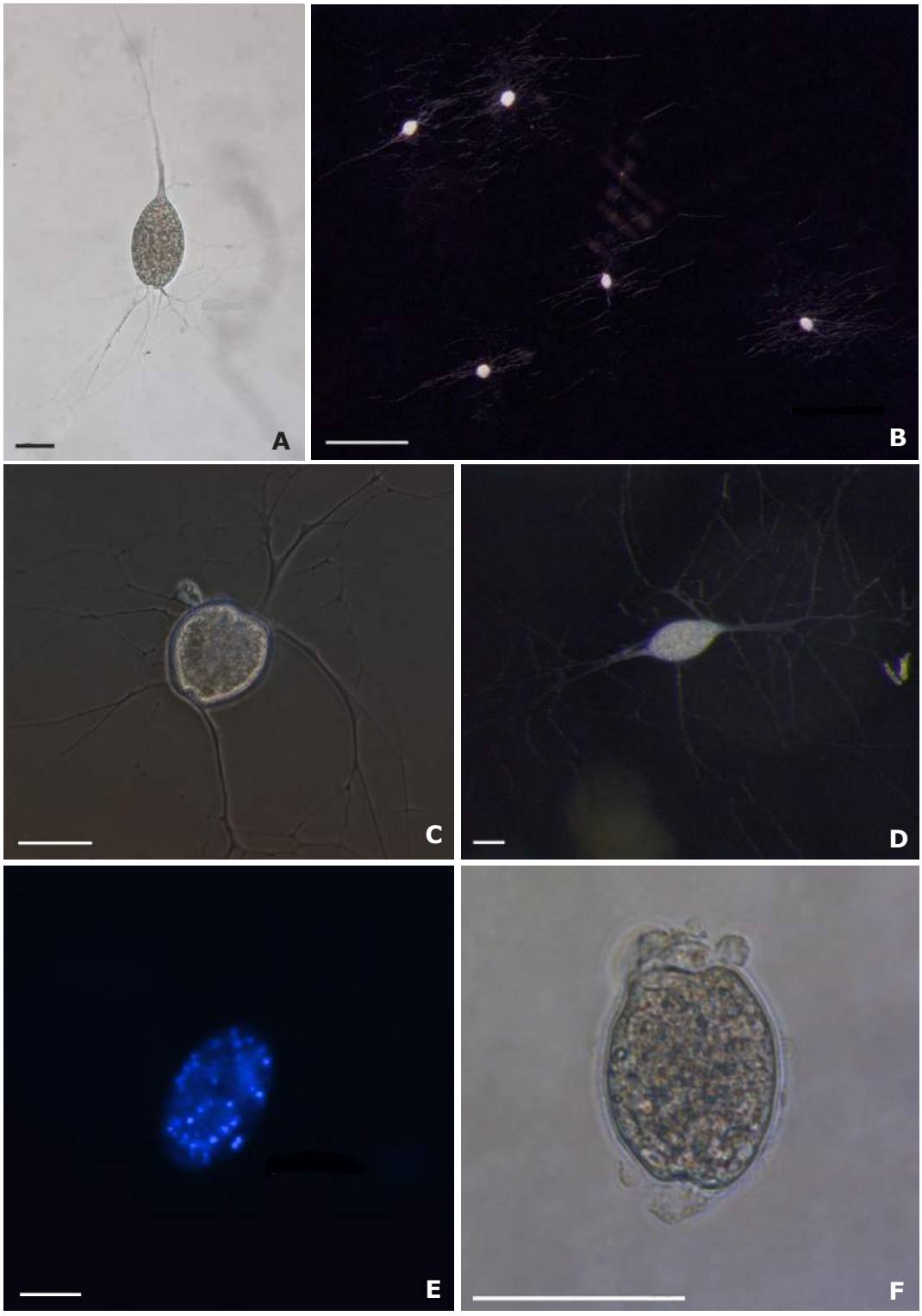


Figure 3.1 Specimens of *Leannia veloxifera* n. gen. et sp. A. Living holotype. B. Paratypes with their expensive granuloreticulopodia's web. C, D. Two living specimens. E. Holotype stained with DAPI (blue) viewed with UV light excitation (460–500 nm). F. Fixed specimen. Scale bar (A, C–F) correspond to 50 µm and scale bar (B) correspond to 500 µm.

3.5.2. Molecular phylogeny (SSU rDNA, actin, ß-tubulin)

To investigate the phylogenetic position of the new species, we performed an analysis of complete SSU rRNA gene sequence (total length 3,033 bp, GC content 32%). Three sequences were aligned to 25 sequences of foraminifera from our database and phylogenetic trees were built using ML and BI methods (Figure 3.2). The tree is rooted at the clade I according to the ß-tubulin phylogeny (FigureS 3.2) and Hou *et al.* (2013)Hou et al. (2013). The new allogromiid sequences form a very long branch (reduced 50% in Figure 3.2) not related to any of the previously described monothalamous clades (Pawlowski *et al.* 2002b). Its position at the base of a clade formed by eight globothalamean species and few monothalamids belonging to clades A, BM, and C is relatively well supported (0.95 PP, 74% BV). Relationships between other monothalamid clades, including *Capsammina patelliformis*, *Allogromia sp*., *Nemogullmia sp*., the clade E and the freshwater foraminifer *Reticulomyxa filosa* are not resolved. The topology of the ML tree differs from the BI tree in the position of C. *patelliformis*, which branches at the base of the tree.

To further refine the phylogenetic position we analysed actin and ß-tubulin genes. In the actin tree (FigureS 3.1), its branch is very long compared to other foraminiferans. The new species groups in the unresolved clade formed by six tubothalameans, *Bathysiphon flexilis* and *R. filosa*. This clade is sister to monothalamous clade M, composed of *Allogromia*, *Edaphoallogromia* and *Bathysiphon sp*. Both clades form a sister group to Globothalamea, which are well supported in Bayesian analyses (1 PP) but not in ML analysis (53% BV). The topology of the ML tree differs slightly, with monothalameous clade M branching at the base of Globothalamea and the clade, to which belongs the new allogromiid species. However, this topology is not supported (less than 25% BV).

In the ß-tubulin tree (FigureS 3.2), the amino acid sequence of the new allogromiid branches as sister to Globothalamea. This relation is strongly supported in Bayesian analysis but not in ML analysis. The topology of foraminiferal tree is characterized by strong support for Globothalamea (1 PP, 93% BV), and paraphyly of monothalamids and tubothalameans. A monothalamid *Astrammina rara* branches at the base of the tree, followed by a clade of *Allogromia* and *Crithionina delacai*. However, none of these branching patterns is strongly supported.

A final analysis was carried out by using the concatenated ß-tubulin and actin genes (Figure 3.3) with radiolarians as outgroup. Within foraminifera, Globothalamea form a distinct group (1 PP, 85% BV) with the new species branching at their base (0.95 PP, 46% BV). The other monothalamids form unsupported branches with Tubothalamea branching within them (0.71 PP, 43% BV). The monophyly of foraminifera is relatively well supported (1 PP, 80% BV).

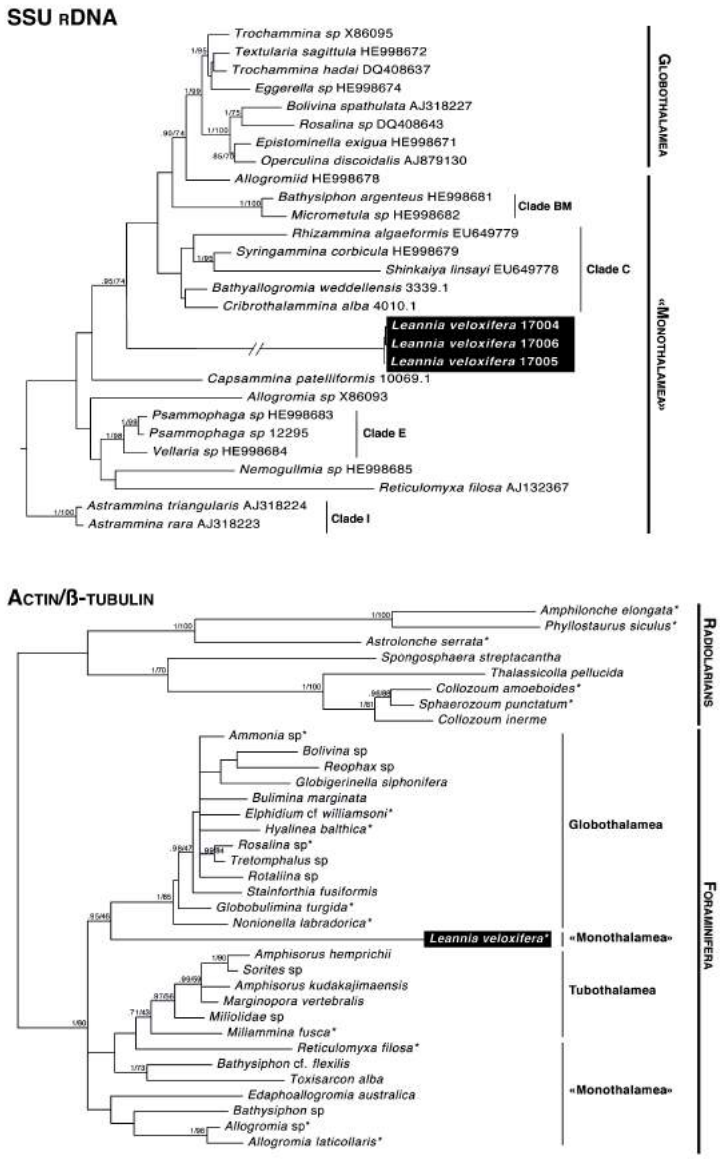


Figure 3.2 Phylogenetic tree of 34 sequences of foraminifera based on complete SSU rDNA sequences, showing the position of *Leannia veloxifera* n. gen. et sp. in a black frame. Support

values are given as MrBayes posterior probabilities/RaxML bootstrap; only values superior or equal to 0.85 for posterior probabilities and 70 for bootstrap values are shown. Concatenated phylogeny of actin and ß-tubulin genes. The analysis was done with 27 sequences of foraminifera sequences with eight sequences of radiolarian used as outgroup. Both genes were retrieved for species with a asterisk (*), ß-tubulin gene are missing for the other. Support values are given as MrBayes posterior probabilities/RaxML bootstrap; only values superior or equal to 0.50 for posterior probabilities and 40 for bootstrap values are shown. The position of *Leannia veloxifera* n. gen. et sp. is shown in a black frame.

In addition to phylogenetic analyses of complete SSU, actin, and ß-tubulin sequences, we also analysed a short fragment of the SSU rDNA, commonly used as foraminiferal barcode (Pawlowski & Holzmann 2014), and for which many environmental sequences are available. In Figure 3.3, we present a tree with 62 selected sequences representing previously described environmental clades (ENFOR), unique environmental lineages (ENV), undetermined monothalamous morphotypes (UNDET) and identified morphospecies. The new allogromiid species did not branch with any of the previously described environmental clades (ENFOR 1-9, Pawlowski *et al.* 2011a). However, it branches with the unique environmental sequence of “uncultured foraminifera” from the Highborne Cay in Bahamas (Bernhard *et al.* 2013). Both sequences differ by only 8% and their relation is highly supported (1 PP, 100% BV). A sequence of another uncultured foraminifera from Sippewissett marshes in Massachusetts (Habura *et al.* 2008) branches at the base of this group.

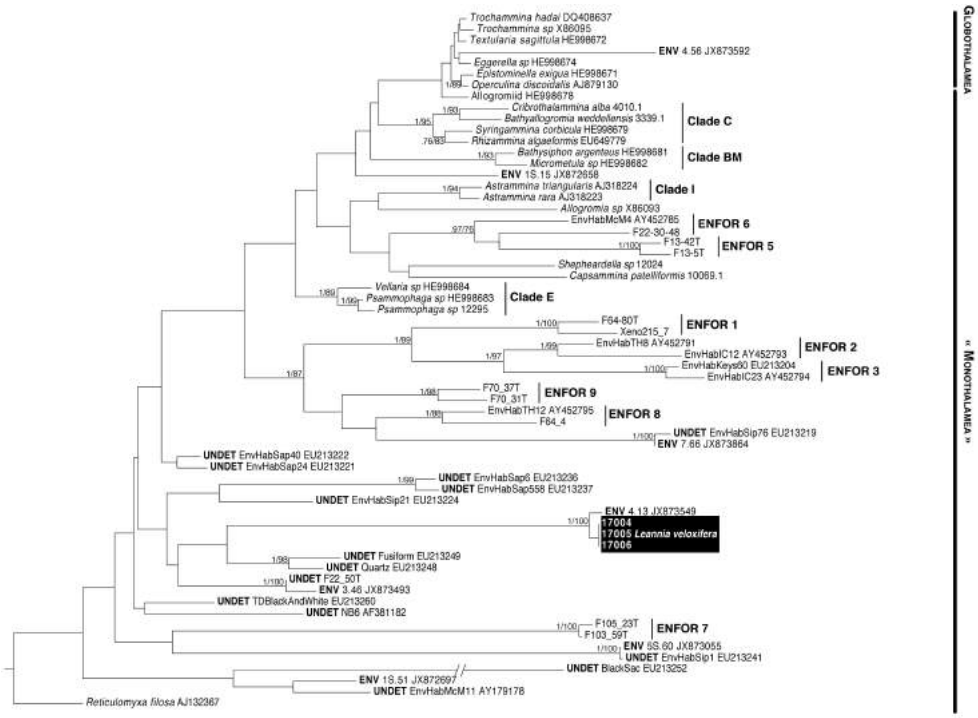


Figure 3.3 Phylogenetic tree of 62 sequences of foraminifera including 37 environmental sequences. *Leannia veloxifera* n. gen. et sp. sequences are framed in black. Support values are given as MrBayes posterior probabilities/RaxML bootstrap; only values superior or equal to 0.75 for posterior probabilities and 75 for bootstrap values are shown.

We also looked for environmental sequences in the large dataset of environmental sequences provided by next-generation sequencing. We found two sequences, both from Marlborough Sounds in New Zealand, which are related to the new clade. One is exactly identical to Leannia sequences (100% identity and 100% coverage) and the other is close (98% identity and 90% coverage) to the sequence from Highborne Cay in Bahamas. However, those sequences are very short (53 and 57bp respectively) and correspond only to one hypervariable region (37F) of the SSU rDNA. Therefore, we did not add them to the tree on Figure 3.3.

3.6. Discussion

The species described here is the second new allogromiid, after *Arnoldiellina fluorescens* (Apothéloz-Perret-Gentil *et al.* 2013 - Chapter 2), reported from the same locality in Gulf of Eilat during the last 3 years. This may sounds surprising given an extensive foraminiferal research that has been conducted in this area over the years and which conducted to an impressive number of publications about Gulf of Eilat foraminifera (reviewed in Reiss & Hottinger 1984; Lee & Anderson 1991; Hottinger *et al.* 1993). However, all these classical work focused on large benthic foraminifera and does not care about the small-sized species. One of us (JP) showed many years ago that the poorly known community of calcareous microforaminifera flourish on the Halophila leaves and coral rubble in the Gulf of Eilat (Pawlowski & Lee 1991, 1992). At that time, however, our attention was focused on tiny calcareous species, which could be identified either directly on dried leaves or in the fine fraction of sediment samples. Two new genera and eight new species of microforaminifera belonging to the families Glabratellidae and Rotaliellidae have been described (Pawlowski & Lee 1991, 1992).

Compared to this work on hard-shelled foraminifera, the isolation and description of new allogromiid species is much more challenging. The organic-walled foraminifera are not preserved in dried samples and can be isolated only from laboratory cultures or formalin-fixed samples. The cultivation approach traditionally used in protistology is seldom applied to foraminiferal species, because they are difficult to maintain in laboratory cultures and their description has to be done rapidly after they have been observed. The allogromiids usually flourish in culture dishes for few weeks, probably in result of asexual reproduction of one or two individuals, and then rapidly disappeared. Only few species adapt to culture conditions and can be maintained for longer periods of time, like for example, *Allogromia laticollaris* or other species of this genus (McEnery & Lee 1976; Parfrey & Katz 2010).

Despite these difficulties, our study shows that the cultivation, even for short periods of time, is essential for taxonomic study of this group. Hundreds of novel lineages have been revealed by eDNA and RNA studies (Pawlowski *et al.* 2011a; Tsuchiya *et al.* 2013; Bernhard *et al.* 2013; reviewed in Pawlowski *et al.* 2014b), but most of them remained microscopically undocumented. The fact that *Leannia veloxifera* branches with one of these enigmatic lineages confirms that at least some of them can be

assigned to tiny allogromiids, which possibly form a rich community in shallow tropical waters. Their inconspicuous presence may also explain the immense diversity of environmental lineages observed at the deep-sea bottom (Pawlowski *et al.* 2011a; Lecroq *et al.* 2011; Lejzerowicz *et al.* 2013). Many of these undetermined sequences have been amplified from samples of xenophyphoreans or other large deep-sea benthic foraminifera, which tests could provide a suitable habitat for tiny allogromiids (Lecroq *et al.* 2009b). More extensive cultivation efforts coupled with a detailed microscopic study could lift the veil on these mysterious "eDNA" foraminiferans.

3.7. Taxonomic summary

Supergroup RHIZARIA Cavalier-Smith, 2002 Phylum FORAMINIFERA D'Orbigny, 1826 Class "Monothalamea" Pawlowski et al. 2003

Leannia n. gen. Apotheloz-Perret-Gentil et Pawlowski 2014

Description. Test free, monothalamous, ovoid shape (ratio length/width between 1 and 2), < 115 µm in length and < 85 µm in width; organic wall transparent from 1 to 3 µm in width. Two opposite apertures, funnel-shaped with a tubular internal extension. Cytoplasm multinucleate (Figure 3.1E) at least in this stage of its life cycle; granular,

with rapid movement (Movie S1). Reticulopodes very active with rapidly forming large reticulopodial network and fast moving granules (Movie S2). Type species. *Leannia veloxifera* n. sp. Apotheloz-Perret-Gentil et Pawlowski 2014

Etymology. The genus was named in honour of first author's daughter.

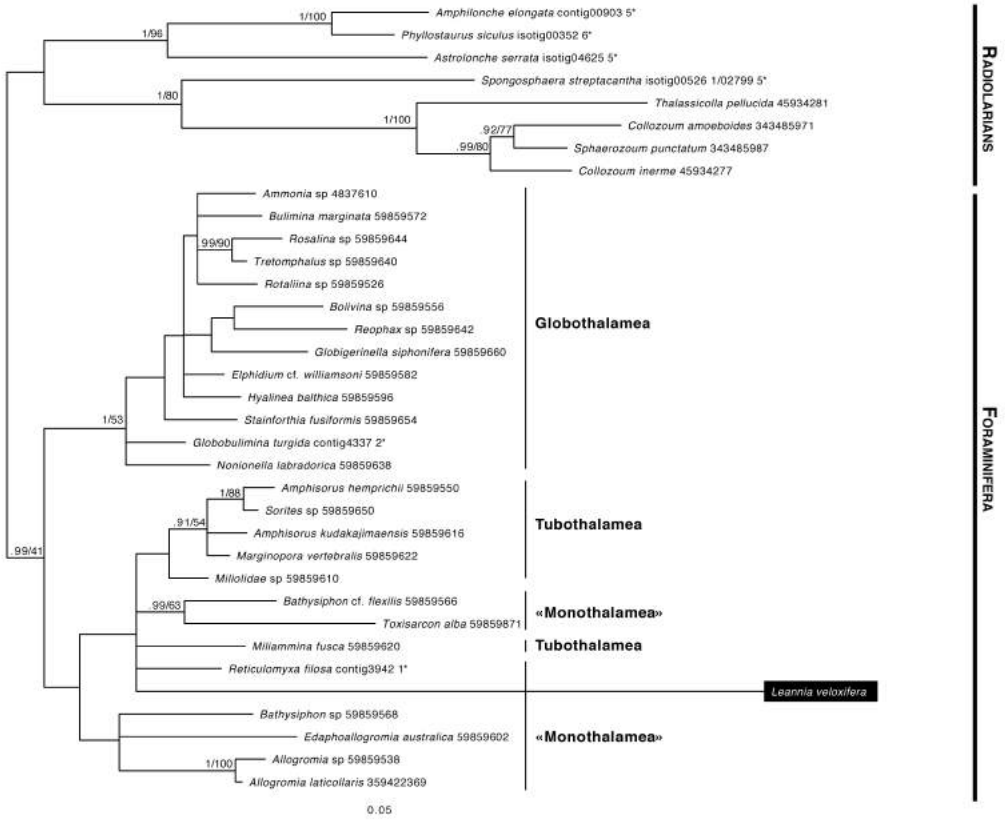
Leannia veloxifera n. sp. Apotheloz-Perret-Gentil et Pawlowski 2014

Description. Same as for genus. DNA/Amino acids sequences. SSU rDNA sequences, Actin and ß-tubulin proteins (GenBank LM994876– LM994880) Type locality. Gulf of Eilat, Red Sea (Halophila sea grass meadow in front of the IUI, Eilat, Israel). Type habitat. Marine Type material. A specimen preserved in formalin was selected as holotype (MHNG INVE 89252) and deposited at the Museum of Natural History in Geneva (MHNG) together with five paratypes (MHNG INVE 89253).

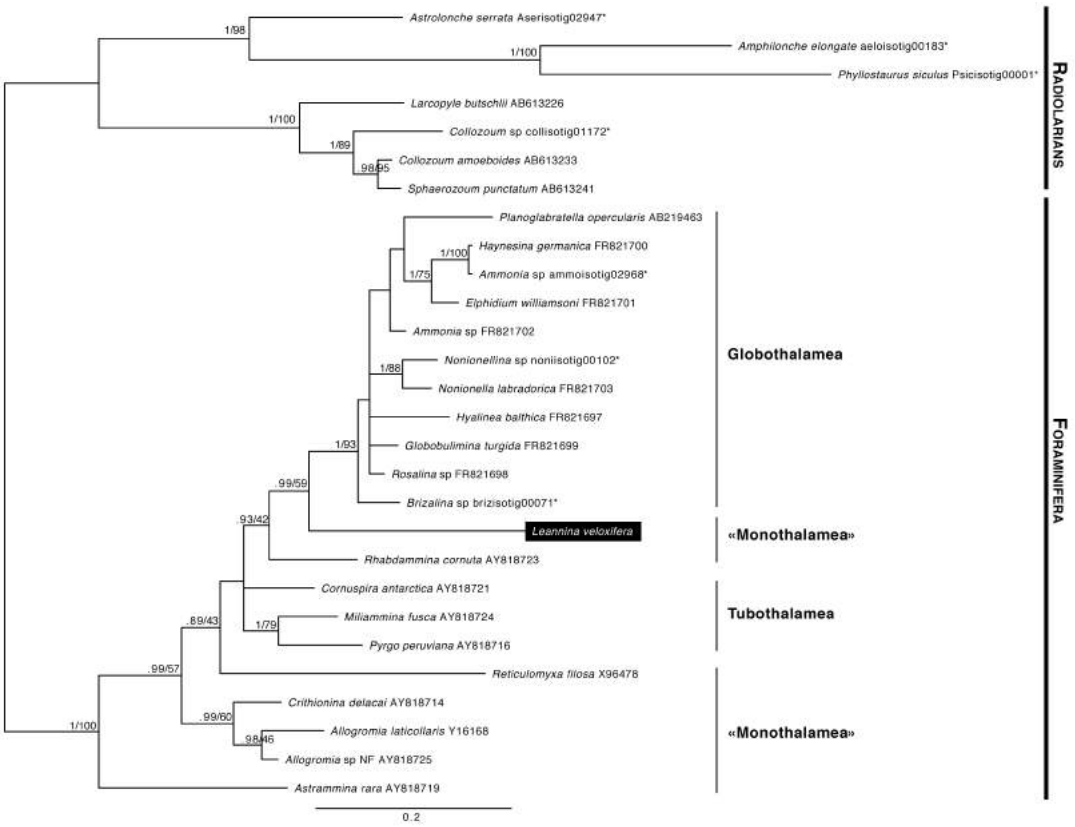
Etymology. The species was named for the extreme rapidity to form its granuloreticulopodial network. Remarks. *Leannia veloxifera* is morphologically similar to *Arnoldiellina fluorescens*, another allogromiid described from the Gulf of Eilat (Apothéloz-Perret-Gentil *et al.* 2013 - Chapter 2). However, *Leannia* had two apertures, while *Arnoldiellina* possesses only one. Moreover, the later species shows green autofluorescence when observed under UV light.

3.8. Supplementary materials

FigureS 3.1 Actin gene phylogeny of 27 sequences of foraminifera with eight sequences of radiolarian used as outgroup. Support values are given as MrBayes posterior probabilities/RaxML bootstrap; only values superior or equal to 0.50 for posterior probabilities and 40 for bootstrap values are shown. The position of *Leannia veloxifera* n. gen. et sp. is shown in a black frame.



FigureS 3.2 ß-tubulin gene phylogeny of 21 sequences of foraminifera with seven sequences of radiolarian are used as outgroup. Support values are given as MrBayes posterior probabilities/RaxML bootstrap; only values superior or equal to 0.50 for posterior probabilities and 40 for boot- strap values are shown. The position of *Leannia veloxifera* n. gen. et sp. is shown in a black frame.



CHAPTER 4

TAXONOMIC REVISION OF FRESHWATER FORAMINIFERA WITH THE DESCRIPTION OF TWO NEW AGGLUTINATED SPECIES AND GENERA

FERRY SIEMENSMA, LAURE APOTHÉLOZ-PERRET-GENTIL, MARIA HOLZMANN, STEFFEN CLAUSS, ECKHARD VÖLCKER & JAN PAWLOWSKI

Published in European Journal of Protistology, **60**, 28-44, 2017

4.1. Project description

For a long time, Maria Holzmann and myself, looked at freshwater sediments in the hope of finding living foraminifera. However, our attempts have always been vain and therefore we have been very excited when Ferry Simensma and Eckard Völcker contacted Maria because they have found an interesting freshwater foraminiferan looking like species and they wanted to sequence it. These observations were the results of special attention and regular screening of the freshwater sediment put into cultivation. Ferry and Eckard performed all the morphological work and isolated the specimens. Maria barcoded them and I performed the molecular analyses, including the data from my freshwater project. We decided to not include all the environmental data in this paper because we wanted to publish them separately.

4.2. Abstract

Most foraminifera inhabit marine habitats, but some species of monothalamiids have been described from freshwater environments, mainly from Swiss water bodies and over 100 years ago. Recent environmental DNA surveys revealed the presence of four major phylogenetic clades of freshwater foraminifera. However, until now only one of them (clade 2) has been associated to morphologically described taxon – the family Reticulomyxidae. Here, we present morphological and molecular data for the genera representing the three remaining clades. We describe two new agglutinated freshwater genera from China and the Netherlands, *Lacogromia* and *Limnogromia*, which represent clades 3 and 4, respectively. We also report the first ribosomal DNA sequences of the genus *Lieberkuehnia*, which placed this genus within clade 1. Our study provides the first morphotaxonomic documentation of molecular clades of freshwater foraminifera, showing that the environmental DNA sequences correspond to the agglutinated monothalamous species, morphologically similar to those described 100 years ago.

4.3. Introduction

Foraminifera are unicellular eukaryotes characterized by the presence of granuloreticulopodia and the possession of a membranous, agglutinated, or calcareous test, which is either monothalamous (single-chambered) or polythalamous (multi-chambered) (Loeblich & Tappan 1988). Within monothalamids some species like *Reticulomyxa filosa* are amoeboid naked forms. Until 1859, the foraminifera were only known from marine habitats, but that year Claparède & Lachmann described a monothalamid foraminifer, *Lieberkuehnia wageneri*, sampled from an unknown water body in Berlin. It had a smooth flexible test with an entosolenian tube that separated the main cytoplasm mass from the pseudopodial peduncle.

In 1886 Henri Blanc , a Swiss scientist, described another freshwater foraminifer, *Gromia brunneri,* which he had collected from the bottom of Lake Geneva. This single-chambered species had an agglutinated test, an organic layer covered and/or embedded with foreign, mainly non-organic, particles. In subsequent years, Eugène Penard, another Swiss protozoologist, described four similar species *Gromia gemma*

and *G. squamosa* (1899), *G. linearis* (1902) and *G. saxicola* (1905) from the same lake. He also described *G. nigricans* (1902), which he found not far from Lake Geneva in Mateg- nin and a marsh near Rouelbeau. Penard made permanent preparations of these foraminifera, which are still preserved and available in the Penard Collection of the Natural History Museum of Geneva.

In 1904, Ludwig Rhumbler erected the subfamily Allogromiinae for monothalamous foraminifera characterized by a more or less flexible organic test wall commonly with one or rarely two terminal apertures at either end of the test. He included all described freshwater species in this taxon. In a recent higher ranked classification of foraminifera based on molecular phylogenies (Pawlowski *et al.* 2013), monothalamous foraminifera were considered as a paraphyletic group that contains agglutinated and organic walled species as well as "naked" amoeboid species and environmental clades with unknown morphological affinities.

Traditionally the organic-walled foraminifera are called allogromiids. Most of them are distributed over a wide range of marine and brackish habitats (Gooday 2002). Freshwater allogromiids with an agglutinated test were originally placed in genus *Gromia* by their discoverers, but as its type species *G. oviformis* is a filose marine species, Rhumbler (1904) transferred three species (*G. squamosa*, *G. nigricans* and *G. linearis*) to *Rhynchogromia* Rhumbler 1894. He further erected a new genus, *Diplogromia,* for the other two species having a double test wall: *G. brunneri* and *G. gemma,* however without designing a type species for the genus (Table 4.1).

Rhumbler 1904	de Saedeleer 1934	Deflandre 1953	Loeblich & Tappan 1960
Rhynchogromia	***Allelogromia***	***Allelogromia***	***Saedeleeria***
- linearis	*- brunneri*	*- brunneri*	*- gemma*
- nigricans	*- nigricans*	*- nigricans*	
- squamosa	*- squamosa*	*- squamosa*	
	- linearis		
Diplogromia	***Diplogromia*** *emend.*	***Diplogromia***	***Diplogromia***
- brunneri	*- gemma*	*- gemma*	*- brunneri*
- gemma			*- squamosal*
			-nigricans
		Penardogromia	***Penardogromia***
		- linearis	*- linearis*
			- palustris (1961)
	G. saxicola	*G. saxicola*	*G. saxicola*

Table 4.1 Classifications of agglutinated freshwater allogromiids

De Saedeleer (1934) revised Rhumbler's classification leaving *D. gemma* in its genus and creating a new genus *Allelogromia* for the *Rhynchogromia* species with *G. brunneri* as type species. Deflandre (1953) erected the genus *Penardogromia* for *G. linearis*, with the argument that it had a homogenous agglutinated test with calcareous particles. Loeblich & Tappan (1960) argued that the classification of De Saedeleer was unacceptable, because *G. brunneri* had been fixed as the type of *Diplogromia* by subsequent designation of Cushman (1928). They created the genus *Saedeleeria* for *G. gemma*, transferring *G. squamosa* and *G. nigricans* also to *Diplogromia,* but without giving any supporting explanations. Another agglutinated allogromiid, *Penardogromia palustris,* was described by Thomas (1961) from a freshwater marsh near Bordeaux (France).

Beside these descriptions there have been some scattered records of agglutinated freshwater allogromiids over the years (Wailes 1915; Hoogenraad & de Groot 1940; Grospietsch 1958; Siemensma 1982; Meisterfeld pers. comm.; Clauss, unpublished) and some photomicrographs available online (Revello 2015; Protist Information Server 2016).

Leidy (1879) was the first who described an allogromiid foraminifer, *Gromia terricola*, from a terrestrial habit. He found this non-agglutinated species “among moist moss in the crevices of pavements, in shaded places, in the city of Philadelphia”. A similar terrestrial organic walled allogromiid Edaphoallogromia australica has been described by (Meisterfeld *et al.* 2001).

Apart from these agglutinated and organic-walled species, some naked amoeboid freshwater species belonging to the family Reticulomyxidae have been described. The best known of these species is *Reticulomyxa filosa* (Nauss 1949), long time considered as an amoebozoan, until its foraminiferal affinity was demonstrated by molecular study (Pawlowski *et al.* 1999). Since then two new species of Reticulomyxidae were described: *Haplomyxa saranae* (Dellinger *et al.* 2014) and *Dracomyxa pallida* (Wylezich *et al.* 2014).

In an attempt to rediscover the allogromiids described by Penard and Blanc, Holzmann & Pawlowski (2002) examined samples from Lake Geneva. They did not succeed in finding any specimens by microscopic observations. However, several foraminiferal DNA sequences were obtained from the same sediment samples that built a monophyletic clade with the marine genera *Ovammina* and *Cribrothalammina* at its base. In a later report numerous environmental rDNA sequences revealed the existence of a large number of freshwater monothalamids branching in several clades. However, none of these clades (except clade 2 that comprises the family Reticulomyxidae) could be linked to known freshwater allogromiids (Holzmann *et al.* 2003). Further studies based on environmental DNA surveys showed that foraminifera are also a ubiquitous component of soil samples (Lejzerowicz *et al.* 2010; Geisen *et al.* 2015).

Here, we describe two new agglutinated freshwater species (*Lacogromia cassipara* gen. nov., sp. nov. and *Limnogromia sinensis* gen. nov., sp. nov.). *Lacogromia cassipara* is commonly encountered in mesotrophic water bodies in the Netherlands. We collected specimens from different locations and found two morphotypes. The other species, *Limnogromia sinensis*, is an isolate from China. We compare both new species with those described by Blanc, Penard and Thomas, with reference to the slides of the Penard Collection in Geneva. In addition, we describe a *Lieberkuehnia* species based on cultured material and report the first DNA data for

this species. Based on these data, we revise the taxonomy of agglutinated freshwater foraminifera and discuss their phylogeny and ecology.

4.4. Materials and methods

4.4.1. Sampling

Sediment samples containing morphotype A of *Lacogromia cassipara* collected weekly from March to May 2016 were taken from the bottom of a mesotrophic pond in the natural reserve Crailoo, 52°14′54.2″N 5°09′57.3″E (The Netherlands). A wide mouth pipette with an internal opening of 5 mm was used to collect the upper layer of the sediment from a depth of 30–40 cm. Every time a wide mouthed bottle was filled with 5 cm of sediment, transported to the lab and kept at room temperature on a windowsill on the north side. Small amounts of sediment were transported to 60 mm Petri dishes and examined with an inverted microscope. A Petri dish contained on average two specimens. 13 specimens were isolated with a micro pipette and kept in RNAlater and over 220 specimens were isolated to be examined, mea- sured and photographed with an upright microscope. A small number were kept in wet mounts in moisture chambers for observations.

One sample of morphotype B of *Lacogromia cassipara* was taken in April 2014 from a mesotrophic ditch in the natu- ral reserve of Laegieskamp, 52°16′39.0″N 5°08′24.7″E (The Netherlands). The ditch had a thick layer of organic sediment. The upper layer of the sediment was collected from a depth of *c*. 20 cm also using a wide mouth pipette. 7 specimens were isolated and preserved in guanidine for subsequent DNA extraction. Over 100 specimens were examined, measured and photographed with an upright microscope.

A small sediment aliquot, <1 cc, with specimens of *Limnogromia sinensis*, was taken from sediment of a shallow pond in the city park of Yangshuo (China) on October 2015 (24°46′48.5″N 110°29′07.1″E) and kept for three weeks in a closed mini tube. We found 11 specimens, 7 of them were isolated, photographed and fixed in RNAlater®, one was prepared as type specimen and the others were used for light microscopic study.

The cultured specimens of *Lieberkuehnia* sp. came from the river Havel in Berlin (Germany).

4.4.2. Morphological analyses

Living specimens of *Limnogromia* and *Lacogromia* were filmed and photographed with a Canon D70 camera using an Olympus BX51 microscope with following objectives: 10XAPLN, 20 × 0.75 APO, 60 × 0.90 APO with correction collar and 100 × 1.30 oil, all with DIC. This equipment was also used for the slides of the Penard Collection. Adobe Photoshop was used for processing and measuring. For searching samples and for isolating specimens of both new allogromiid species a Leitz Diavert inverted microscope was used.

Living cells of *Lieberkuehnia* sp. were filmed and pho- tographed with a Nikon TE2000U inverse microscope and a Jenaval microscope with DIC.

4.4.3. DNA extraction, amplification, cloning and sequencing

DNA was extracted using guanidine lysis buffer (Pawlowski 2000) for 22 specimens of *L. cassipara*, 4 specimens of *L. sinensis* and 16 specimens of *Lieberkuehnia* sp. DNA isolate numbers and accession numbers are given in Table 4.2. Semi-nested PCR amplifications of the 5' terminal barcoding fragment of small-subunit (SSU) rDNA were performed using primer pairs s14F3 (acgcamgtgtgaaacttg) – sB (tgatccttctgcaggttcacctac) and 14F1 (aagggcaccacaagaacgc) - sB.

The amplified PCR products were purified using High pure PCR Purification Kit (Roche Diagnostics) cloned with the TOPO TA Cloning Kit (Invitrogen) following the manufacturer's instructions and transformed into competent *E. coli.* Sequencing reactions were performed using the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) and analyzed on a 3130XL Genetic Analyzer (Applied Biosystems).

Species	Isolate	Accession numbers
Lacogromia cassipara	18849	LT576147 - LT56154
	18990	LT576139
	18991	LT576155
	18992	LT576140
	18993	LT576141
	18994	LT576142
	18995	LT576156
	18996	LT576143
	18997	LT576157
	18998	LT576146 - LT576158
	18999	LT576144
	19000	LT576159
	19001	LT576160
	19002	LT576145
	19179	LT604807
	19180	LT604808
	19181	LT604809
	19184	LT604813
	19185	LT604810
	19186	LT604811
	19188	LT604812
Limnogromia sinensis	18810	LT222211
	18811	LT222212- LT222213
	18812	LT222214 - LT222216
	18813	LT222217 - LT222219
Lieberkuehnia sp.	19189	LT604814
	19191	LT604815
	19192	LT604816
	19193	LT604817
	19194	LT604818
	19197	LT604819
	19198	LT604820
	19199	LT604821
	19200	LT604822
	19201	LT604823
	19202	LT604824
	19203	LT604825
	19205	LT604826
	19207	LT604827
	19208	LT604828
	19209	LT604829

Table 4.2 Isolate and accession numbers of sequenced freshwater foraminifera

4.4.4. Phylogenetic Analysis

The obtained sequences were manually aligned to 65 other foraminiferal sequences (43 freshwater sequences and 22 marine sequences) using Seaview software (Gouy *et al.* 2010). After elimination of the highly variable regions, 721 sites were left for analysis. The phylogenetic tree was constructed with maximum likelihood method based on the GTR + G model with 1'000 bootstrap replicates, using PhyML algorithms as implemented in the Seaview software.

We built a phylogenetic tree based on partial 18S rDNA with marine monothalamous foraminifera from several clades (Pawlowski *et al.* 2002b) and environmental freshwater and soil sequences. Moreover, the sequences from two formerly described freshwater/soil species (*Reticulomyxa filosa* and *Edaphoallogromia australica*) were added to the analysis. The tree was arbitrarily rooted on monothalamous clades A-C.

4.5. Results and Discussion

4.5.1. Taxonomic descriptions

Supergroup Rhizaria Cavalier-Smith 2002

Phylum Foraminifera (d'Orbigny 1826)

Monothalamids (Pawlowski et al. 2013)

Clade 3

***Lacogromia* gen. nov.**

Diagnosis: Test elongated to broadly pyriform or lens- or spindle-shaped, with a layer of small siliceous particles and commonly with some organic particles of debris. Test colourless or yellowish to almost black; aperture straight or oblique; test up to 1000 µm long. Generally with 1-8 nuclei, sometimes up to 30. Nuclei spherical, ovular. Peduncle and entosolenian tube asymmetrical.

Etymology: the prefix *Laco,* Latin for "pond", in reference to its freshwater habitat. The suffix -*gromia* refers to its relationship with *Allogromia*.

Type species: *Lacogromia cassipara*

New combinations:

Lacogromia squamosa (Penard, 1899) comb. nov.
Basionym *Gromia squamosa* Penard (1899)

Lacogromia brunneri (Blanc, 1886) comb. nov.
Basionym *Gromia brunnerii* Blanc (1886); synonym *Gromia gemma* Penard (1899)

Lacogromia palustris (Thomas, 1961) comb. nov.
Basionym *Penardogromia palustris* Thomas (1961)

***Lacogromia cassipara* sp. nov**. (Figure 4.1- 4.4)

Diagnosis: Test broadly ovoid to elongated pyriform, sometimes lens- or spindle-shaped, with a layer of small siliceous particles and usually with more or less organic particles from sediment. Test slightly flexible, colourless or light yellow, ochre, brown or almost black, 50-560 µm long; aperture oblique. Some specimens have a double ring around the aperture. Cell usually with 1-8 nuclei, sometimes up to 30. Nuclei spherical, with irregular but rounded pieces distributed throughout the nucleus with slightly more nucleoli in the periphery. No resting stages have been observed.

Etymology: *cassipara* is the Latin epitheton for "making a spider's web" that refers to the large web like granofilose reticulum of this species.

Type locality: Organic sediment, 40 cm deep, freshwater pond in nature reserve Crailoo in the central area of the Netherlands, located at 52°14'54.2"N 5°09'57.3"E.

Type specimen: The type specimen has been deposited in the Natural History Museum of Geneva (holotype in alcohol nr. MHNG-INVE-97019; 3 paratypes in alcohol, nr. MHNG-INVE-97020 and 5 paratypes in slides, nr. MHNG-INVE-97021, embedded in HYDRO-Matrix®).

Description: The general shape and structure and the corresponding terminology of agglutinated allogromiids with Lacogromia as an example are summarized in Figure 4.1.

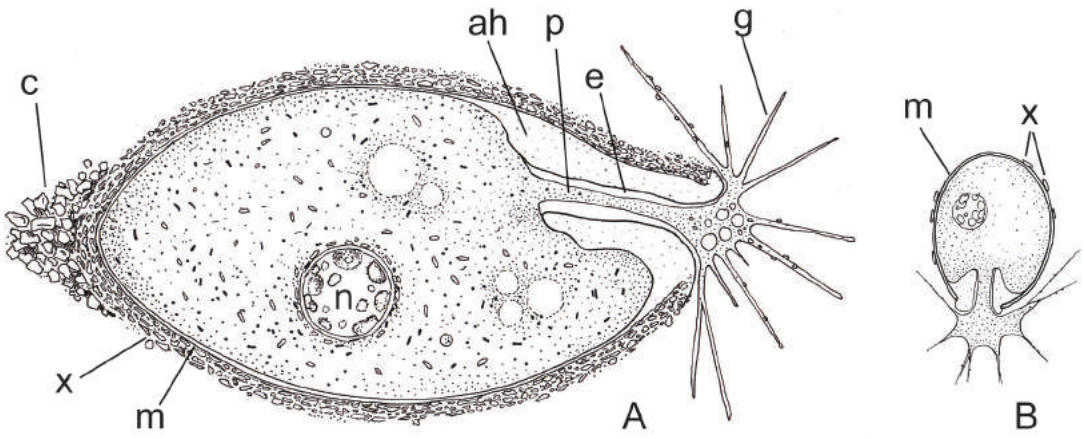


Figure 4.1 General morphology of Lacogromia cassipara. (A) Adult cell. (B) Young cell, test c. 50 µm. Abbreviations: ah - apertural hyaloplasm; p – peduncle; e - entosolenian tube; g – granuloreticulopodia; n – nucleus; m - membrane; x – xenosomes; c - cap of adhering bunch of particles.

The shape of the test is variable, ranging from broadly ovoid to elongated pyriform (Figure 4.2 A-F). Some, usually larger, specimens are rather subglobular (Figure 4.2 F), while other large specimens can have a more lens- or spindle-shaped outline (Figure 4.2 C). Smaller specimens, up to circa 160 µm, are always elongated ovoid (Figure 4.2 D).

The proximal end can be broadly rounded (Figure 4.2 A,E,F) or more conical (Figure 4.2 B,C). All tests are bilaterally symmetrical, usually with one side more curved than the other (Figure 4.2 B,E,F). Sometimes the less curved side bends slightly upwards towards the aperture (Figure 4.2 B,C,E). All tests are circular or nearly circular in cross section.

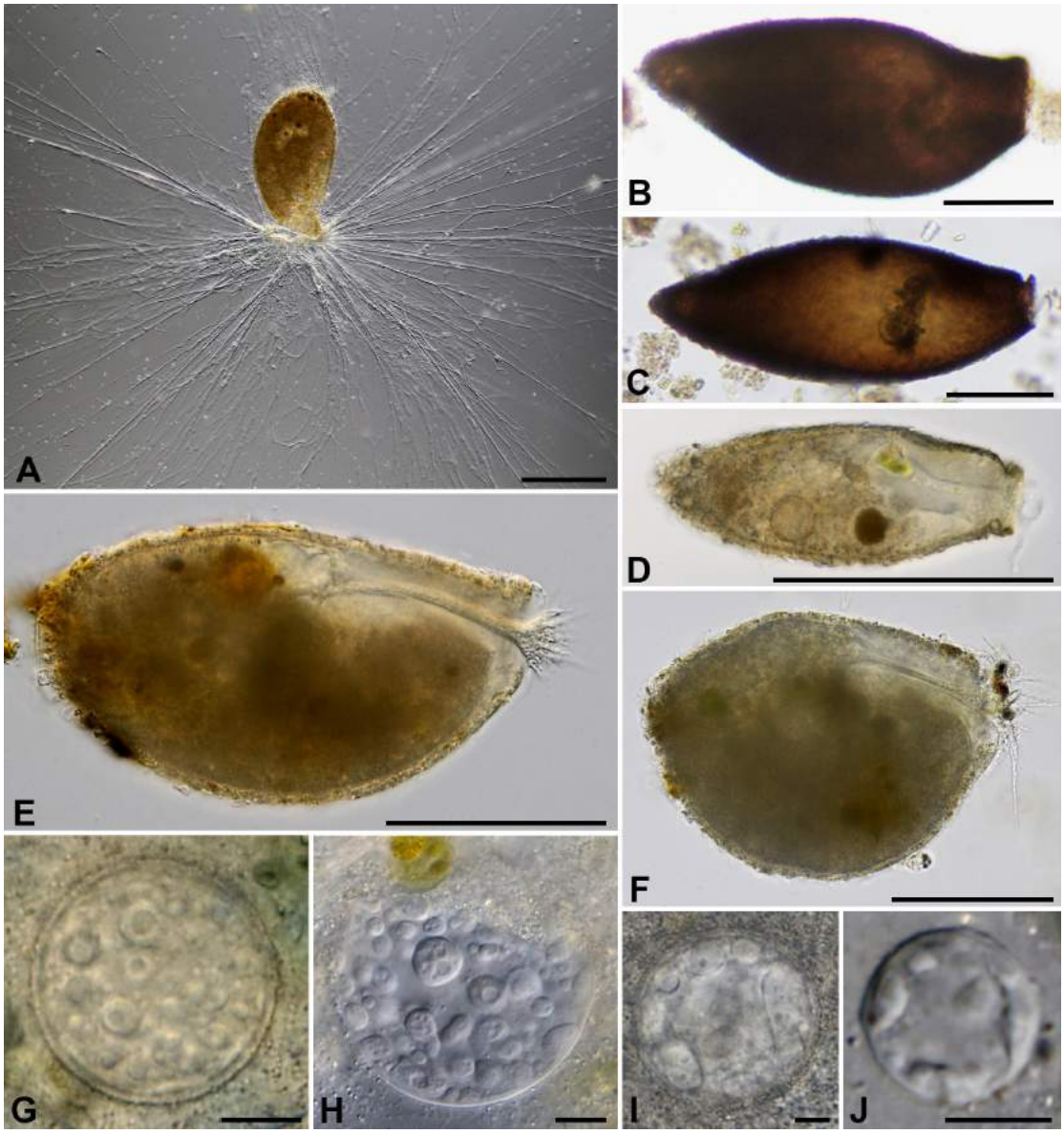


Figure 4.2 *Lacogromia cassipara*. (A) Cell with fully employed granuloreticulopodium. (B-D) Tests of morphotype B. (E-F) Tests of morphotype A. (G-J) Nuclei. Scale bars: (A) 200 µm, (B-F) 100 µm, (G-J) 10 µm.

The test wall is a thin membrane, not always visible and usually colorless, more or less flexible and covered with a layer of very small, irregularly shaped, usually flattened, particles, mainly siliceous, but organic material can also be present (Figure 4.3 C,J-K). The agglutinated layer is about 4-8 µm thick with the proximal area usually being thicker. The size of these particles is variable (*c*. 1-3 µm). Size and density of the particles may vary per specimen (Figure 4.3 J-K). Some particles could be identified as fragments of diatom shells. All these particles are probably held together by a kind of cement.

Specimens that were kept for many weeks in petri dishes with a small layer of sediment, had a thinner layer of particles than freshly collected specimens, probably because building material became scarce. Because all particles are more or less of the same size, it is likely that the material is selected by the foraminifera.

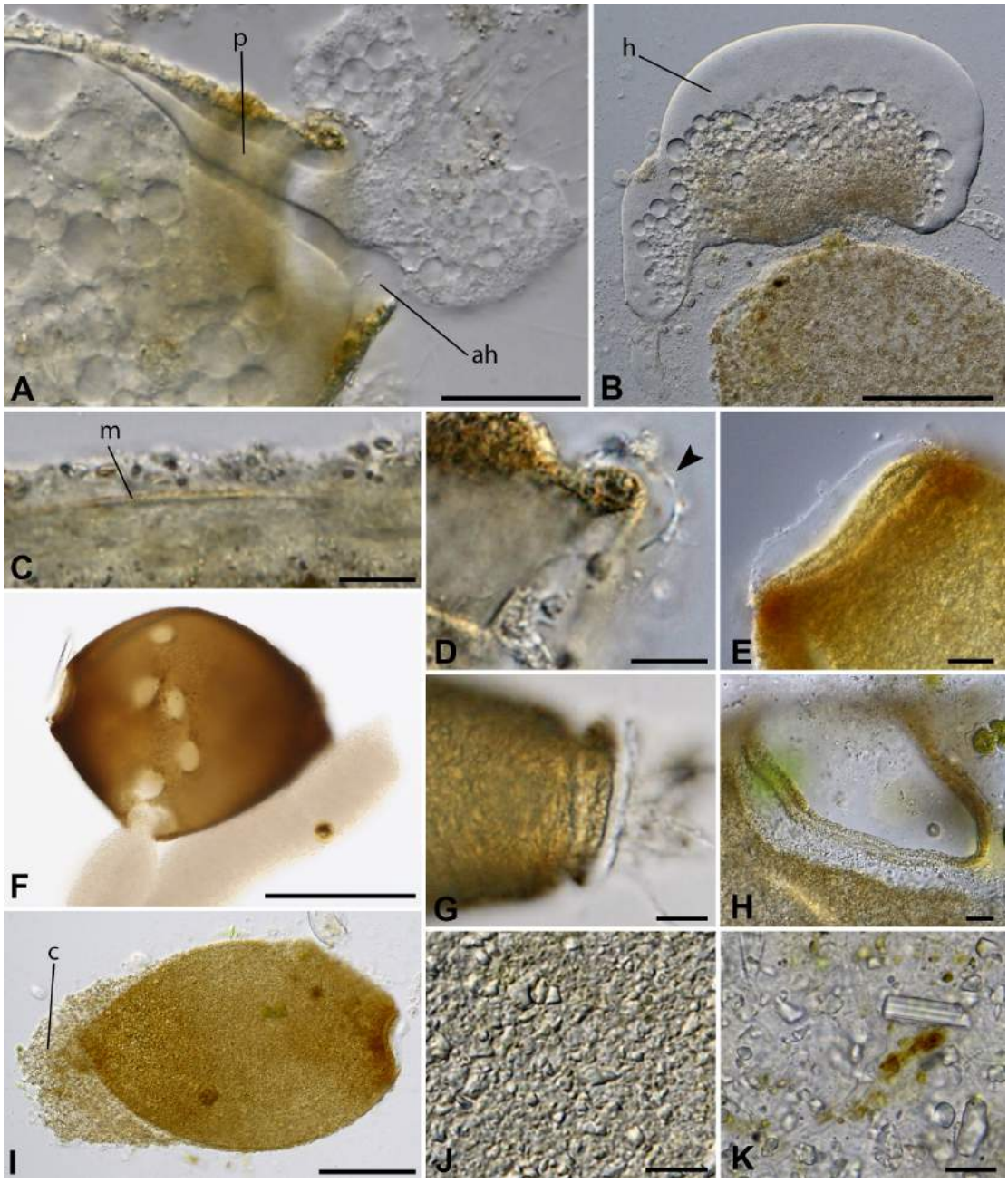


Figure 4.3 *Lacogromia cassipara*. (A) Apertural region with peduncle and entosolenian tube; specimen strongly flattened, pressed by cover glass. (B) Mass of cytoplasm, pressed out of the test. (C) Test wall with layer of xenosomes. (D) Detail of aperture; optical section with hyaline collar arrowed. (E) Apertural hyaline ring and double ring. (F) Empty test with holes,

probably made by offspring; when pressed, fine granular cytoplasm streamed out. (G) Test with constriction behind collar of hyaline material. (H) Aperture with double ring, strongly flattened. (I) Flattened test with cap of adhering particles and double ring. (J-K) Detail of surface of a test. Abbreviations: ah = apertural hyaloplasm; c - cap of adhering particles; co - collar. Scale bars: (A) 20 µm; (B,F,I) 100 µm; all other bars 10 µm.

Morphological variations: We found morphological differences between populations from different locations and consider them as different morphotypes (A and B). Cells of type A (Figure 4.2 A,E,F) look greyish or brownish when observed under transmitted light. The colour depends on the kind of food in the cytoplasm, the number of crystal-like particles and the colour of the agglutinated material in the test wall. Mineral material is commonly colorless, but organic particles are mostly ochre yellow, brown or black. Tests of type B vary in color, those of younger, smaller specimens are light ochre yellow (Figure 4.2 D), and tests of older, larger specimens are darker ochre yellow or reddish brown and black (Figure 4.2 B,C). The colour is not always evenly distributed. Usually the proximal and apertural regions are darker (Figure 4.2 C, Figure 4.3 F).

Another difference between the two types is the covering of the proximal part. Tests of morphotype B have an extra layer of loosely attached particles, resembling a kind of cap, while tests of morphotype A do not have any extra covering. Agglutinated particles of these caps are larger than the regular ones, up to 10 µm. Differences between both morphotypes are summarized in Table 4.3.

L. cassipara	**Morphotype A (n=224)**	**Morphotype B (n=109)**
Aperture	No pronounced collar, smooth	Distinct collar with double ring, often with constriction
Shape	Broadly ovoid-pyriform, fundus broadly rounded	Elongated ovoid-elongated pyriform, or spindle-shaped; fundus conical, rounded
Fundus	Without extra cap of xenosomes	Usually with cap of larger xenosomes
Structure	Particles loosely attached	Particles close to each other
Color	Colorless or light ocre yellow	Dark brown, ocre yellow or black
L/B ratio	1.1-2.4, mean 1.5	1.7-3.5, mean 2.3
Length	91-530 µm, mean 262 µm	123-560 µm, mean 267 µm
Width	48-407 µm, mean 173 µm	46-202 µm, mean 117 µm
Nuclei,	18-66 µm, mean 38.6 µm	8.7-77 µm, mean 29.0 µm

Table 4.3 Morphological differences between morphotypes of *Lacogromia cassipara*.

The length of all observed tests, both alive or empty, varied between 91 and 560 µm (mean 264 µm, std. dev. 77, n=333); with a width of 48-407 µm (mean 154 µm). The average length/breadth ratio is 1.8, with extremes between 1.1 and 3.5. Biometrical analysis showed differences in this ratio between both morphotypes (Table 4.3, Figure 4.4).

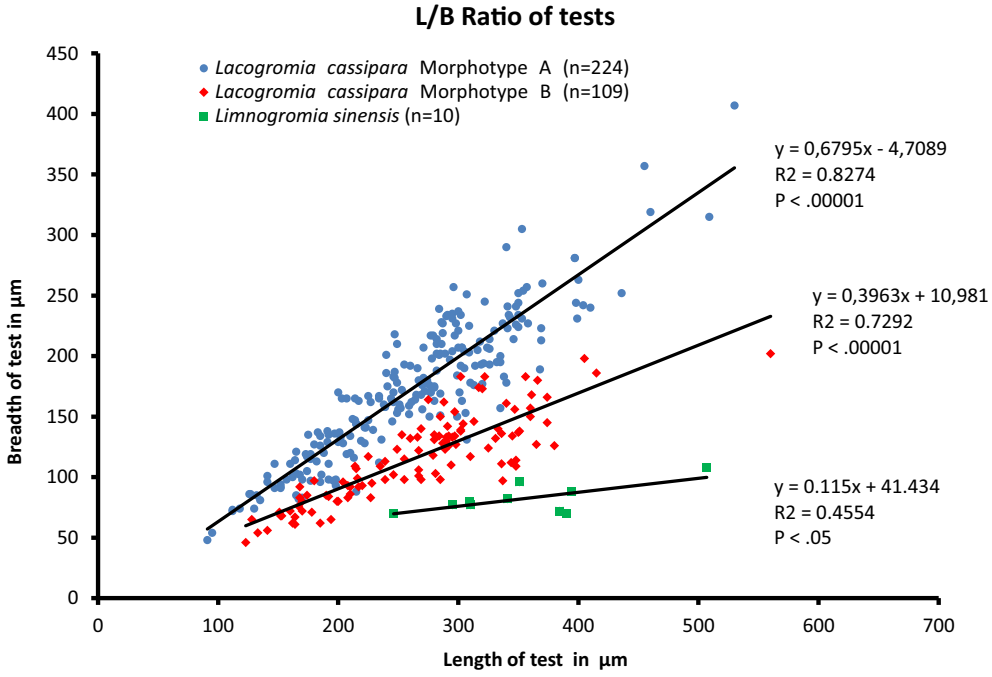


Figure 4.4 Biometric analysis of the length/breadth ratio of tests of *Lacogromia cassipara,* morphotype A and B, and *Limnogromia sinensis*.

Aperture. The test has one circular aperture, commonly at its smallest end, and usually cut obliquely. The diameter of the aperture is highly variable per test, between 9 and 133 µm, mean 42 µm. Specimens of morphotype B have a double ring around the aperture (Figure 4.3 E,H,I) built of particles and commonly with a more or less clear constriction behind this collar (Figure 4.2 C,D, 3G). The second ring of this collar is a little broader than the first one and also more pointed in cross section.

The granular cytoplasm is separated from the aperture by an area of extremely hyaline material, resembling a pierced rubber stopper. This hyaline material, which we call here apertural hyaloplasm, is attached to the rim of the aperture (Figure 4.3

D,E,G). It is translucent and only detectable by small granules and bacteria attached to its surface (Figure 4.3 D). In tests of morphotype B this hyaloplasm is attached to the second ring and in cross section visible as a clear curl (Figure 4.3 A,D).

The apertural hyaloplasm surrounds the entosolenian tube, which connects the granuloplasm with the surrounding environment. The narrow stream of cytoplasm flowing through this tube, the peduncle, is usually small in lateral view and broader in dorsal view. Sometimes two or more peduncles are present. The entosolenian tube is located eccentrically, usually on the less curved side, and becomes funnel-shaped towards the aperture, with the peduncle following its shape (Figure 4.2 E, 3A).

Although the almost featureless apertural hyaloplasm is difficult to detect visually, the presence within it of an entosolenian tube can be detected indirectly when larger particles are pushed through it, e.g. when the cell is pressed by the cover glass. In such a case, we observed that nuclei blocked the opening or passed the tube like a balloon which is pressed through a tube. The flexibility of the entosolenian tube could be observed when large food remnants were exported out of the cell. The same is true for phagocytose. Many cells contained food particles like rotifers and algae that were much larger than the diameter of the aperture and the entosolenian tube, so the cell must widen its aperture and entosolenian tube to engulf these large objects.

Based on our observations a cell can change the shape and amount of its apertural hyaloplasm dynamically. When a cell is disturbed it can decrease the amount of the hyaloplasm rather quickly.

Cytoplasm: The cytoplasm is granular with a large number of yellowish birefringent rod-like particles, probably crystals, about 1.3 µm long. One or more vacuoles of different size are present and smaller ones may fuse. We could not observe any contractile vacuole, probably because of the constantly moving plasm and the opaqueness of the test. When cytoplasm is pressed or squeezed out of the test, a zone of viscous hyaloplasm is formed together with a large number of non-contractile vacuoles (Figure 4.3 A,B). Pseudopodia are granuloreticulopodia with bidirectional streaming as is characteristic for foraminifera. They emerge from the peduncle.

Nucleus: About 26% of the cells (n=333) had one nucleus, while the other cells had 2-8 nuclei, except two cells which had over 20 and 30 nuclei respectively. The nuclei vary in diameter from 8.7-77 µm. Uninucleate cells have the largest nuclei while

multinucleate cells have smaller ones, this probably corresponds to different life stages as described for *Allogromia laticollaris* in Parfrey & Katz (2010). The nucleoli are irregularly rounded and distributed throughout the nucleus with slightly more at the periphery (Figure 4.2 G-J). These nucleoli are about 1.4-14.6 µm in diameter. Large nucleoli may show one or more small lacunae (Figure 4.2 G-H). The amount of nucleoli in a nucleus may strongly differ per cell.

The nuclei are constantly rotating, with frequent changes in direction. In living cells, nuclei are difficult to observe in detail because of the opaqueness of the agglutinated wall. When nuclei are squeezed out of the test, they usually escape through the smaller entosolenian tube or a tear or rupture in the test and get damaged. Within a minute, the nucleoli disintegrate and a weakly granular nucleus remains.

Reproduction: We could not observe the complete life cycle, but did observe an isolated specimen that divided overnight in two daughter cells. In the past, we have observed schizogony with multiple fissions of a specimen which we now recognize as *L. cassipara*. In this “medium sized” specimen (about 250 µm long), we observed the large nucleus dividing into over 30 nuclei. The following day, 36 small daughter cells were observed around the empty test (Siemensma 1982). All daughter cells were about 50 µm long. They had a smooth membrane which became covered with particles during the next days (Figure 4.1 B).

In the recent sample about 28% of all observed tests were empty. Empty tests were on average larger, 317 µm in length, compared with 241 µm for living cells. Most empty tests showed holes in their wall, usually in the median area (Figure 4.3 F). We assume that these holes were made by offspring when leaving the test.

Phylogenetic position: Based on partial SSU rDNA sequences, *Lacogromia cassipara* branches within group 3 (Figure 4.5). This clade is composed exclusively of environmental sequences obtained mainly from samples collected in Geneva basin.

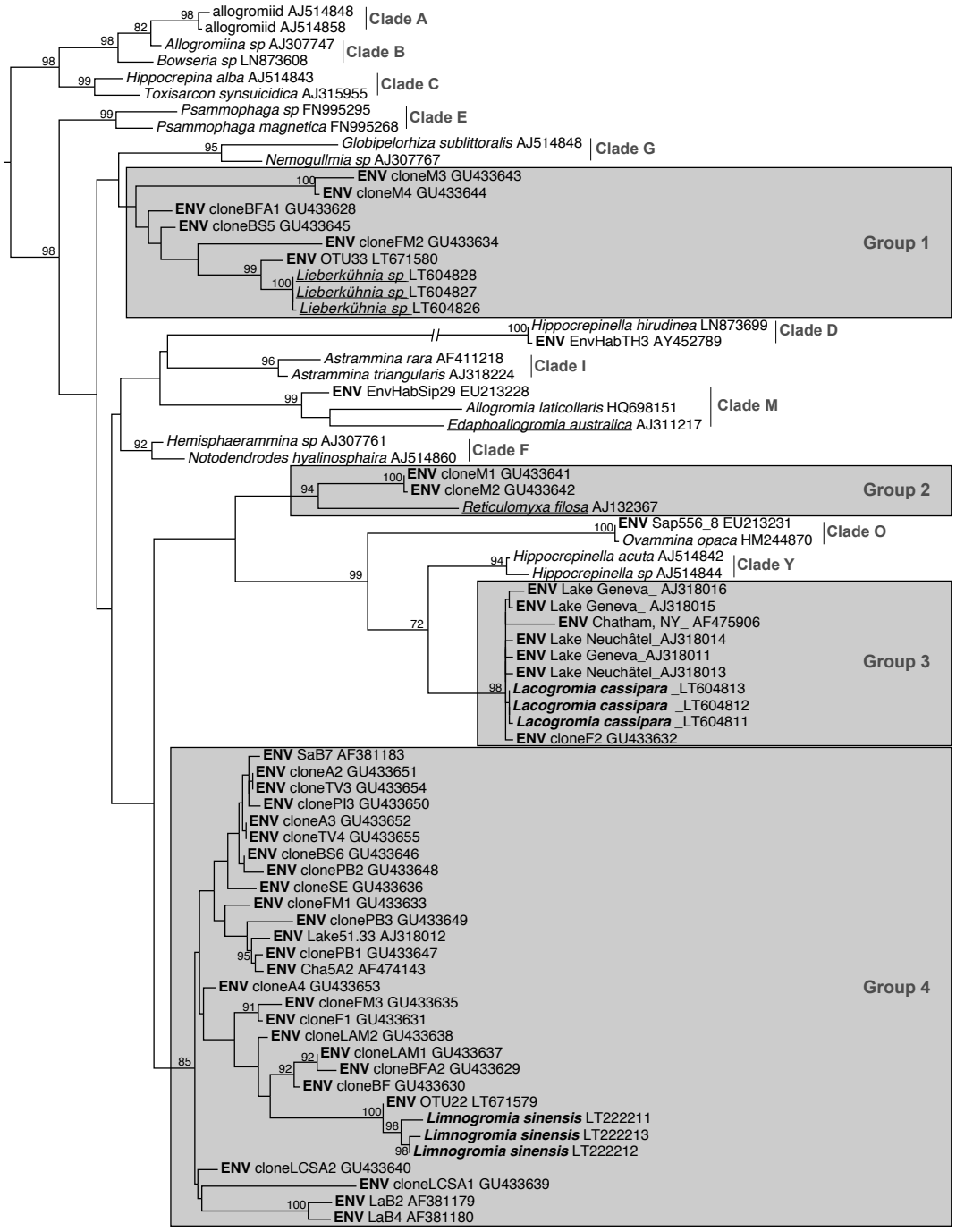


Figure 4.5 Phylogenetic tree based on partial 18S rDNA of 74 sequences of foraminifera, including 44 environmental sequences (ENV) and 3 freshwater species (underline). Newly described species are highlighted in bold. Grey boxes correspond to the four freshwater clades. Only bootstrap values greater than 70% are shown.

Ecology: One population of *L. cassipara*, morphotype A, was present in the surface layer of organic sediment in a shallow mesotrophic freshwater pond. This pond is

part of Zanderij Crailoo in the Netherlands, an area were sand has been excavated between 1870 and 1971. The area is fed by ground water and now a nature reserve. Common amoeboid organisms in the sample were *Pelomyxa flava*, *Difflugia binucleata*, *D. pyriformis* and *Centropyxis ecornis*. Characteristic algae were *Micrasterias americana* and *M. rotata*.

Lacogromia cassipara feeds on diatoms (e.g. *Navicula spp.*, *Diatoma vulgarum*, *Tabellaria spp.*), blue and green algae (*Ankistrodesmus spp.*, *Phacus triqueter*, *Euglena acus*, *Cosmarium spp.*), filamentous algae (*Hyalotheka spp.*) and fungal spores. We also noticed rotifers and small testate amoebae (*Euglypha rotunda*, *Cryptodifflugia oviformis*) in cells and once a small nematode had been engulfed. Generally speaking one can say that *L. cassipara* feeds on anything it can get; it is omnivorous.

The observed population of morphotype B was isolated from a ditch in the nature reserve Laegieskamp, also in the Netherlands, and about 6 km away from Zanderij Crailoo, with similar environmental conditions. Both populations were discovered in early spring, specimens were abundant in April and disappeared end of May. Other findings of *L. cassipara* also come from shallow mesotrophic water bodies, like ditches in the Hol, Naardermeer and Westbroekse zodden, all old peat bogs in the central area of the Netherlands. It was also found in the flood plain of a small oligotrophic stream near Renkum, the Netherlands. Another location, which is also oligotrophic, is the Diepveen, a fen in the northern part of the Netherlands, 200 km distant from Zanderij Crailoo. However, our findings over the years are very scarce.

Remarks: *Lacogromia cassipara* resembles in its pyriform shape *Gromia brunneri*, but it differs from it in the structure of the nucleus and the much thinner test wall. It differs from *Gromia squamosa* in several aspects. *G. squamosa* is much larger, always spindle-shaped, with a thick layer of particles, and in cross section its test is more elliptical than circular and sometimes strongly compressed. Measured specimens from Penard's permanent slides show that *G. squamosa* has an average L/B ratio of 3.2 vs. 1.5 and 2.4 for morphotypes A and B of *L. cassipara* respectively. The structure of its nucleus is quite different from all other known freshwater allogromiids. It has an internal layer, by which the nucleus resembles "a very thick ring bordered on its inner contour with a clear, dark line (...) which consists of small elongated flakes" (Penard 1902), a phenomenon that has never been observed in *L.*

cassipara. *G. nigricans, G. linearis* and *G. saxicola* differ from *L. cassipara* by their elongated tubular and much more flexible test. It differs from *P. palustris* in its general shape and the straight aperture of the latter.

Monothalamids (Pawlowski *et al.* 2013)

Clade 4

***Limnogromia* gen. nov.**

Diagnosis: Test cylindrical tot elongated cylindrical, agglu- tinated, encrusted with a large number of small siliceous particles. Test very flexible, extendible and pliable. Up to 200 ovular nuclei. Peduncle and entosolenian tube asymmetrical.

Type species: *Limnogromia sinensis*

Etymology: the prefix limnos of the genus name refers to the freshwater habitat. The suffix - gromia refers to the relationship with Allogromia.

New combinations:

Limnogromia saxicola (Penard, 1905) comb. nov.
Basionym *Gromia saxicola* Penard (1905)

Limnogromia nigricans (Penard, 1902) comb. nov.
Basionym *Gromia nigricans* Penard (1902)

Limnogromia linearis (Penard, 1902) comb. nov.
Basionym *Gromia linearis* Penard (1902)

***Limnogromia sinensis* sp. nov.** (Figure 4.6)

Diagnosis: Test cylindrical, agglutinated, encrusted with a large number of small siliceous particles. Test very flexible, extensible and pliable; neck can bend very strongly and the proximal end can be stretched like a spine. Multinucleate, up to 200 nuclei; nuclei very small, usually spherical but sometimes ovoid with nucleolar material laying close to the nuclear membrane. Test 235–411 m long (mean 345 m) and 65–75 m broad (n = 11); nuclei 6.0–8.2 m in diameter.

Etymology: *sinensis* is a toponym with suffix -ensis which refers to the country of the type locality, China.

Type locality: 24°46'48.5"N 110°29'07.1"E, city park of Yangshuo, China (October, 2015).

Type material: The type specimen has been deposited in the Natural History Museum of Geneva (holotype in alcohol, nr. MHNG-INVE-97022; 3 paratypes in alcohol, nr. MHNG-INVE-97023).

Description: Cells of *L. sinensis* have a cylindrical yellowish to brownish test with an organic wall, encrusted with a large number of very small siliceous particles, laying closely packed together (Figure 4.6 H). Tests are 235-411 µm long (mean 345 µm) and 65-75 µm broad (n=11). The L/B ratio is 4.3 (3.5-5.6). Though the tests are tubular, they are not of equal width throughout. The area around the aperture is pliable and extensible and the neck can bend very strongly, through nearly 180 degrees (Figure 4.6 C,D). On one occasion we observed a fold in the neck region indicating that the neck was twisted (Figure 4.6 F). The proximal end is usually rounded, but a specimen, kept in a petri dish for some weeks, showed an extensible proximal end which was pulled out far, shaped like a spine (Figure 4.6 E,G). The same specimen could also widen its aperture to resemble a funnel (Figure 4.6 E-F). One specimen was squeezed between cover and object glass, which caused most nuclei to be ejected. We counted up to 150 nuclei and estimated the total number around 200. The nuclei were 6.0-8.2 µm in diameter, usually spherical or ovoid, with small pieces of nucleolar material laying close to the nuclear membrane (Figure 4.6 B). No resting stages have been observed.

Figure 4.6 *Limnogromia sinensis.* (A) Common habitus. (B) Nucleus. (C-D) Micrographs showing the flexibility of the neck. From C to D is 3 minutes. (E) Specimen with elongated fundus. (F) Same specimen with twisted and folded anterior part. (G) The same specimen, S-shaped. (I) Detail of test. (H) Unknown agglutinated freshwater species from Uruguay (photomicrograph Revello, 2015). Scale bars: (B) 5 µm; H- 10 µm; all other bars 100 µm.

Phylogenetic position: *Limnogromia sinensis* branches within group 4, close to OTU22, an environmental sequence found in a river located in the Geneva basin (Figure 4.5).

Ecology: We have only restricted observations of *Limnogromia sinensis* because of the small amount of specimens we had. Observed food were diatoms and blue algae, but probably it is omnivorous.

Remarks: Besides the molecular data, its morphology characterizes *Limnogromia sinensis* as a new genus, and consequently new species. The overall shape and structure of the cell and the number of small nuclei are quite different from any other described freshwater foraminifer, except *G. saxicola* (Figure 4.8 G-J). Both species have a tubular, but highly flexible test which can stretch, bend and twist and which can form trails in viscous appearance, and an aperture that can be transformed into a funnel. Both species have up to 200 small nuclei. Nuclei of *L. sinensis* have a diameter of 6.0-8.2 µm, which is comparable to the measurements given by Penard (1905) for nuclei of *G. saxicola* (6-8 µm). However, nuclei in preserved cells of *G. saxicola* (slide 437 of the Penard Collection) are 3.3-4.5 µm in one cell and 4.9-6.2 µm in another cell.

There are other differences. *G. saxicola* has a blackish test, according to Penard resembling a *Difflugia* species, while tests of *L. sinensis* are yellowish and smooth. *G. saxicola* was found at a depth of 20 à 40 m., while *L. sinensis* was isolated from very shallow water.

Morphologically *Limnogromia* and *Lacogromia* differ mainly in three characters: *Limnogromia* has a tubular test which is very flexible with up to 200 nuclei. *Lacogromia* has a more or less pyriform test, ranging from nearly circular to spindle-shaped, that is more or less flexible, and contains a cell body with rarely more than 30.

An interesting observation appeared end of 2015 on YouTube, where Carlos Revello published a video of what seems to be an unknown *Limnogromia* species (Figure 4.6 I). It was found in a freshwater brook near San José, Uruguay (Revello, pers. comm.). Similar micrographs were published on the Japanese Protist Information Server (2016), showing specimens from the USA which also resemble *Limnogromia*. A flexible test has also been observed in some marine agglutinated monothalamids such as Cedhagenia saltatus (Gooday *et al*. 2010)

Monothalamids (Pawlowski *et al.* 2013)

Clade 1

***Lieberkuehnia* sp.** (Figure 4.7 A-C)

Morphology: The genus *Lieberkuehnia* foraminiferal specimens with an ovoid or spherical flexible organic-walled test with a single aperture. An entosolenian tube built of hyaloplasm separates the main cytoplasm mass from the pseudopodial peduncle. The pseudopodial peduncle is at the origin of the pseudopodial network as well as a cytoplasm layer which surrounds the cell.

The cytoplasm of *Lieberkuehnia sp*. is colourless, yellowish, brownish or greenish, shows continuous cytoplasm streaming and contains over 100 nuclei and many vacuoles. Nuclei are granular with usually two or three relatively small rounded nucleoli with one or two lacunae each. Well-fed cells are filled with cytoplasm and also have a layer of hyaloplasm completely surrounding the tests. Starving cells often lack from that surrounding cytoplasm and sometimes it does not even fill the test completely. The test of a well-fed cell is not always easy to detect as the inner cytoplasm is difficult to differentiate from the one surrounding the cell. We have observed cells with a test size ranging from 50 m to 300 m. The pseudopodial network can be very large, extending some millimetres over the substrate. Sometimes the main cell body is covered by detritus.

Reproduction: We have observed two different modes of reproduction. Most often the main cell body divides into several cells (up to 5). Although each new cell has its own pseudopodial peduncle, young cells often share at least parts of the pseudopodial network. Sometimes a second form of reproduction was observed. Within the pseudopodial network a blob of plasma is formed. This blob then forms a new test and a new peduncle. Initially, the new cell is connected with the pseudopodial network of the old cell.

Phylogenetic position: *Lieberkuehnia* sp. branches within clade 1 in the SSU rDNA phylogenetic tree (Figure 4.5). It is closely related to the OTU33 from Geneva basin.

Other OTUs present in this clade have been reported from soil samples (Lejzerowicz *et al.* 2010).

Ecology: In our cultures *Lieberkuehnia* sp. fed mainly on diatoms and green algae.

Remarks: For the moment we leave this *Lieberkuehnia* species in open nomenclature, because we are not fully convinced that it is identical to *L. wageneri* as described by Claparède & Lachmann (1859) and re-described by Penard (1907) and Mrva (2009). *Lieberkuehnia* sp. resembles *L. wageneri* in some aspects(general shape, entosolenian tube and pseudopodial peduncle), but also shows some diffferent morphological features. We will address the species problem in this genus in a separate paper including additional sequences from different *Lieberkuehnia* strains, which we have in culture.

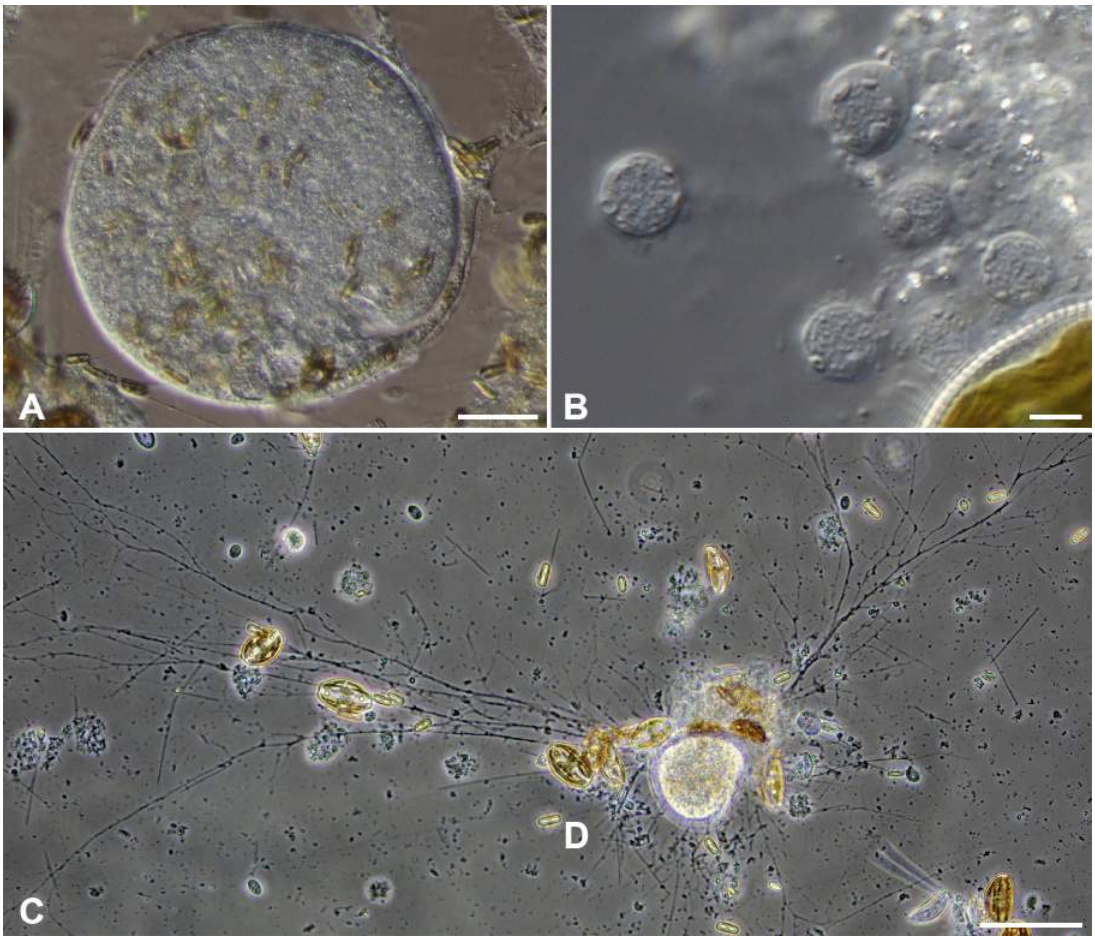


Figure 4.7 *Lieberkuehnia* sp. *(*A) Common habitus. (B) Nuclei. (C) Cell with pseudopodial network. Scale bars: (A) 50 µm. (B) 10 µm. (C) 200 µm.

4.5.2. Taxonomic revision of some historical freshwater foraminiferal species and genera

The known agglutinated freshwater allogromiids share morphological similarities, which lead to some complications when reading the original descriptions and viewing Penard's slides as well as the illustrations made by Blanc (1888), (Penard 1899, 1902, 1905) and (Thomas 1961). Penard (1899), a careful observer and describer, already recognized the difficulty of differentiating between several species. Therefore the question is: how well defined are those classical species and genera?

***Gromia brunneri* Blanc 1886**. Penard (1899) states that Blanc's *G. brunneri* might in fact represent three different species: *G. brunneri*, *G. gemma* and *G. squamosa*. However, based on the descriptions of Blanc (1886, 1888), we cannot agree with Penard. Blanc describes the largest specimens of *G. brunneri* (500-1000 µm) as being ovoid to almost spherical, and the smallest specimens (200 µm) as spindle- or bottle-shaped. This description does not fit the features of *G. squamosa*, which is a large spindle-shaped species (Figure 4.8 C), up to 1000 µm. Morphotype A of *L. cassipara* is in this respect similar to Blanc's *G. brunneri*, with larger specimens being almost spherical and smaller specimens being spindle-shaped.

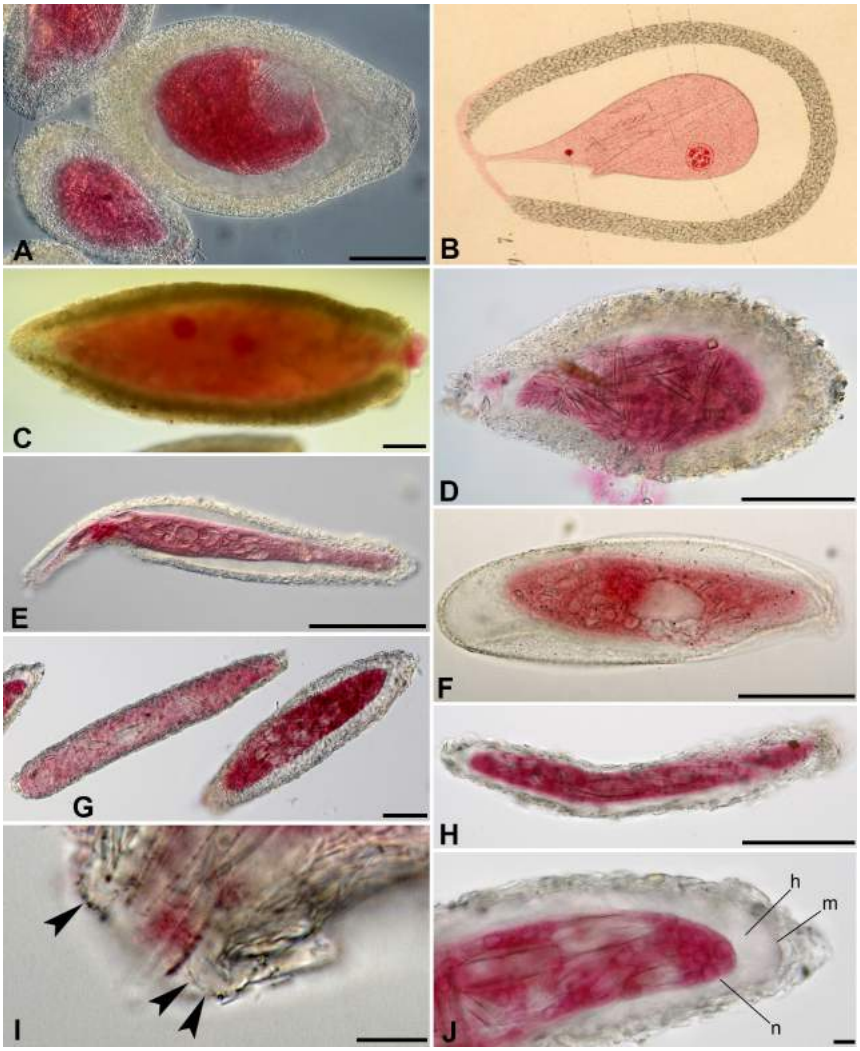


Figure 4.8 (A) *Gromia brunneri.* (B) *Gromia brunneri,* in Blanc 1888. (C) *Gromia squamosa.* (D) *Gromia gemma.* (E-F) *Gromia linearis.* (G-J) *Gromia saxicola.* Hyaline collar arrowed. All images, except B, are from the Penard Collection in Geneva. According to Penard all specimens were treated with alcohol, stained with borax-carmine and embedded in Canada balsam. Abbreviations: h = hyaline plasm; m – cell membrane; n – nuclei. Scale bars: (I-J) 10 µm; all other bars 100 µm.

***Gromia gemma* Penard 1899**. In the original description of *G. gemma*, Penard (1899) mentioned the thick internal mucous layer as an important character. In a later publication (Penard 1902) he remarks that this layer is not visible in living cells, but only in stained preparations. In 1905 Penard also observed such an internal mucous layer in *G. brunneri*. We were able to repeat his experiment of pressing cells out of their tests, but what Penard describes as a mucous layer is in our opinion just a layer of viscous hyaloplasm (Figure 4.3 B). In a later description of *G. gemma* Penard (1905) does not even mention this mucous layer, which should be so characteristic.

In summarizing the main differences between his *G. brunneri* and *G. gemma* he only mentioned the size of the test, the thickness of the test wall and the oblique aperture. According to Penard (1905) further differences concern the test wall that is much thinner in *G. brunneri* than in *G. gemma*. His 1902 illustration of *G. brunneri* shows an extremely thin wall, almost a membrane with some attached particles. However, the numerous specimens in his two slides labeled "*G. brunneri*", have a very thick test wall, between 20 and 77 µm (Figure 4.8 A). We compared the only specimen of *G. gemma* in the Penard Collection with those of *G. brunneri*, and found no significant difference (Figure 4.8 A,D). All these specimens are also very similar to the drawings given by Blanc (Figure 4.8 B). Penard (1905) remarked that *G. brunneri* and *G. gemma* might be one species, as he considered the three main differences mentioned above as not very important. Based on Penard's statement and our observations of his slides we consider *G. gemma* as a junior synonym of *G. brunneri*. Penard also supposed that *G. gemma* is an adult stage of *G. brunneri*, but that seems to be less likely considering the dimensions given by Blanc for *G. brunneri* (200-1000 µm) and those by Penard for *G. gemma* (200-600 µm). The more than 30 specimens of *G. brunneri* preserved in the Penard Collection measure 160-670 µm.

Gromia squamosa Penard 1899. In our opinion a well described species. Large and robust, spindle-shaped, with typically its broadest part at one third of the test measured from the aperture. Tests in the Penard Collection are 383-783 µm long (Figure 4.8 C).

Gromia linearis Penard 1902. Penard's description of *G. linearis* seems clear. Slide 433 of the Penard collection contains four specimens, all labeled "*G. linearis*" (Figure 4.8 E,F), but one of them is very different in shape and structure (Figure 4.8 F). It has a thin test wall and an elongated ovoid shape. The nuclei of the four specimens have the same structure but they most probably do not belong to the same species.

Gromia nigricans Penard 1902. This species resembles in its general shape *G. squamosa* and smaller specimens of *Lacogromia cassipara*, but differs strongly from both species by its highly flexible and pliable test, which resembles those of *G. linearis*, *G. saxicola* and *Limnogromia sinensis*. *G. nigricans* has also been found by Hoogenraad & de Groot (1940) and their observations correspond to those of Penard. The four specimens observed by Wailes (1915) and labeled *G. nigricans* represent probably a *Lacogromia* species.

Gromia saxicola Penard 1905. In our opinion a well described species, morphologically closely related to *Limnogromia sinensis*.

Penardogromia palustris Thomas 1961. According to Thomas (1961) the test is covered with calcareous particles, but we doubt if this is specific to this species and therefore distinctive feature. In fact, we find the description of *P. palustris* insufficient to distinguish it from other related species. Though Thomas describes the test as elongated tubular, he did not mention anything about the flexibility, extensibility and pliability of the test, which is so characteristic for tubular species. Based on the original drawing (Thomas 1961) the species resembles much more a small *Lacogromia* than a *Limnogromia* species. The test in this drawing (Thomas 1961) also resembles the deviating specimen in slide 433 of the Penard Collection (Figure 4.8 F).

Rhynchogromia Rhumbler 1894. Rhumbler transferred *G. squamosa*, *G. nigricans* and *G. linearis* to *Rhynchogromia* based on the assumption that the small particles in the test wall of these species are mainly secreted. However, he stated that there is an important difference between his *Rynchogromia variabilis* and the three *Gromia* species, because Penard and Blanc both described the test wall particles as siliceous plates and rods, while the particles of *R. variabilis* are not of siliceous origin. There is no reason to assume that the particles in the test walls of the three *Gromia* species are secreted. Firstly, neither Penard nor Blanc mentioned this option. Secondly, in the preserved *Gromia* specimens from the Penard collection, the small particles were comparable with those in the test wall of *L. cassipara*, including diatom frustules. Because the particles in all examined agglutinated freshwater foraminifera are true xenosomes, these species cannot be assigned to *Rhynchogromia*, as originally defined.

Diplogromia Rhumbler 1904. This genus is characterized by the presence of an internal mucous test wall. Its type species is *G. brunneri*, according to Loeblich & Tappan (1960), but this species does not have such a layer. What Blanc (1888) considered to be a second internal layer, is just the cell membrane, as is clearly visible in his drawings (Figure 4.8 B). Therefore we reject *Diplogromia* as a legal genus.

Allelogromia De Saedeleer 1934. The genus *Allelogromia* has been rejected by Loeblich & Tappan (1960) as being a junior synonym for *Diplogromia*.

Penardogromia Deflandre 1953. This genus was designed by Deflandre for species with a homogenous agglutinated test with calcareous particles, similar to some tests of agglutinated miliolids. He based the introduction of this new genus on his observations of Penard's slide of *G. linearis* in polarized light, but without giving any additional information. We also observed Penard's slides in polarized light, but did not find any significant difference between the material in the tests of all preserved species. According to Penard (1902) the test of *G. linearis* is comparable with those of *G. squamosa* and *G. brunneri*. We agree with Penard and therefore we do not accept this genus.

Saedeleeria Loeblich & Tappan 1960. This genus was designed for *G. gemma*, but as we consider this species as a junior synonym of *G. brunneri*, it is rejected.

4.5.3. General remarks on morphology, ecology, and taxonomy of freshwater agglutinated foraminifera

Morphology: This is the first time that both morphological and molecular data of agglutinated foraminiferal freshwater species could be acquired and used to revise the taxonomy of this poorly known group. The obtained results allow an increased understanding of the morphological variation within the different freshwater foraminiferal clades. Both new described species closely resemble in their general morphology the classical ones described by Blanc, Penard and Thomas. All species have an agglutinated test with an entosolenian tube and a peduncle. Though an entosolenian tube has only been described for *G. gemma* by Penard (1899), we could also detect it in two stained specimens of Penard's slides: in *G. brunneri* and *G. saxicola* (Figure 4.8 I), where small particles attached to the surface of apertural hyaloplasm made the tube visible, just as in *L. cassipara* (Figure 4.3 D). Because all known species have the same overall structure, we assume that all classical species have such a tube. Penard (1902) described how difficult it is to detect this tube, because the surrounding material is "as clear as water" as we confirmed. He also noted that the tube is only visible in stained preparations and never in living cells. Blanc (1888) remarked that *G. brunneri* does not have such a tube, but that is unlikely, given the presence of a tube in his drawing of this species (Figure 4.8 B). In the same publication he mentioned the opaqueness of the test which prevented any

clear observation and which might be the reason why he was not able to detect a tube.

The function of an entosolenian tube might be to protect the cell against penetration by predators and/or parasites, comparable with the diaphragms and/or narrow apertures in some testate amoebae, e.g. *Lesquereusia*, *Zivkovicia* and *Cucurbitella*, which prevents rotifers from laying their eggs inside (de Smet 2006).

With the exception of *L. sinensis* and *G. saxicola*, all known agglutinated freshwater foraminifera are mononucleate, having one large nucleus, usually 60-77 µm in diameter, or multinucleate, with smaller nuclei, usually 2-8 but sometimes more than 30 in number. Only *L. sinensis* and *G. saxicola* have a large number of small nuclei, up to 200. The number of nuclei in a cell might be related to different life stages as has been described for some other monothalamids (Goldstein & Barker 1990; Parfrey & Katz 2010). Due to the limited number of specimens available for observation we cannot exclude that *L. sinensis* and *G. saxicola* also possess uninucleated specimens. Comparing the nuclei of the two newly described species with those preserved on slides is also difficult as nuclei disintegrate rapidly once removed from the cytoplasm. Penard squeezed tests to get the nuclei out of it and also stained and observed them, so we do not know if damaged ones have been described.

Differences between both morphotypes in *L. cassipara* could be induced by environmental factors; for example, the amount of iron could affect the color of the test, as has been described for *Gromia oviformis* (Hedley 1960).

Ecology: Freshwater foraminifera seem to be rare, given the very scarce microscopic records over the years. However, molecular data show a rich diversity of freshwater and soil foraminiferans (Pawlowski & Holzmann 2002; Lejzerowicz *et al.* 2010). The close relationship between *Lieberkuehnia* sp. and *Limnogromia sinensis* with environmental sequences (OTU33 and OTU22) suggests that same kinds of morphotypes might also live in the Geneva basin. Members of clades 3 and 4 represented by *Lacogromia* and *Limnogromia* respectively seem to be present in all types of habitats tested molecularly (lake, small and big river, pond, soil) in the Geneva area. Group 3 and 4 are represented by more sequences than group 1 and 2 (Chapter 6), which suggests that the species described by Penard might still occur in the Geneva basin.

Taxonomy: Based on molecular phylogenetic data, we could place a morphologically described species in each of the major freshwater foraminiferal clades. *Lieberkuehnia* clusters with clade 1, *Reticulomyxa* with clade 2, *Lacogromia* is a member of clade 3 and *Limnogromia* is a representative of clade 4. As non of the classical genera are well established, we transfer the classical species to either *Lacogromia* (*G. brunneri*, *G. squamosa* and *P. palustris*) or *Limnogromia* (*G. linearis*, *G. saxicola* and *G. nigricans*). As criteria we choose the flexibility and shape of the test. We are aware that our choice is arbitrary, but for the moment it is the only useful morphological character.

CHAPTER 5

ENVIRONMENTAL DNA METABARCODING REVEALS HIGH DIVERSITY OF FRESHWATER FORAMINIFERA IN THEIR TAXONOMIC HOME

LAURE APOTHÉLOZ-PERRET-GENTIL, JAN PAWLOWSKI

Manuscript in prep.

5.1. Project description

This project was the main topic of my research at the beginning of my thesis together with the attempt to unveil the morphology of freshwater foraminifera. I sampled the Geneva basin regularly for 3 years, looked at fresh sediment under the microscope for uncountable hours and spend about 2 years trying to develop a fluorescence in situ hybridization method (FISH) to bring to light those unknown species. Unfortunately this method never succeeded with freshwater sediment, either because no living specimens were present in the samples or their cells were not visible through agglutinated tests. However, the FISH method worked well on marine species, proving that the method was effective and could be used in the future.

5.2. Abstract

Freshwater foraminifera are the most mysterious part of this highly diverse group of mainly marine rhizarian protists. All described freshwater foraminiferal species belong to monothalamids, a group of single-chambered foraminifera having organic or agglutinated wall. Several of these species have been reported for the first time from Geneva basin by Swiss protistologists at the turn of the XIX and XX centuries. However, no microscopic observations on this group have been conducted since then and the only evidence of their presence there nowadays are few environmental DNA sequences obtained from Geneva lake sediments. Here, we present the results of an extensive eDNA metabarcoding study targeting freshwater foraminifera in various water bodies in Geneva Basin, conducted in 2014-2016. Our study reveals the presence of foraminiferal rDNA sequences in almost all sampling sites and at different seasons, suggesting that the group is much more abundant and diverse than was previously thought. We identified 48 foraminiferal OTUs branching within 5 clades comprising only freshwater species. Phylogenetic analyses suggest that foraminifera colonized freshwater habitats several times during their evolution, becoming an important albeit largely unrecognized component of freshwater protist community.

5.3. Introduction

The first foraminifera have been described from freshwater habitats more than a century ago. The first freshwater foraminiferal species was *Lieberkuhnia sp* (Claparède & Lachmann 1859) described in 1859 from an unnamed water body in Berlin. In subsequent years, Henri Blanc and Eugène Penard, both Swiss protistologists, described 6 different foraminiferal species from the Geneva basin. All these species have been placed in the genus *Gromia* considered to belong to foraminifera at that epoch.

Since then, only few new freshwater foraminiferal species have been sporadically described. Nauss (1949) described a large naked plasmodial protist *Reticulomyxa filosa*, later identified as a foraminiferan (Pawlowski *et al.* 1999). In 1961, Thomas described *Penardogromia palustris* from freshwater near Bordeaux, France, while Meisterfeld and colleagues (2001) described the first organic-walled soil foraminiferal species, *Edaphoallogromia australica*, from tropical forest in Queensland. More

recently, three new Reticulomyxidae: *Haplomyxa saranae* (Dellinger *et al.* 2014), *Dracomyxa pallida* (Wylezich *et al.* 2014), *Wobo gigas* (Hülsmann 1986, 2006, http://www.arcella.nl/wobo) have been described from laboratory cultures. The description of two other agglutinated foraminiferal species from freshwater bodies is in preparation (Siemensma *et al*. submitted – see Chapter 4).

At the same time, the studies of environmental DNA (eDNA) unveil the presence of foraminiferal sequences in sediments of Swiss lakes (Holzmann & Pawlowski 2002; Holzmann *et al.* 2003) and soil habitats (Lejzerowicz *et al.* 2010; Geisen *et al.* 2015). The phylogenetic analyses of these sequences reveal four clades of freshwater foraminifera (Clade FW1-4). One of them is associated to morphologically described family Reticulomyxidae. The other clades remained morphologically undocumented until the recent description of two new agglutinated freshwater species *Lacogromia cassipara* and *Limnogromia sinensis* and the sequencing of *Lieberkuenia sp*. (Siemensma *et al*. submitted – Chapter 4). These new descriptions contribute to characterize morphologically the major freshwater foraminiferal groups, which diversity, however, remains poorly explored.

In this study, we attempt to fill this gap by extensively screening the eDNA samples from various sites in historical location of freshwater foraminifera descriptions, the Geneva basin. We performed 93 samplings in 46 locations during different periods of the year. We collected surface sediment samples from rivers, streams, lakes and ponds as well as some soil samples. We succeeded to amplify and sequence foraminiferal barcoding region of 18S rDNA for 56 samples representing 29 different locations. The resulting sequences were clustered into 48 Operational Taxonomic Units (OTUs) at 98% of identity. We investigated the phylogenetic position of the obtained OTUs and their relationships to other monothalamous clades by analysing the complete 18S rDNA sequences. Additionally, we performed a high-throughput sequencing (HTS) survey on 68 sediment and 43 biofilm samples representing 67 locations to investigate the distribution of freshwater foraminifera in the Geneva basin.

5.4. Materials and methods

5.4.1. Sampling

46 sampling spots (river, lake, pond and some soil sample) from all over the Geneva basin were sampled at different periods of the year (TableS 5.1) between 2010 and 2013. About 0.25g of surface sediment or soil sample were collected in a sterile tube and kept at 4°C until processing in the lab (between 0 and 2 days). 4 locations were investigated for DNA and RNA every two months during one year. For those samples, about 15g of surface sediment was collected and frozen at -80°C until processing. For each sample, 3 replicates were performed.

5.4.2. DNA/RNA extraction, PCR amplification and Sanger sequencing

In total, 314 extractions were performed using either DNeasy PowerSoil or RNeasy PowerSoil Total RNA Kits from Qiagen following the manufacturer's instructions. Following each RNA extraction, the DNA was eluted using the RNeasy PowerSoil DNA Elution Kit (Qiagen). cDNA was then synthetised using the iScript Select cDNA synthesis Kit (BioRad, Hercules, CA) with random primers. PCR amplifications of a partial fragment of the 18S rDNA were done using the primer pair s14F3 and sB. PCR products were re-amplified using the nested primer s14F1 (see TableS 5.2 for primer sequences).

Some samples were also amplified in order to obtain the complete 18S rDNA sequences. This was done with two more PCR steps. The second fragment was amplified using the primer pair 6F and 17. PCR products were re-amplified using the nested primer 15A. The third fragment was amplified using the primer pair A10 and 12R. PCR products were re-amplified using the nested primer 7R. The sequenced fragments have been assembled to retrieve the complete 18S rRNA gene.

The amplified PCR products were purified using High Pure PCR Purification Kit (Roche Diagnostics) and cloned with the TOPO10 kit from Invitrogen (Thermo Fisher Scientific, Waltham, MA). Sequencing reactions were performed using the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) and analysed on a 3130XL Genetic Analyser (Applied Biosystems).

5.4.3. Clustering, sequence alignment and phylogenetic analysis

The obtained sequences of the partial SSU rDNA were clustered at 98% using mothur (Schloss *et al.* 2009) and the resulting OTUs were aligned to the local database using Seaview Software (Gouy *et al.* 2010). After removing the hypervariable region, a selection of 1061 sites on a total alignment of 2478 was used to build a ML phylogeny using RAxML v.7.4.2 (Stamatakis 2014) with GTR + G as model of evolution and 100 replicates for the bootstrap analysis.

The complete 18S rDNA sequences were aligned to the local database of foraminifera using the same software as mentioned above. After removal of hypervariable regions, a selection of 2253 sites on a total alignment of 6723 was used to build a ML phylogeny following the same conditions as for the short fragment.

5.4.4. PCR amplification, HTS sequencing and bioinformatics

Samples used for the HTS analysis were amplified by PCR using the primer pair s14F3 and 17 and PCR products were re-amplified using the tagged nested primer s14F1 and 17. Individual tags are composed of 8 nucleotides attached at each primer 5'- extremities and each sample is tagged with a unique combination of forward and reverse primers allowing the multiplexing of all the samples in one library. For each sample, PCR replicates were performed for each extraction in order to have 6 replicates with the same combination of tags that were pooled. PCR products were purified with Sephadex G-50 superfine resin (GE Healthcare) and quantified using QuBit HS dsDNA (Invitrogen). The same amount of each sample was pooled and a final purification step was performed with High Pure PCR Product Purification kit (Roche Applied Science). The library was prepared with Illumina TruSeq® PCR free Preparation Kit. Final library was quantified with qPCR using KAPA Library Quantification Kit and sequenced on a MiSeq instrument using paired-end sequencing for 500 cycles with Nano kit v2.

The foraminiferal OTUs were obtained following the method described in Pawlowski *et al.* (2014a). After quality filtering and assembly steps, the run from this study was combined to the foraminifera run on biofilm sample of the Chapter 9. Then, de-replication was performed in order to obtain Individual Sequence Units (ISUs). An abundance threshold of 10 was used for the minimum number of reads required for

each ISU (Bokulich *et al.* 2013). They were then grouped at 98% using complete-linkage clustering method. Finally, we removed chimeric sequences found with manual inspection of Uchime (Edgar *et al.* 2011) candidates. OTUs were assigned to the first hit using nBLAST (Altschul *et al.* 1990) against a local database containing all monothalamous foraminifera sequences as well as the Sanger sequences from this study with at least 95% of identity. All sequences blasting with marine clade were checked manually with phylogeny to correct a possible wrong assignation to a marine clade. Computer analyses where performed using R (R Core Team 2013). For the repartition of the clade, only sites with more than 1000 reads were kept. The non-metric multidimensional scaling plot was performed using vegan package (Oksanen *et al.* 2013).

5.5. Results and discussion

5.5.1. Phylogeny

A short fragment of the 18S rDNA, used as foraminiferal barcode (Pawlowski & Holzmann 2014), was amplified in 56 out of the 93 sites sampled in Geneva area. Cloning and sequencing of PCR products led to obtaining 269 sequences that were clustered in 48 OTUs at 98% of identity. To examine where these sequences branched among other freshwater foraminifera, we built a ML phylogeny with 103 sequences including 72 environmental foraminiferal sequences from previous studies (Holzmann *et al.* 2003; Pawlowski *et al.* 2011b) and 10 sequences from described freshwater foraminifera (Figure 5.1). The tree was rooted on marine clades A, B and C. Those clades branch as a sister groups to multi-chambered globothalamous foraminifera (Pawlowski *et al.* 2013).

Almost all obtained OTUs branched in the four previously described freshwater clades (Lejzerowicz *et al.* 2010), called here FW1-4. The *Reticulomyxa*-bearing clade FW2 is strongly supported with a bootstrap value (BV) of 100 while the clades FW3 and FW4 show a moderate support (77BV and 61BV respectively). The Clade FW1 did not show a significant support. Some OTUs branch separately: 2 OTUs branch with *E. australica* in the clade M (comprising the marine genus *Allogromia*) and one OTU (OTU41) branches close to the marine clade E, represented by *Psammophaga spp*. In both cases, the relationships of freshwater OTUs are well supported (81BV

and 99BV respectively). Finally, 2 OTUs are assembled in a new clade (100BV) that branches separately, between Clades A, B, C and the rest of monothalamids. We have tried ML phylogeny with several site selections and those two OTUs always branch together but their position in the tree is variable. We propose to call this new clade FW5 to be consistent with the nomenclature of other freshwater groups.

The OTU41 that branches with the marine clade E as well as the two OTUs forming the new clade FW5 were inspected manually to check whether they are not chimeras or result of a contamination. However, the 3 OTUs were very different from all the sequences that were previously amplified in the lab according to our local database. Therefore we decided to keep them as valuable data. Moreover, we built a phylogeny of the OTU41 with all known sequences from the clade E (FigureS 5.1 A). This analysis shows that the OTU41 branches indeed inside the clade E and close to the sequence of an undescribed “chocolate silver saccamminid” from Habura *et al.* (2008) isolated from a salt marsh environment in Georgia.

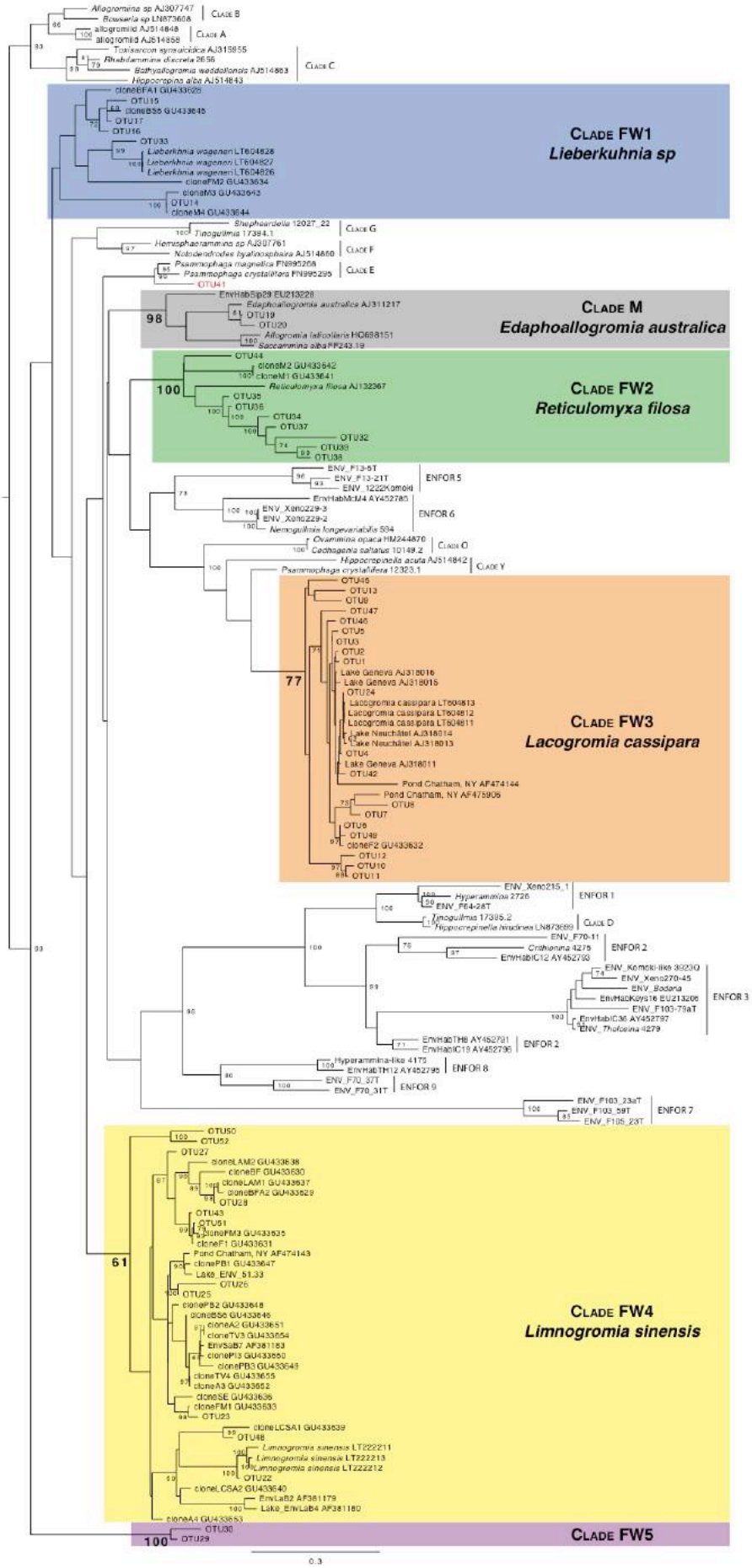

CLADE FW1
Lieberkuhnia sp
CLADE M
Edaphoallogromia australica
CLADE FW2
Reticulomyxa filosa
CLADE FW3
Lacogromia cassipara
CLADE FW4
Limnogromia sinensis
CLADE FW5
ENFOR 1
ENFOR 2
ENFOR 3
ENFOR 5
ENFOR 6
ENFOR 7
ENFOR 8
ENFOR 9
CLADE A
CLADE B
CLADE C
CLADE D
CLADE E
CLADE F
CLADE G
CLADE O
CLADE Y
0.3

Figure 5.1 Maximum likelihood phylogenetic tree of the 48 OTUs aligned with 103 sequences of monothalamous foraminifera including 29 environmental sequences from marine habitats and 40 from freshwater or soil habitats. The tree is based on the partial 18S rDNA sequences. The formally described freshwater species: *Liberkuhnia sp, E. australica, R. filosa, L. cassipara* and *L. sinensis* are indicated in each representative clade. Support values are RAxML bootstraps. Only values superior to 60 are shown.

In parallel, we obtained complete 18S rDNA sequences for 14 isolates from 12 sampling stations. The 14 sequences correspond to 8 OTUs and were aligned to 33 foraminiferal sequences including 25 monothalamids and a ML phylogeny was built (Figure 5.2). The tree was rooted on clade I as in Pawlowski *et al.* (2013). Unfortunately we were not able to amplify complete 18S sequences from the clades FW2 and FW4, as those two groups show large A and T insertions in their hypervariable regions, which renders sequencing very difficult. The clade FW3 branches as sister to clades O and Y with a very high support (100BV), while the clade FW2 branches between them and clade G. The clade FW1 branches between crown groups A+B+C + BM and clade E, however there is no support for this topology. Nevertheless, the monophyly of the clades FW1 and FW3 is strongly support with 100BV.

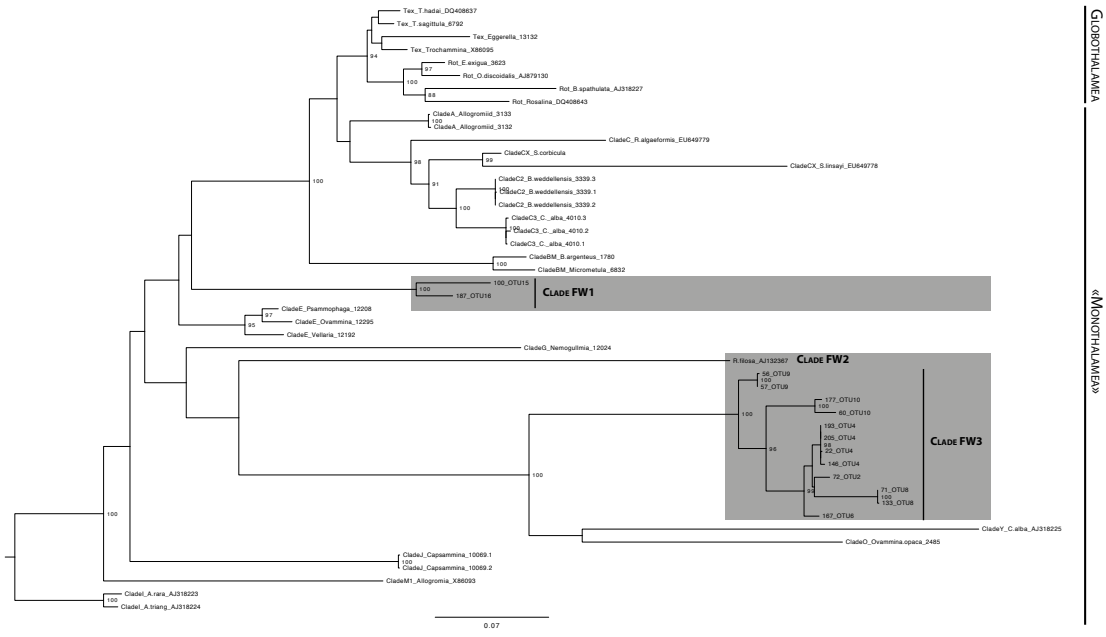


Figure 5.2 Phylogenetic tree of 48 sequences of foraminifera, including 15 from environmental freshwater samples, based on complete 18S rDNA.

5.5.2. Diversity and Ecology

In addition to phylogenetic study, we also analysed the diversity of freshwater foraminifera at spatial scale. We sampled a large area of the Geneva basin, including 68 sediment samples from 27 different location sites and 61 epilithic biofilm samples from different locations, among which 17 have been sampled twice in a year. 55% of these locations (27/62 locations in sediment and 40/61 locations in biofilm) were tested positive for foraminiferal DNA (FigureS 5.2 and FigureS 5.3), suggesting that the foraminifera are present in the entire basin.

In order to investigate the distribution of the freshwater foraminifera, we sequenced all positives samples using a HTS approach. The number of reads for the two Illumina runs as well as the filtering process are summarised in the TableS 5.3. In total, 6'635'241 sequences representing 15'978 ISUs were clustered at 98% into 1653 OTUs. After removal of chimeric sequences, obvious contamination (99-100% sequence identity with a marine sequence from our local database) and the ones without blast hit 1045 OTUs remained and were used for analyses. 98% (1020 OTUs) matched to the freshwater clades, 1% (12 OTUs) to marine monothalamous clades E and M, and for 1% (13 OTUs) no assignments to known clades were found. The 2 OTUs belonging to clade M branch with *E. australica*, while the 10 OTUs branching within clade E are close to the freshwater OTU41 (FigureS 5.1 B).

The four freshwater clades (FW1-4) are represented more or less equally (between 129 OTUs for FW2 and 313 for FW3). The clade FW5 is represented by only 17 OTUs. As previously said, the clade FW2 was difficult to sequence due to large AT insertions, which are probably the cause of a limited sequencing success and thus a relatively lower number of OTUs found in this clade. In Figure 5.3, the relative frequency of each clade is represented per site. The clades FW1, FW3 and FW4 are found in all types of samples with clade FW3 (in orange) representing more than the half of the dataset (Figure 5.3 B). The clade FW1 can be very abundant and unique representative of freshwater foraminifera in one soil sample, two small river, two medium river and two big river samples. It is also present in mots of big river samples. The clade FW4 is mainly present in small and medium rivers, but also dominate in two soil samples. The clade FW2 seems to be more present in still water, like ponds or lakes, and rivers with reduced flow velocity. The clade FW5 is very rare and is present in more than 20% in three samples only, collected from soil,

pond and biofilm. The distribution and localisation of the reads assigned to the clade FW5 are shown in the FigureS 5.4 and the description of all freshwater clades are summarised in the TableS 5.4.

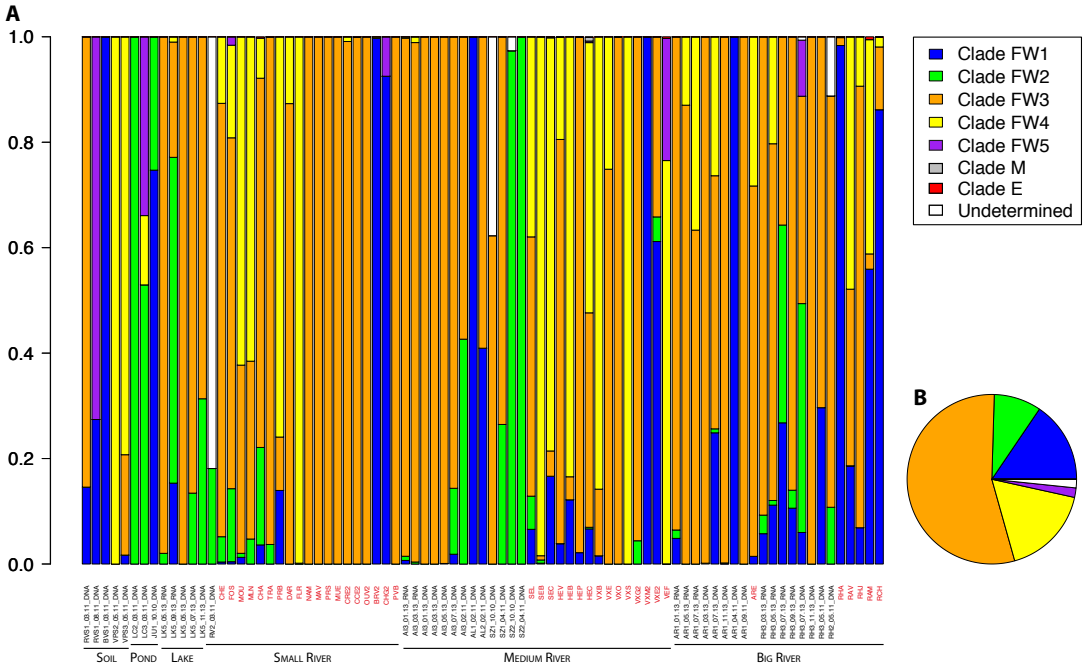


Figure 5.3 A. Relative frequencies of each freshwater foraminifera clades per site. Sites are classified in function of their ecosystem. Biofilm samples are indicated in red. B. Sum of the relative frequency of each clade for the entire dataset.

We investigated also the repartition of the communities across the samples using non-metric multidimensional scaling (Figure 5.4). We used 4 localities that were sampled every two months during one year. Those samples are coming from the lake Geneva, the two biggest rivers in Geneva (Rhône and Arve) and a smaller river (Aire). The foraminiferal communities in Lake Geneva and the Aire show small variations during the year and are different from other localities. On the other hand, the Rhône and the Arve rivers seem to share similar community assemblages, especially during spring. However, an important variation in the assemblage of each of those rivers is observed during the year.

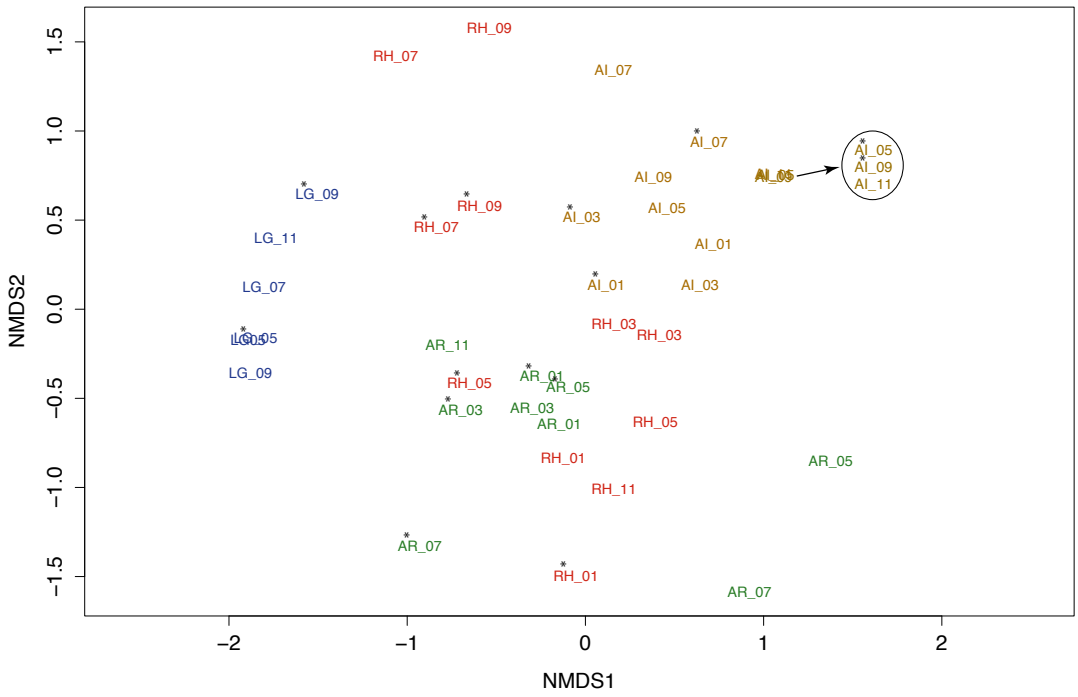


Figure 5.4 Non-metric multidimensional scaling of the four monitored localities. The analysis is based on the normalised abundance matrix of the OTUs and the distances were computed using Bray-Curtis dissimilarity. The lake Geneva (LG) is marked in blue, the Rhône (RH) in red, the Arve (AR) in green and the Aire (AI) in gold. The number written next to each sample corresponds to the month of sampling. RNA samples are marked with a star (*).

5.6. Conclusions

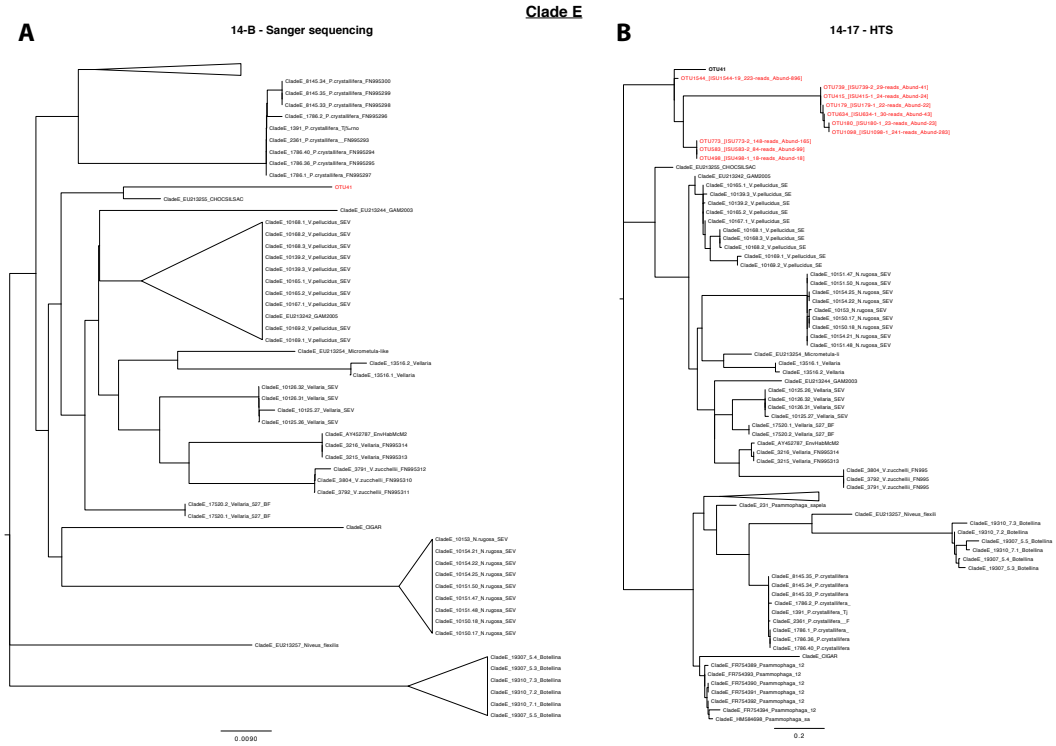
This study confirmed that freshwater foraminifera are widely distributed and polyphyletic. The adaptation from marine to freshwater environment seems to occur several times during the evolution of foraminifera. Although most of the OTUs form uniquely freshwater clades, there two clades (M and E) that comprise both freshwater/soil and marine OTUs. The case of *E. australica*, isolated from Australian rain forest soil was confirmed by the presence of closely related OTUs isolated from Geneva basin. These species as well as those branching within the marine clade E may represent recent transitions from marine to freshwater habitats.

In total, our study revealed 48 OTUs of freshwater foraminifera by cloning and Sanger sequencing, and 1045 in the HTS analysis. Even if some of the later could be due to a high level of polymorphism in the sequenced taxa, still the diversity of

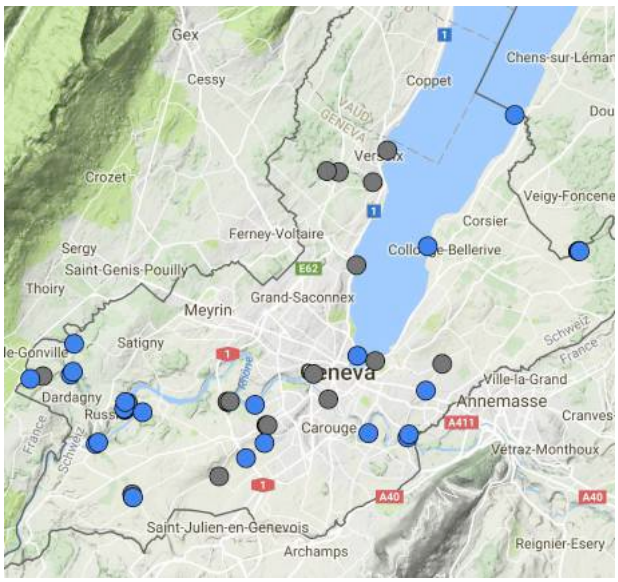
freshwater foraminifera largely exceeds the number of seven species formerly described from the Geneva basin. It is expected that the OTUs identified here represent only the peak of a largely hidden diversity of freshwater foraminifera and that further studies of other areas will contribute to unravel the true diversity of this group, not only at OTU level but also at higher taxonomic level as illustrated by the finding of a new freshwater clade FW5.

5.7. Supplementary data

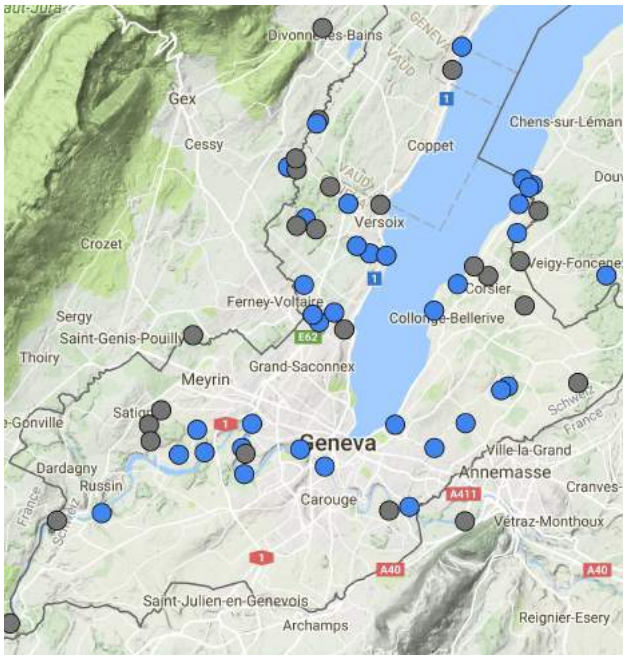
FigureS 5.1 Maximum likelihood tree of the clade E based on the 14-B fragment (about 1000bp) with the OTU41 (A). Based on the 14-17 fragment (about 300bp) with the sequences assigned to Clade E in the HTS analysis (B).



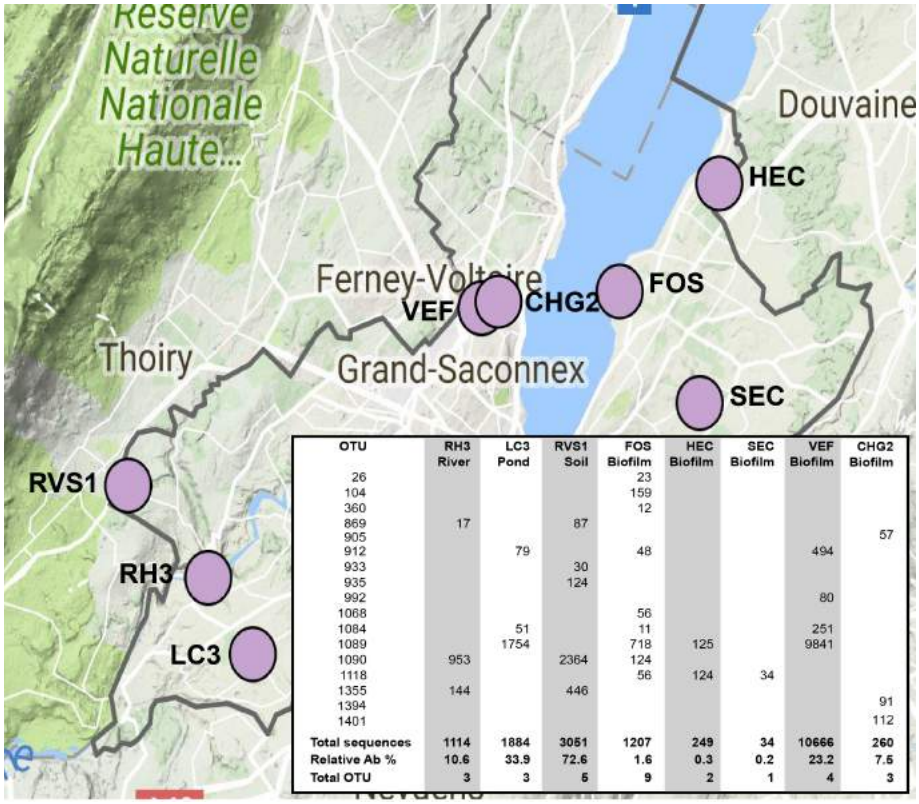
FigureS 5.2 Sampling map of the surface sediment samples. The locations in blue correspond to the locations where the DNA amplification of foraminifera succeeds.



FigureS 5.3 Sampling map of the biofilm samples. The locations in blue correspond to the locations where the DNA amplification of foraminifera succeeds.

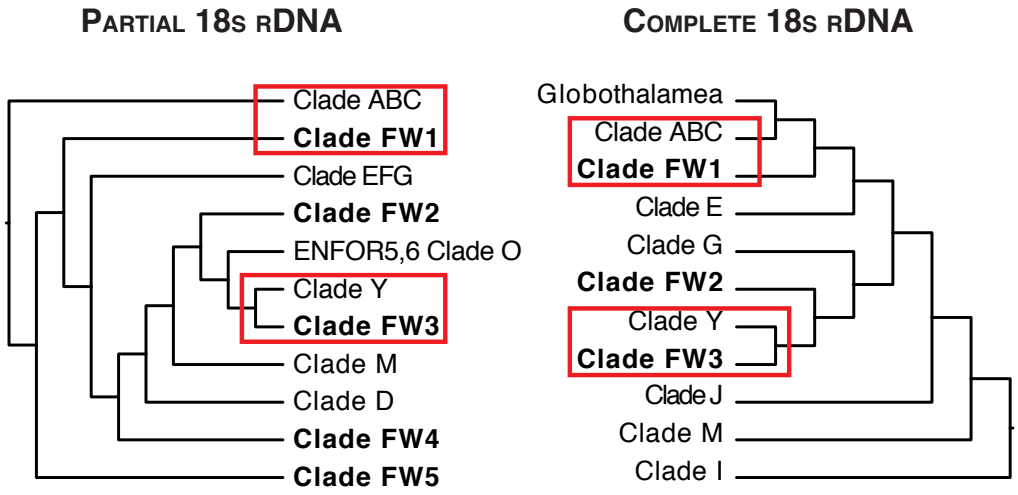


FigureS 5.4 Repartition and localisation of the reads assigned to the clade FW5 in the HTS dataset



OTU	RH3 River	LC3 Pond	RVS1 Soil	FOS Biofilm	HEC Biofilm	SEC Biofilm	VEF Biofilm	CHG2 Biofilm
26				23				
104				159				
360				12				
869	17		87					
905								57
912		79		48			494	
933			30					
935			124					
992							80	
1068				56				
1084		51		11			251	
1089		1754		718	125		9841	
1090	953		2364	124				
1118				56	124	34		
1355	144		446					
1394								91
1401								112
Total sequences	**1114**	**1884**	**3051**	**1207**	**249**	**34**	**10666**	**260**
Relative Ab %	**10.6**	**33.9**	**72.6**	**1.6**	**0.3**	**0.2**	**23.2**	**7.5**
Total OTU	**3**	**3**	**5**	**9**	**2**	**1**	**4**	**3**

FigureS 5.5 Simplified topologies of the two phylogenetic trees



TableS 5.1 List of sample with GPS coordinate and sampling date. The presence of the partial and/or complete 18S sequence and their used in the HTS analysis was also indicated for each sample.

Name	Locality	Coordinates	sample type	Date mm yy	type	partial 18S	18S	HTS
BM1	Bois des mouilles	46°11'34"N 6°04'54"E	sediment	10 10	DNA			
BM2	Bois des mouilles	46°11'35"N 6°04'54"E	sediment	10 10	DNA			
BM3	Bois des mouilles	46°11'35"N 6°04'59"E	sediment	10 10	DNA			
LC1	Etang de Laconnex	46°09'26"N 6°01'43"E	sediment	10 10	DNA			
LC2	Etang de Laconnex	46°09'22"N 6°01'45"E	sediment	10 10	DNA			
LC3	Etang de Laconnex	46°09'22"N 6°01'44"E	sediment	10 10	DNA	+		+
PB1	Parc Brot	46°11'02"N 6°06'13"E	sediment	10 10	DNA			
PB2	Parc Brot	46°11'02"N 6°06'12"E	sediment	10 10	DNA			
PB3	Parc Brot	46°11'02"N 6°06'14"E	sediment	10 10	DNA			
SZ1	Seymaz	46°10'49"N 6°10'57"E	sediment	10 10	DNA	+	+	+
SZ2	Seymaz	46°11'50"N 6°11'31"E	sediment	10 10	DNA	+		+
SZ3	Seymaz	46°12'27"N 6°12'04"E	sediment	10 10	DNA			
JU1	Etang de Jussy	46°15'03"N 6°16'39"E	sediment	10 10	DNA	+		+
JU2	Etang de Jussy	46°15'04"N 6°16'38"E	sediment	10 10	DNA			
JU3	Etang de Jussy	46°15'03"N 6°16'36"E	sediment	10 10	DNA	+		+
BM4	Bois des mouilles	46°11'34"N 6°04'54"E	sediment	11 10	DNA			
BM5	Bois des mouilles	46°11'35"N 6°04'54"E	sediment	11 10	DNA			
BM6	Bois des mouilles	46°11'35"N 6°04'59"E	sediment	11 10	DNA			
AL1	Allondon	46°12'55"N 5°59'47"E	sediment	02 11	DNA	+	+	+
AL2	Allondon	46°12'16"N 5°59'44"E	sediment	02 11	DNA	+	+	+
AL3	Allondon	46°10'38"N 6°00'35"E	sediment	02 11	DNA	+		+
AI3	L'Aire	46°09'51"N 6°04'37"E	sediment	02 11	DNA			

AI2	L'Aire	46°10'16"N 6°05'31"E	sediment	02 11	DNA	+		+
AI3	L'Aire	46°10'37"N 6°06'08"E	sediment	02 11	DNA	+	+	+
BM7	Bois des mouilles	46°11'34"N 6°04'54"E	sediment	03 11	DNA			
BM8	Bois des mouilles	46°11'35"N 6°04'54"E	sediment	03 11	DNA			
BM9	Bois des mouilles	46°11'35"N 6°04'59"E	sediment	03 11	DNA			
LC4	Etang de Laconnex	46°09'26"N 6°01'43"E	sediment	03 11	DNA			
LC5	Etang de Laconnex	46°09'22"N 6°01'45"E	sediment	03 11	DNA	+		+
LC6	Etang de Laconnex	46°09'22"N 6°01'44"E	sediment	03 11	DNA	+		+
VX1	Versoix	46°16'54"N 6°08'13"E	sediment	03 11	DNA			
VX2	Versoix	46°16'53"N 6°0837"E	sediment	03 11	DNA			
VX3	Versoix	46°16'39"N 6°09'45"E	sediment	03 11	DNA			
BVS1	Praire-Barrage	46°11'34"N 6°01'29"E	Soil	03 11	DNA	+	+	+
BVS2	Praire-Barrage	46°11'31"N 6°01'30"E	Soil	03 11	DNA			
BVS3	Praire-Barrage	46°11'33"N 6°01'34"E	Soil	03 11	DNA			
RV1	Roulavaz	46°12'06"N 5°58'19"E	sediment	03 11	DNA			
RV2	Roulavaz	46°12'10"N 5°58'44"E	sediment	03 11	DNA	+		+
RV3	Roulavaz	46°12'12"N 5°59'40"E	sediment	03 11	DNA	+		
RVS1	Roulavaz	46°12'06"N 5°58'19"E	Soil	03 11	DNA	+		+
RVS2	Roulavaz	46°12'10"N 5°58'44"E	Soil	03 11	DNA			
RVS3	Roulavaz	46°12'12"N 5°59'40"E	Soil	03 11	DNA			
SZ4	Seymaz	46°10'49"N 6°10'57"E	sediment	04 11	DNA	+	+	+
SZ5	Seymaz	46°11'50"N 6°11'31"E	sediment	04 11	DNA	+		+
SZ6	Seymaz	46°12'27"N 6°12'04"E	sediment	04 11	DNA			
AR1	L'Arve	46°10'45"N 6°10'54"E	sediment	04 11	DNA	+		+
AR2	L'Arve	46°10'51"N 6°09'36"E	sediment	04 11	DNA	+	+	+
AR3	L'Arve	46°11'38"N 6°08'16"E	sediment	04 11	DNA			
VPS1	Barrage	46°11'23"N 6°01'28"E	Soil	05 11	DNA			
VPS2	Barrage	46°11'25"N 6°01'27"E	Soil	05 11	DNA	+		+
VPS3	Barrage	46°11'24"N 6°01'27"E	Soil	05 11	DNA	+		+
PIS1	Aire-la-Ville	46°11'20"N 6°02'04"E	Soil	05 11	DNA	+		+
PIS2	Aire-la-Ville	46°11'20"N 6°02'04"E	Soil	05 11	DNA			
PIS3	Aire-la-Ville	46°11'20"N 6°02'04"E	Soil	05 11	DNA	+	+	+
RH1	Rhône (Jonction)	46°12'13"N 6°07'46"E	sediment	05 11	DNA			
RH2	Rhône (Onex)	46°11'30"N 6°05'49"E	sediment	05 11	DNA	+	+	+
RH3	Rhône (Allondon)	46°10'36"N 6°00'30"E	sediment	05 11	DNA	+		+
LK1	Lac (Versoix)	46°17'23"N 6°10'14"E	sediment	06 11	DNA			
LK2	Lac (Vengeron)	46°14'44"N 6°09'12"E	sediment	06 11	DNA			
LK3	Lac (Paquis)	46°12'38"N 6°09'14"E	sediment	06 11	DNA	+	+	+
LK4	Lac (baby-plage)	46°12'31"N 6°09'50"E	sediment	06 11	DNA			
LK5	Lac (Bellerive)	46°15'10"N 6°11'35"E	sediment	06 11	DNA	+	+	+
LK6	Lac (Hermance)	46°18'12"N 6°14'29"E	sediment	06 11	DNA	+		
RVS1	Roulavaz	46°12'06"N 5°58'19"E	Soil	08 11	DNA	+		+
AR1	L'Arve	46°10'45"N 6°10'54"E	sediment	09 11	DNA	+	+	+
AR1	L'Arve	46°10'45"N 6°10'54"E	sediment	10 11	DNA	+		

AI3	L'Aire	46°10'37"N 6°06'08"E	sediment	10 11	DNA	+	
SZ1	Seymaz	46°10'49"N 6°10'57"E	sediment	10 11	DNA	+	
SZ1	Seymaz	46°10'49"N 6°10'57"E	sediment	02 12	DNA	+	
AR1	L'Arve	46°10'45"N 6°10'54"E	sediment	06 12	DNA	+	
AI3	L'Aire	46°10'37"N 6°06'08"E	sediment	08 12	DNA	+	+
AR1	L'Arve	46°10'45"N 6°10'54"E	sediment	01 13	RNA	+	+
RH3	Rhône (Allondon)	46°10'36"N 6°00'30"E	sediment	01 13	RNA	+	+
AI3	L'Aire	46°10'37"N 6°06'08"E	sediment	01 13	RNA	+	+
AR1	L'Arve	46°10'45"N 6°10'54"E	sediment	01 13	DNA	+	+
RH3	Rhône (Allondon)	46°10'36"N 6°00'30"E	sediment	01 13	DNA	+	+
AI3	L'Aire	46°10'37"N 6°06'08"E	sediment	01 13	DNA	+	+
LK5	Lac (Bellerive)	46°15'10"N 6°11'35"E	sediment	03 13	RNA		+
AR1	L'Arve	46°10'45"N 6°10'54"E	sediment	03 13	RNA	+	+
RH3	Rhône (Allondon)	46°10'36"N 6°00'30"E	sediment	03 13	RNA	+	+
AI3	L'Aire	46°10'37"N 6°06'08"E	sediment	03 13	RNA	+	+
LK5	Lac (Bellerive)	46°15'10"N 6°11'35"E	sediment	03 13	DNA	+	+
AR1	L'Arve	46°10'45"N 6°10'54"E	sediment	03 13	DNA	+	+
RH3	Rhône (Allondon)	46°10'36"N 6°00'30"E	sediment	03 13	DNA	+	+
AI3	L'Aire	46°10'37"N 6°06'08"E	sediment	03 13	DNA	+	+
LK5	Lac (Bellerive)	46°15'10"N 6°11'35"E	sediment	06 13	RNA		+
AR1	L'Arve	46°10'45"N 6°10'54"E	sediment	06 13	RNA	+	+
RH3	Rhône (Allondon)	46°10'36"N 6°00'30"E	sediment	06 13	RNA		+
AI3	L'Aire	46°10'37"N 6°06'08"E	sediment	06 13	RNA	+	+
LK5	Lac (Bellerive)	46°15'10"N 6°11'35"E	sediment	06 13	DNA	+	+
AR1	L'Arve	46°10'45"N 6°10'54"E	sediment	06 13	DNA	+	+
RH3	Rhône (Allondon)	46°10'36"N 6°00'30"E	sediment	06 13	DNA	+	+
AI3	L'Aire	46°10'37"N 6°06'08"E	sediment	06 13	DNA	+	+
RH3	Rhône (Allondon)	46°10'36"N 6°00'30"E	sediment	07 13	RNA		+
AI3	L'Aire	46°10'37"N 6°06'08"E	sediment	0713	RNA		+
LK5	Lac (Bellerive)	46°15'10"N 6°11'35"E	sediment	07 13	DNA		+
AR1	L'Arve	46°10'45"N 6°10'54"E	sediment	07 13	DNA		+
RH3	Rhône (Allondon)	46°10'36"N 6°00'30"E	sediment	07 13	DNA		+
AI3	L'Aire	46°10'37"N 6°06'08"E	sediment	07 13	DNA		+
LK5	Lac (Bellerive)	46°15'10"N 6°11'35"E	sediment	09 13	RNA		+
RH3	Rhône (Allondon)	46°10'36"N 6°00'30"E	sediment	09 13	RNA		+
AI3	L'Aire	46°10'37"N 6°06'08"E	sediment	09 13	RNA		+
LK5	Lac (Bellerive)	46°15'10"N 6°11'35"E	sediment	09 13	DNA		+
RH3	Rhône (Allondon)	46°10'36"N 6°00'30"E	sediment	09 13	DNA		+
AI3	L'Aire	46°10'37"N 6°06'08"E	sediment	09 13	DNA		+
LK5	Lac (Bellerive)	46°15'10"N 6°11'35"E	sediment	11 13	DNA		+
AR1	L'Arve	46°10'45"N 6°10'54"E	sediment	11 13	DNA		+
RH3	Rhône (Allondon)	46°10'36"N 6°00'30"E	sediment	11 13	DNA		+
AI3	L'Aire	46°10'37"N 6°06'08"E	sediment	11 13	DNA		+

TableS 5.2 List of primer used in this study

Primer	Sequence 5'-3'	Direction
A10	CTC AAA GAT TAA GCC ATG CAA GTG G	for
6F	CCG CGG TAA TAC CAG CTC	for
7R	CTG RTT TGT TCA CAG TRT TG	rev
12R	GKT AGT CTT RMH AGG GTC A	rev
14F1	AAG GGC ACC ACA AGA ACG C	for
14F3	ACG CAM GTG TGA AAC TTG	for
15A	CTA AGA ACG GCC ATG CAC CAC C	rev
17	CGG TCA CGT TCG TTG C	rev
B	TGA TCC TTC TGC AGG TTC ACC TAC	rev

TableS 5.3 Filtering process on the Sediment library and the two biofilm librairies

Statistics parameter	Sediment	Biofilm 1	Biofilm 2	Total
Total number of reads	1149796	4329912	3935450	
Reject ambiguous forward	0	0	0	
Reject ambiguous reverse	0	0	0	
Low mean quality forward	58462	170385	264605	
Low mean quality reverse	87595	241854	351466	
Low mean quality contig	0	0	0	
Low base quality contig	50273	322940	323360	
Not enough matching contig	2712	22173	34800	
No primers forward	19264	139069	119482	
No primers reverse	16179	103908	85469	
Mismatch found in primers	342398	12378	11127	
Insufficient sequence length (dimers)	4	7	7	
Total number of good reads	572909	3317198	2745134	6635241
Number of ISUs				15978
Number of OTUs 98%				1653
Number of OTUs without chimera				1257

TableS 5.4 Summary of the five freshwater clades including their representative species with their morphology, the phylogenetic support of the clade with the number of OTUs for the barcode fragment (14-B) found in this study. The table also includes the percentage of sequences identity for the small fragment (% ID 14-17), the preferred habitats and the abundance through the entire dataset

Clade	**Representative species**	**morphology**	**monophyly**	**OTUs 14-B**	**% ID 14-17**	**habitats**	**abundance**
FW1	*Lieberkühnia sp*	organic walled	not well supported	5	88%	all	++
FW2	*Reticulomyxa filosa*	naked	well defined	8	94%	still waters	+
FW3	*Lacogromia cassipara*	agglutinated	moderately supported	19	89%	all	+++
FW4	*Limnogromia sinensis*	agglutinated	moderately supported	11	84%	mostly soil and biofilm	++
FW5	-	-	well defined	2	97%	mostly soil and pond	rare

CHAPTER 6

SINGLE CELL HIGH-THROUGHPUT SEQUENCING UNVEILS DIFFERENT PATTERNS OF INTRAGENOMIC POLYMORPHISM IN RIBOSOMAL RNA GENES OF FORAMINIFERA

Project in progress

6.1. Project description

This project was designed by Jan Pawlowski to follow up on a previous project on intragenomic polymorphism conducted by my master student Alexandra Weber. Maria Holzmann and Ivan Voltsky were first involved in this project and they performed all the lab work to obtain the HTS single-cell data. I was then enrolled to analyse the sequences. This project is not yet over and we are planning another Illumina run with other species targeted to answer some specific questions related to ecology and biogeography. At the end, we are planning to divide the dataset in two parts, with the *Ammonia* species analysed separately from the other foraminifera.

6.2. Abstract

Nowadays, the assessment of microbial eukaryotes diversity is commonly done using metabarcoding approach based on high-throughput amplicon sequencing. In metabarcoding studies, the species or Operational Taxonomic Units (OTUs) are usually distinguished based on more or less fixed thresholds that define the level of intraspecific variations. However, metabarcoding analyses rarely take into account the intragenomic variations, often considered as result of technical errors. Here, we use single-cell high-throughput sequencing approach to evaluate the level of intragenomic polymorphism (IGP) in 18S rRNA gene of 130 specimens of benthic foraminifera, representing different taxa and living in different habitats. Our study confirms previously shown widespread occurrence of IGP in foraminifera. We report different patterns of IGP, including the single-nucleotide polymorphisms (SNP) and expansion segments polymorphisms (ESP), resulting in occurrence of numerous haplotypes, which divergence may reach up to 5%. Interestingly, while SNPs are present in all examined species, the ESPs are found only in multi-chambered calcareous species that generally show much higher IGP level than single-chambered species. We also observe a significant difference of IGP level between shallow-water coastal and deep-sea taxa; the later showing surprisingly low level of intra-individual sequence divergence. Although the origins of the IGP patterns are difficult to explain, we found some evidences that they may represent biological variations rather than technical errors. The high level of IGP in some taxonomic groups could be related to sexual reproduction, hybridization or genomic recombination. As shown by our study, the IGP analysis may provide important information about population composition and structure. Moreover, single-cell HTS testing of IGP level may help avoiding overestimation of environmental diversity in metabarcoding studies.

6.3. Introduction

In recent years, the HTS-based metabarcoding became a standard procedure to evaluate the level of environmental diversity in different groups of microbial eukaryotes. From the beginning, the metabarcoding studies have radically changed our view of eukaryotic diversity, revealing the huge species richness in any kind of explored environment. Metabarcoding studies of microbial eukaryotes allow

assessing the global diversity of marine plankton (de Vargas *et al.* 2015), reveal the uncharted diversity of marine benthos (Forster *et al.* 2016), highlight the importance of rare biosphere (Logares *et al.* 2015), uncover the diversity of poorly documented taxonomic groups (Lecroq *et al.* 2011; Hartikainen *et al.* 2014), document seasonal changes in protist communities (Egge *et al.* 2015a) and promote the use of protist-based diversity indices in biomonitoring (Pawlowski *et al.* 2014b).

All these studies are based on high-throughput sequencing of variable regions of 18S rRNA genes considered as the universal DNA barcodes for most groups of protists (Pawlowski *et al.* 2012). In general, the rRNA genes are characterized by low level of intra-specific variations due to the mechanism of concerted evolution. However, there are several studies showing high level of intraspecific variations in rRNA genes of eukaryotes, e.g. in free-living nematodes (Bik *et al.* 2013; Dell'Anno *et al.* 2015) or in metazoan parasites (Resende *et al.* 2011; Cooper *et al.* 2016). Even more disturbing are studies demonstrating the presence of intragenomic variations of rRNA genes in some taxa (Rooney & Ward 2005; reviewed in Weber & Pawlowski 2014). Among protists, such variations have been observed in ciliates (Gong *et al.* 2013) and radiolarians (Decelle *et al.* 2014), but not in choanoflagellates (Nitsche & Arndt 2015). High intragenomic variation was also found in 28S and ITS rRNA genes of nematodes (Pereira & Baldwin 2016).

In foraminifera, the IGP was shown to be widespread in different taxonomic groups (Weber & Pawlowski 2014). This study based on cloning and sequencing of partial 18S rRNA genes from 16 species, found the high levels of intragenomic variability of up to 5.15% in some species. The authors observed frequent single nucleotide polymorphisms (SNP) and expansion segments polymorphisms (ESP), which however, do not seem to have impact on secondary structure of rRNA (Weber & Pawlowski 2014). In the case of one species (*Elphidium macellum*), for which many specimens have been examined morphologically and genetically, it has been proposed that the IGP is the result of hybridization between closely related species (Pillet *et al.* 2012). However, no further study of this phenomenon has been conducted in foraminifera.

In the previous studies, the IGP in foraminifera was analysed based on cloning and Sanger sequencing of amplicons issued from single-cell PCRs. Here, we used single-cell HTS to increase the sequencing depth and obtain much more accurate view of variations that occur within foraminiferal genomes. We analysed over 7

million sequences from 130 specimens, ranging from 1'000 to 190'000 reads per specimen, with an average value of 55'000 reads. Our results confirm the presence of IGP in the majority of examined species. However, we observe different levels of IGP depending on taxonomy and habitats where the species were collected. We discuss the possible origins of these variations and their implications for the HTS-based metabarcoding studies of eukaryotic diversity.

6.4. Materials and methods

6.4.1. DNA extracts

130 DNA extracts used in this study came from different location and were collected across several years by lab members. In all cases, the DNA was extracted from single specimens. However, in the case of Xenophyophora, which are large size deep-sea protists, only a fragment of the entire specimen has been taken for extraction. For all species more than one DNA isolate was examined. The number of the isolates is indicated next to the species name in the TableS 6.1.

In this study, two independent illumina runs were planned. The Abyss run used DNA extraction from specimens of the two ABYSSLINE cruises performed in 2013 and 2015 (details in Gooday *et al.* 2017) and the Poly run used specimens collected and extracted between 1994 and 2015 in our lab. All specimens were identifying morphologically. Sampling year and location for each specimens used are indicated in the TableS 6.1.

6.4.2. PCR amplification and sequencing

The samples were amplified by PCR using the primer pair s14F3 (ACG CAM GTG TGA AAC TTG) and s17 (CGG TCA CGT TCG TTG C) and re-amplified using the nested primer s14F1 (AAG GGC ACC ACA AGA ACG C) and s17. 35 and 25 cycles were performed for the initial and the nested PCR respectively with an annealing temperature of 50°. A unique combination of tags was used for each sample in order to multiplex them into Illumina library. Individual tags are composed of 8 nucleotides attached at each primer 5'- extremities. PCR were then purified using the Sephadex G-50 superfine resin (GE Healthcare) and quantified using QuBit HS dsDNA (Invitrogen). The same amount of each sample was pooled and a final purification

step was performed with High Pure PCR Product Purification kit (Roche Applied Science). In total, six libraries were prepared with the Illumina TruSeq® PCR free Preparation Kit and quantified with qPCR using KAPA Library Quantification Kit. The three Poly libraries were pooled into a single run and were sequenced on a MiSeq instrument using paired-end sequencing for 500 cycles with a standard kit v2, expecting 14 million reads. The three Abyss libraries were sequenced independently for 500 cycles with a nano kit v2, expecting 1 million reads.

6.4.3. HTS data analysis

Filtering, assembly and de-multiplexing steps were performed following the method described in Pawlowski *et al.* (2014a). The Poly and Abyss libraries were pooled and then a strict de-replication was performed in order to obtain Individual Sequences Units (ISUs). Filtering process for both libraries is summarised in the TableS 6.2. Not all the specimens sequenced were used in this study; only those with more than 1000 reads were kept. For each sample, an alignment with all ISUs was assembled and sequences were manually checked for contaminations.

In order to find the different ribotypes in the two hypervariable regions (37f and 41f), the sequences from all specimens of the same species were combined and analysed manually using Seaview (Gouy *et al.* 2010). The Single Nucleotide Polymorphisms (SNPs) were pictured using R software (R Core Team 2013). For each specimen, the most abundant sequence was plotted at the base and for each sites, the number of substitution was counted. The relative abundance of each substitution was calculated and represented by the length of the bar. The colours used were the same as in the Seaview software: A - red, T - blue, G - yellow, C - green, gap - grey. Identity matrix for each sample and each species were calculated using R with seqinr package (Charif & Lobry 2007). The mean of all pairwise distances was kept. ISUs were combined into one file and clustered in OTUs using two common methods, Swarm2 (Mahé *et al.* 2015) and UPARSE (Edgar 2013). OTUs were then reassigned to species using BLAST (Altschul *et al.* 1990) with a local database containing the ISUs representing each species. rRNA secondary structures were constructed using mfold (Zuker 2003) with default parameters.

6.5. Results

6.5.1. Molecular dataset

In order to investigate the intragenomic and intraspecific diversity of the 18S rDNA, 130 specimens of foraminifera representing 23 species and 19 genera were sequenced using Illumina Miseq. Each species was represented by 2 to 25 specimens. Among the 23 species, 14 belong to the class Globothalamea, coming from shallow and deep sea habitats (11 and 3, respectively) while the nine remaining species belong to the single-chambered (monothalamous) foraminifera, coming from shallow, deep sea and freshwater habitats (4, 3 and 2, respectively). In total, 7'184'136 good reads were obtained for a partial region of the 18S rDNA. This fragment comprises two hypervariable regions specific to foraminifera (37f and 41f, Pawlowski & Lecroq 2010) separated by a conserved region and is about 300 bp long. The number of good reads kept for each specimen is indicated in the Table S1.

The 23 species analysed in this study (Figure 6.1) were separated depending on their degree of polymorphism. In 12 species (indicated with a circle in the Figure 6.1) only SNPs were found. All these species have low level of intra-specific divergence (99%) and IGP ranging from 99% to 100%. All species from the deep-sea belong to this category, which also comprises few shallow water species (mainly monothalamiids) and the two freshwater species.

The remaining 11 species possess distinct expansion segment polymorphism (ESP) in hypervariable regions (indicated with a triangle in the Figure 6.1). We consider as the ESP a sequence with at least 3 mutations that always occur together. All species in this category belong to the multi-chambered class Globothalamea. Among them, there are three Ammonia species for which we obtain a lot of data coming from all around the world. The most variable species in our dataset was *Ammonia T1* with an intraspecific identity of 83%, however this species (or species complex) was also the most represented with 25 different specimens. We investigate the number of OTUs generated by this dataset using popular clustering methods: Swarm and Uparse. The sequences were clustered into 618 and 63 OTUs respectively. The largest amount of OTUs assigned to the same species was for Ammonia T1 who reaches 322 and 22 OTUs with Swarm and Uparse, respectively. Among the 23 species present in the dataset, only two species (*Psammophaga* and *Lieberkuhnia*) were represented by only one OTU with Swarm against 12 for Uparse clustering method (Figure 6.1).

	Species	intraspecific identity	Intragenomic identity	habitat	Clustering OTUs SWARM	Clustering OTUs UPARSE
	Ammonia T1 △	83%	95%-99%	Shallow	366	22
	Cymbaloporetta squammosa △	91%	96%-98%	Shallow	52	6
	Ammonia aberdoveyensis △	95%	97%-99%	Shallow	47	5
	Buccella frigida △	97%	97%-98%	Shallow	27	2
	Elphidium macellum △	97%	97%-98%	Shallow	14	3
	Ammonia aoteana △	98%	97%-99%	Shallow	19	3
	Rosalina sp. △	98%	96%-99%	Shallow	21	1
	Buccella peruviana △	98%	98%	Shallow	6	2
	Leptohalysis scotti △	98%	98%-99%	Shallow	5	2
	Planorbulinella sp. △	98%	98%-99%	Shallow	11	2
	Oridorsalis ○	98%	99%	Deep sea	9	1
	Notorotalia finlayi △	99%	99%	Shallow	5	1
	Aschemonella monile ○	99%	99%	Deep sea	7	1
	Cedhagenia saltatus ○	99%	99%	Shallow	3	1
	Lacogromia cassipara ○	99%	99%	Freshwater	4	2
	Epistominella exigua ○	99%	99%-100%	Deep sea	3	1
	Semipsammina sp1 ○	99%	99%	Deep sea	3	1
	Micrometula ○	99%	99%	Shallow	6	2
	Psammophaga magnetica ○	99%	99%	Shallow	1	1
	Allogromia laticollaris ○	99%	99%	Shallow	4	1
	Nuttalides ○	99%	99%	Deep sea	2	1
	Lieberkühnia sp ○	99%	99%	Freshwater	1	1
	Aschemonella aspera ○	99%	99%	Deep sea	2	1

Figure 6.1 List of the species used in this study with their level of pairwise intraspecific identity, the range of intragenomic identity, habitats and the number of OTUs generated by SWARM and UPARSE. The level of polymorphism is indicated next to the species name: circle – single nucleotide polymorphism, triangle – haplotype.

6.5.2. Single Nucleotide Polymorphism (SNP)

12 species show single or double mutations at some positions. Among them we can distinguish 3 groups. The first group is composed of species for which all specimens show a different pattern, this group is framed in red in the Figure 6.2 and FigureS 6.1. This is the case of the two Xenophyophoran species *Aschemonella monile* (Figure 6.2) and *Semipsammina sp1* (FigureS 6.1). In each case, several extractions (2 for *Semipsammina* and 4 for *A.monile*) have been done from fragments of a single specimen. Indeed xenophyophoran are big organisms from which we can extract only a part of the cytoplasm. All these "replicats" gave exactly the same sequencing pattern while the other specimens show a very different SNP pattern.

The second group is composed of species that share the same polymorphism pattern between their specimens; this group is framed in green in the Figure 6.2 and FigureS 6.1. Eight species were concerned: *Epistominella exigua*, *Psammophaga magnetica* and *Aschemonella aspera* presented in the Figure 6.1 and *Micrometula*, *Nuttalides umbonifera*, *Allogromia latticolaris*, *Lieberkuehnia sp* and *Lacogromia cassipara* in the FigureS 6.1. The mutation rate can change between the different specimens (e.g. *Micrometula*) however the position and the type of the mutation did not change. The highest mutation rate in this group was found in *N.umbonifera*, in one specimen, 40% of the reads have an insertion of a G in a specific position. On the other hand, in *E.exigua* the highest mutation rate hardly reached 1.5% of the reads.

The third group is composed of two species with more complex SNP patterns. The concerned species are: *Oridorsalis umbonatus* (Figure 6.2) and *Cedhagenia saltatus* (FigureS 6.1). In *O. umbonatus*, represented by 9 specimens, three different SNP patterns are found, two (I and II) are shared by 4 specimens each and the last pattern (III) is present in one specimen only. All those mutations occur in the second hypervariable region (41f). The two polymorphic sites in the pattern I did not occur in the same position than the two of the pattern II. However, those both couple of mutations are compensatory, as shown by the predicted secondary structure (Figure 6.3). Moreover, 2 nucleotides are different at position 68 and not shared by the two patterns. The pattern III shares the same mutation sites as the pattern I with 7 additional mutated sites. The two bases at position 68 that are different in the two first patterns are represented in the pattern III (60% for C and 40% for TT, Figure3). Interestingly, the isolate 18658 (pattern I), shows an additional mutation in the 37f hypervariable region present in almost 50% of the reads. This region appears to be conserved in all other tested specimens. For *C. saltatus*, the two specimens share

the same SNP pattern and the third specimen shows a slightly different one. However three mutation sites are conserved between all the specimens.

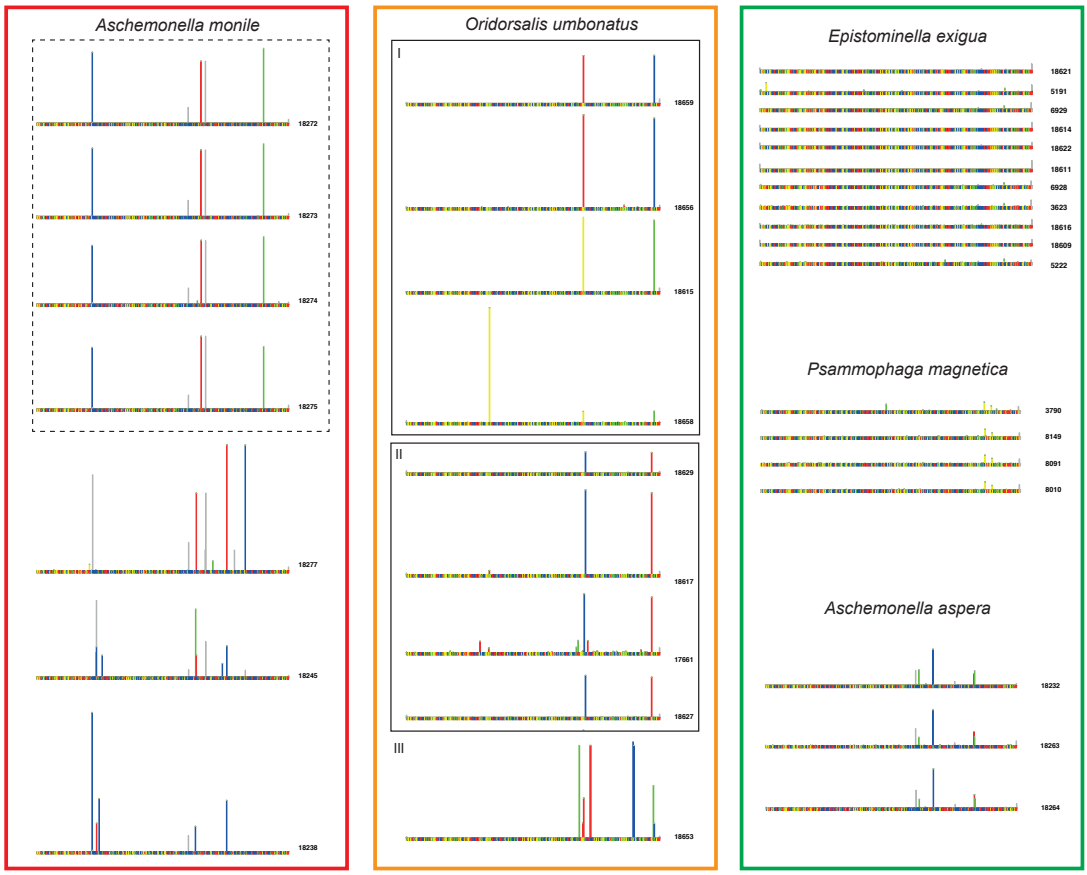


Figure 6.2 SNP patterns for 5 different species. The consensus sequence for each specimen represents the most abundant ISU. Coloured boxes correspond to the three groups described in the results section. For each specimen, the number of extraction is indicated next to the consensus sequence. Dotted box corresponds to several extractions from the same specimen and plain box represents the different SNP patterns found in *O. umbonatus*, indicated with roman numbers.

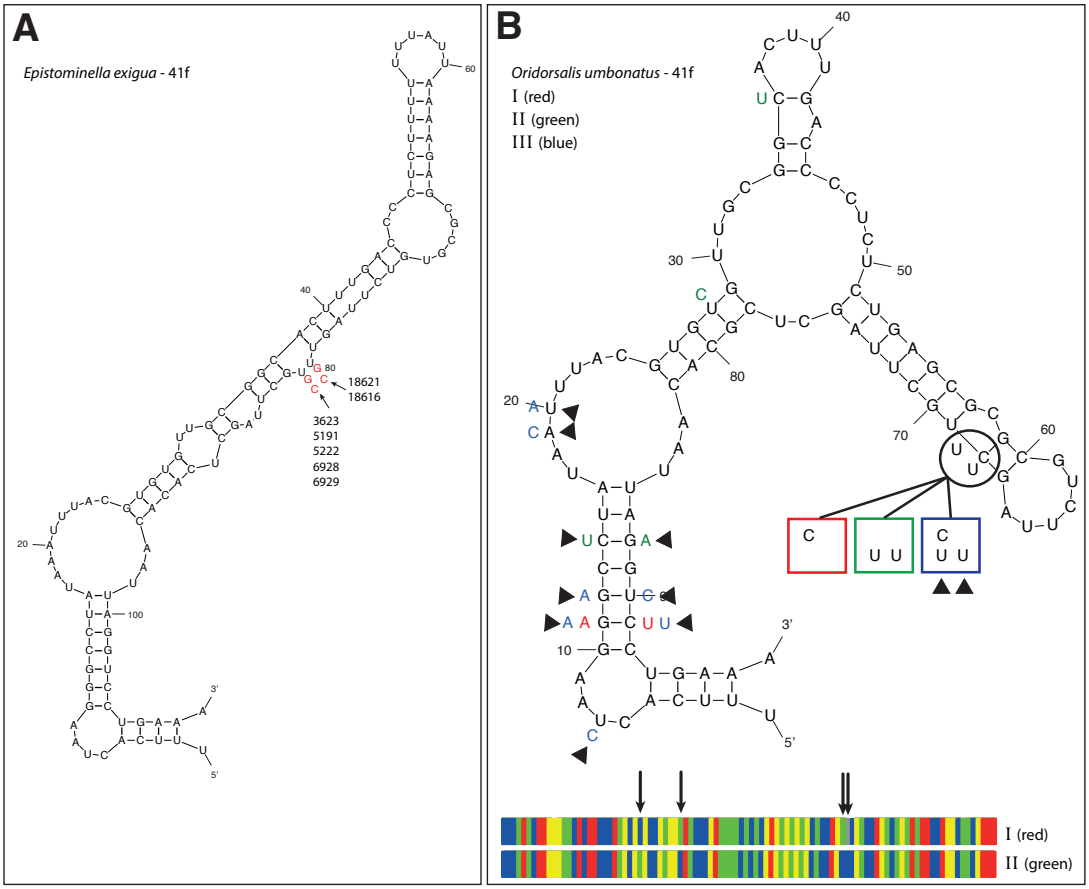


Figure 6.3 Secondary structures of the rRNA 41f hypervariable region of *Epistominella exigua* (A) and *Oridorsalis umbonatus* (B). The SNPs described in the Figure 6.2 are indicated with colors. A. The number of isolate corresponding to each mutation is indicated. B. For *O.umbonatus*, high rate polymorphisms are indicated with a plain triangle. The aligments at the bottom of the figure B represents the sequences of the 41f hypervariable region of the patterns I and II. Arrows represent the sites that are different and not variable between the two patterns.

A lot of small and rare mutation events were observed in all isolates despite the abundance threshold applied to the dataset (ISUs with 10 or less reads were discarded). We plotted the number of sites that vary at least one time in each specimen as function of the total number of reads, both for the entire fragment (300bp) and for the first conserved region (68bp) (Figure 6.4). Our analysis shows that starting from about 60'000 reads, 100% of the sites varied at least once. This tendency was the same for the entire fragment and for the conserved region.

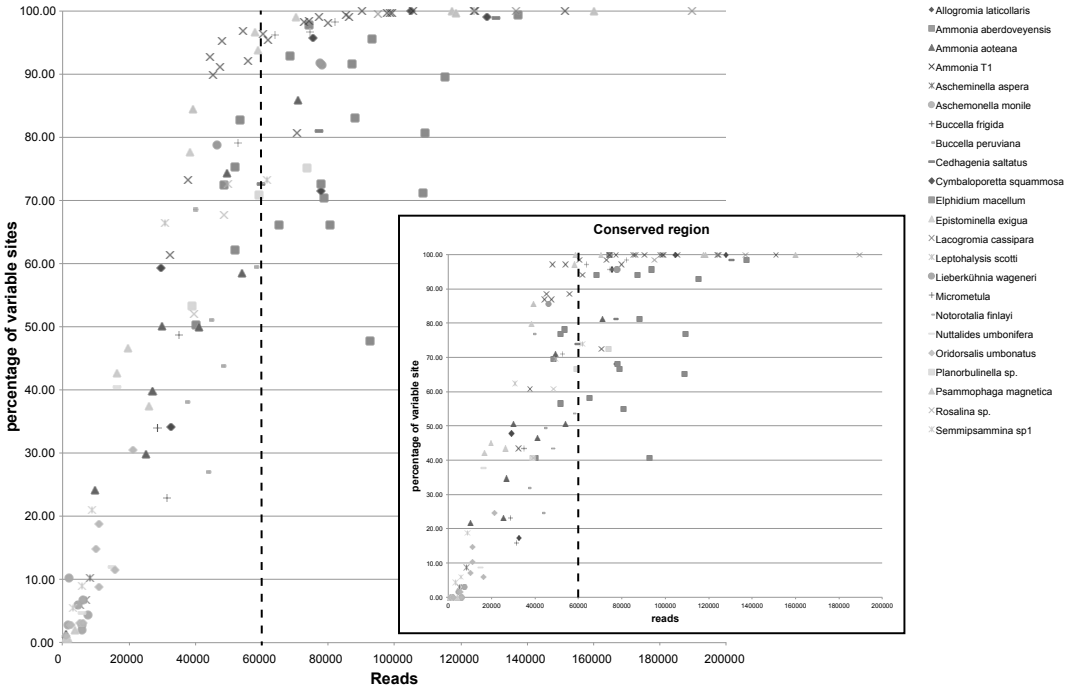


Figure 6.4 Percentage of variable site for the 130 samples plotted as function of the total number of reads on the entire fragment (300 bp). The boxed plot is calculated on the first conserved region (68 bp). Dashed lines at 60'000 reads are plotted to represent the threshold where all sites are mutated at least one time.

To investigate deeper the amount of technical errors generated by our methodology, we took as example the species *Epistominella exigua*, which appears to be very stable with no obvious intragenomic polymorphism. Eleven specimens were sequenced with different sequencing depth (ranging from 1118 to 70385 reads per specimen). The Figure 6.5 represents the mutation rate of each specimen across the entire fragment. Seven specimens with more than 10'000 reads were kept for this analysis. The remaining 4 specimens, with only 2 and 4 variable sites were considered not representative for this analysis (data not shown). Our analysis shows that higher mutation rate occurs at the beginning and at the end of the fragment, probably due to the primer trimming during sequence processing. Otherwise, mutations seem to be distributed uniformly across the fragment except for 2 peaks that occur within the 41f hypervariable region and that are present in each samples (mutation rate between 0.5% and 1%). As shown in the Figure 6.5, the first peak corresponds to a deletion event in a repetition of 7 T, while the second peak corresponds to a substitution of one T with a G or a C (almost with equal rate). Those substitutions occur in a two-nucleotide bulge and are probably not subjected to high

constraints (Figure 6.3). Excluding the beginning and the end of the sequence as well as the two peaks in 41f region that may correspond to biological polymorphism, the average mutation rate never exceed 0.3% for each base position. The substitution rate, calculated on the first conserved region of *E.exigua*, is therefore of one change every 1000-2000 sequenced bases.

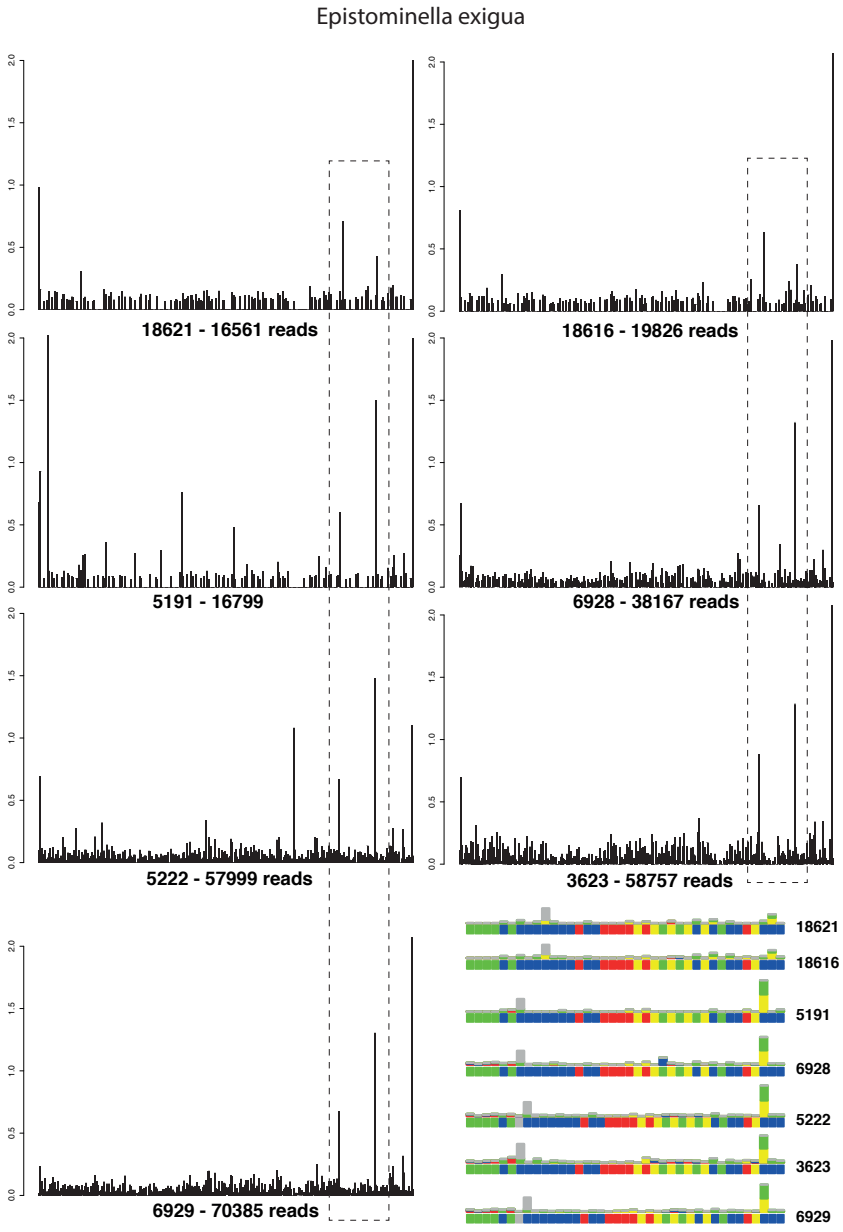


Figure 6.5 Substitution rates (in percentage) for each site in seven specimens of *Epistominella exigua*. The number of reads is indicated next to the isolate number, only specimens with more than 10'000 reads are shown. The boxed region is enlarged for each sample with the representative colours for each nucleotide (A - red, T - blue, G - yellow, C - green, gap - grey).

6.5.3. Expansion segments polymorphism (ESP)

Eight species showed ESPs in their hypervariable regions: *Elphidium macellum*, *Planorbulinella sp* and *Notorotalia finlayi* (Figure 6.6), and *Cymbaloporetta squammosa*, *Rosalina sp*, *Buccella frigida*, *Buccella peruviana* and *Leptohalysis scotti* (FigureS 6.2). Six of them present substitutions in both hypervariable regions (37f and 41f). *Planorbulinella sp* and *N. finlayi* showed variation only in one of the two hypervariable regions sequenced (37f and 41f, respectively). In all cases, the different ESPs do not seem to be uniformly distributed across the specimens and some specimens show patterns that are not shared by the other specimens from the same species. We investigate deeper the ESPs of two species, *Planorbulinella sp* and *N. finlayi*. For *Planorbulinella sp*, four different haplotypes were found, consisting of two different patterns (Figure 6.7). One pattern consists of an ESP in the hairpin loop of the 37f region, comprising either the small ESP shared by the haplotype 1 and 4 (white and pink on Figure 6.6) or the long ESP shared by the haplotype 2 and 4 (blue and green on Figure 6.6). The second pattern, occurring in haplotype 3 and 4, is a double insertion of two nucleotides into the double strand, which are compensatory. Two other compensatory positions in the double strand tend to change between the four haplotypes. Changes in the sequences of the haplotypes 1 and 2 as well as the SNP in the haplotype 4 appears to stabilise the double strand (Figure 6.7). The other species, *N. finlayi*, presents two haplotypes in the hypervariable region 41f composed of two different hairpin loop ESPs and two compensatory SNPs shared by haplotypes.

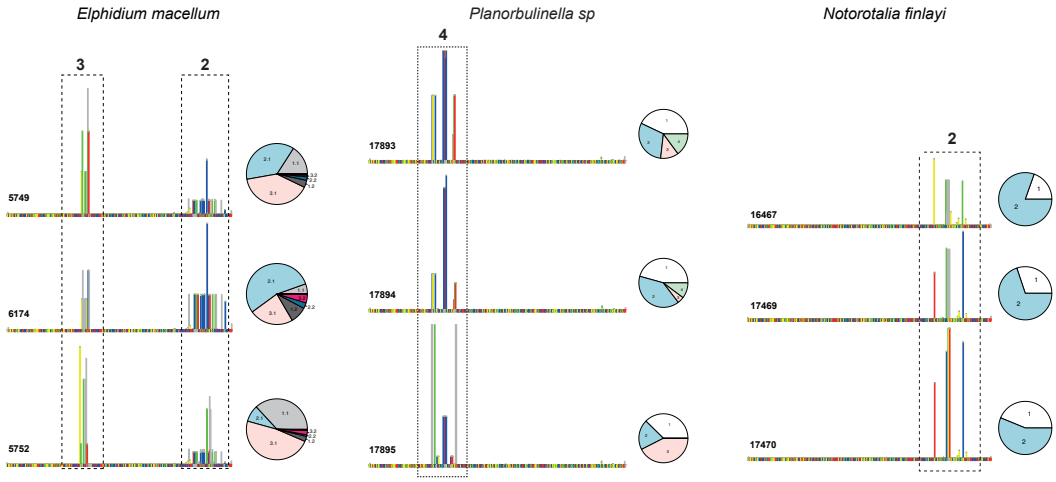


Figure 6.6 Expansion segments polymorphism of three species. The consensus sequence for each specimen represents the most abundant ISU. Pie charts show the distribution of each

ESP within one specimen. The number in top of the dashed box represents the number of ESP found for this region. For each species, the number of isolates is indicated just above the consensus sequence.

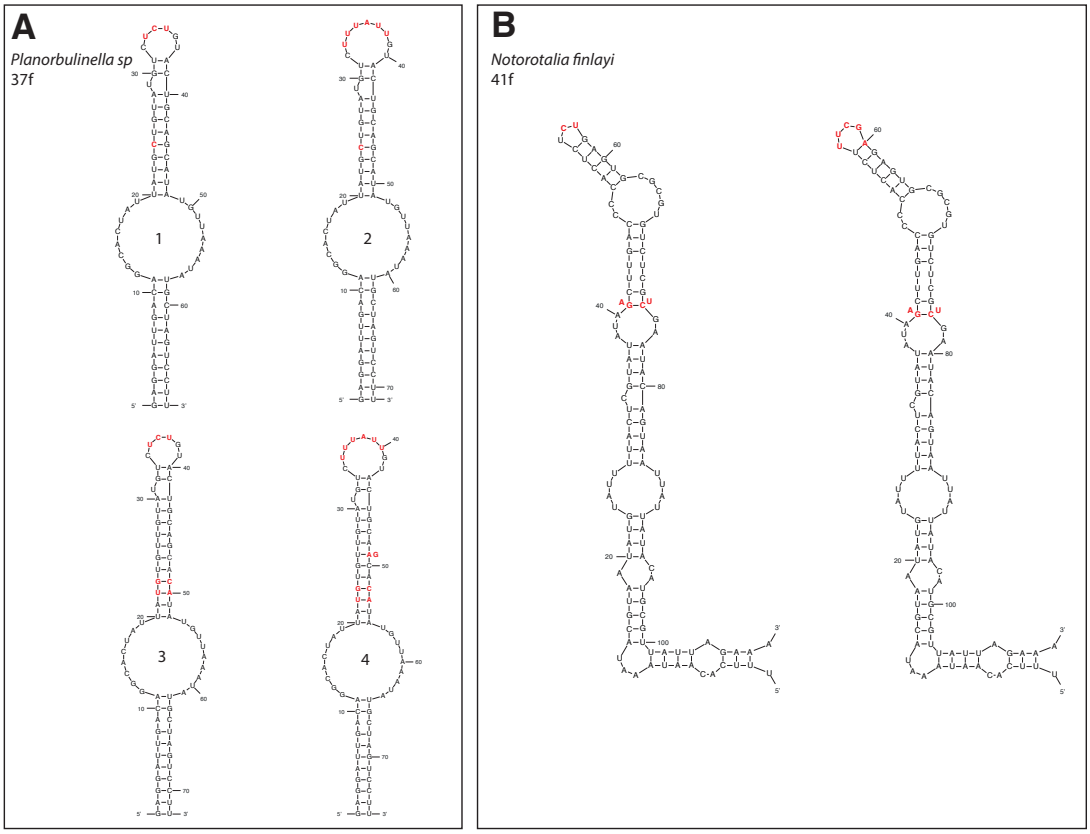


Figure 6.7 Secondary structures of the different ESPs in *Planorbulinella sp* (A) and *Notorotalia finlayi* (B) for the rRNA hypervariable region 37f and 41f, respectively. ESPs and SNPs are indicated in red.

In addition, we analysed three *Ammonia* species known to show different ESP patterns. We analysed specimens from different geographic localities in order to examine the biogeographic distribution of different haplotypes.

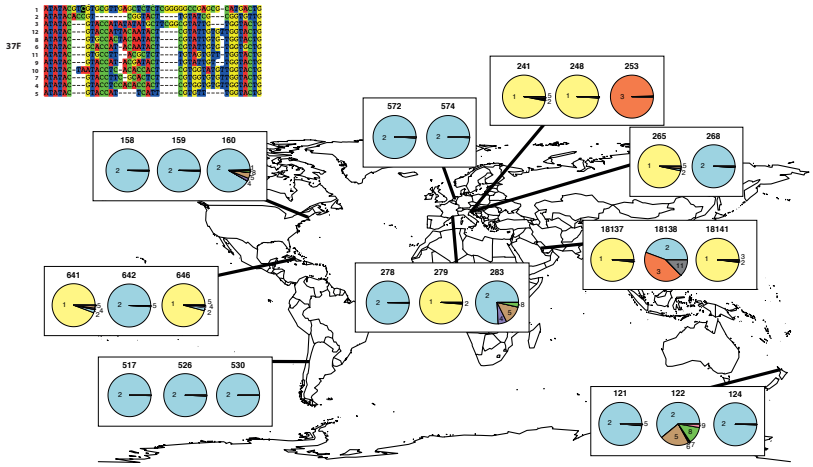
Ammonia T1 shows 12 different ESPs in the 37f region. Three ESPs are dominant and highly different from each other (1 – yellow, 2 – blue and 3 – orange, Figure 6.8). The haplotype 1 was most abundant in eight specimens, while the haplotype 2 dominates in 15 specimens. The haplotype 3 was much more rare and was found as unique haplotype in one specimen, and shared with the haplotype 2 in another one. No obvious geographical pattern was found for this species, although the haplotype 1 seems to be present in Northern hemisphere only. Interestingly, the haplotype 1 occurs in the same localities as haplotype 2, but no specimens were found to contain

both of them, moreover the pairwise distance between those 2 haplotypes shows a barcoding gap (FigureS 6.3 SNP pattern for the species *Ammonia T1*. Coloured boxes correspond to the most abundant ESP of the specimen. The frequency of the pairwise distance matrix is ploted in the bottom left corner.). The SNP patterns of the 25 specimens of Ammonia T1 are presented in the FigureS 6.3. This shows that even if ESPs have been found only in 37f, the 41f region also varies in some samples.

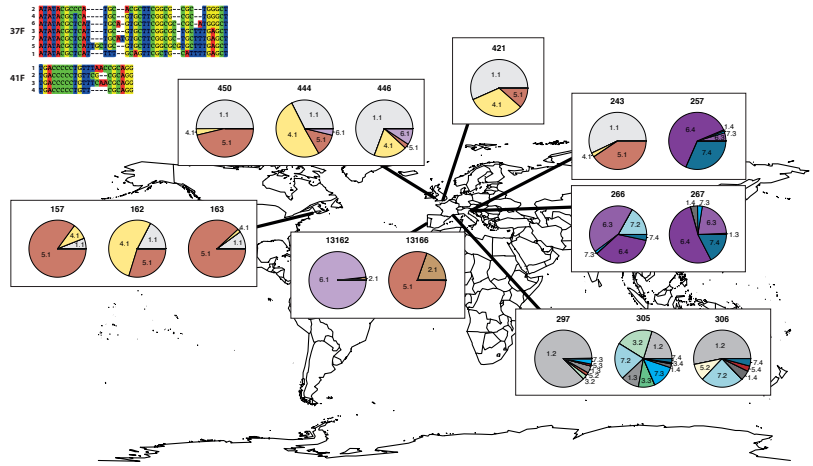
For *Ammonia aberdoveyensis*, we found seven ESP in the 37f region and four ESP in the 41f region. In contrast to *Ammonia T1*, most of the specimens of *A.aberdoveyensis* show high intragenomic diversity, sharing different ESP in the same specimens. However, while the different 37f ESP are shared in all examined localities, the variation of 41f ESP is geographically restricted and shared only by specimens collected in the Mediterranean Sea (Camargue and Adriatic sea). The ESP 37f-6 (purple in Figure 6.8) seems also present essentially in the Adriatic Sea and in the Western Mediterranean Sea. Specimens from Camargue in France show higher diversity than in other locations. The SNP patterns of the 16 specimens of *A.aberdoveyensis* are present in the FigureS 6.4.

Compared to the other two species *Ammonia aoteana* has much more limited distribution and is found only in the Southern hemisphere. We found 3 ESP in the 37f region in this species. Two of them (1 and 2) were shared by specimens from NZ and Chile, while the third ESP (3) was present exclusively in two specimens from Australia. Interestingly, the sequences of the Australian ESP 3 (pink) are closer to the ESP 2 (blue), than the later one is to the ESP 1 (white) (12 nucleotides differences between 1 and 2 against only 7 between 2 and 3). The FigureS 6.5 shows that in *A.aoteana* SNP patterns are present in the 41F hypervariable region. However these different mutations do not seem to occur together. Indeed, all the different combinations of each mutation were found in the dataset.

Ammonia T1



Ammonia aberdoveyensis



Ammonia aoteana

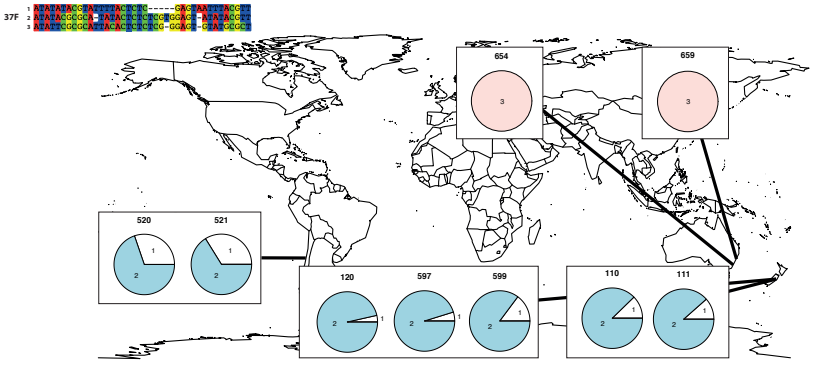


Figure 6.8 Mapping of the different specimens of the three *Ammonia* species sequenced (T1, *aberdoveyensis* and *aoteana*). For each specimen, the distribution of each ESP is represented by a pie chart. The sequence of each ESP is indicated in the upper left corner for each species.

6.6. Discussion

6.6.1. Technical vs biological origin of IGP

The relatively high rates of IPG found in our study are congruent with previous analyses performed using cloning and Sanger sequencing. We have found the same number of ESP in the 37f and 41f regions of *Elphidium macellum* and in the 41f region of *Psammophaga magnetica*, as in the studies of Pillet *et al.* (2012) and Weber & Pawlowski (2014), respectively. In some cases we found more ESPs and SNPs than previously, which could be easily explained by sequencing depth. For example, in the case of *Ammonia T1*, 3 ESP were found in the previous study, against 12 here. However, only 3 ESPs were abundant in the HTS dataset, and therefore appears in the Sanger sequencing. On the contrary, for *Leptohalysis scotti*, the Sanger analysis reveals an ESP that we miss in the HTS dataset. This could be explained by the fact that only 2 of the 3 specimens sequenced by Sanger were used in this study.

One of the major concerns about studies involving sequencing, and more particularly high-throughput amplicon sequencing, is the rate of technical errors engendered by the different steps. Some authors (Weber & Pawlowski 2014) consider the singleton mutations as technical errors. These authors found that technical errors range from 0.3% and 3.35% (1 error every 30-300 bases). However, Sanger sequencing is considered as an accurate sequencing method with an error rate of 1 base every 100’000 (Ewing & Green 1998), which is interestingly the same error rate of the Taq polymerase during the PCR reaction (Tindall & Kunkel 1988; Cline *et al.* 1996; McInerney *et al.* 2014). Given the high fidelity rate of Sanger sequencing, we can deduce that technical error rates have been largely overestimated in this study and that the level of polymorphism is higher than previously thought. In our case, we showed that the technical errors increase in function of the sequencing depth and we estimated the error rate as 1 error per 1000-2000 sequenced bases (0.05-0.1%), which is consistent with the study of Schirmer *et al.* (2015) for the MiSeq Illumina technology.

The difference between a technical error and a biological polymorphism is not always easy to pick up. We often prefer to discard polymorphisms than to keep technical errors. However, for some species the polymorphism rate can easily be confused with technical errors. For example, in the case of *Epistominella exigua*, we have

found a SNP that barely reaches 1% of the reads, but which appears to be a real polymorphism, because it is shared by several specimens and did not affect the secondary structure of the rRNA. Some other SNPs are also found at this relative abundance but are not shared by the other specimens. This example highlights the issue related to the fixed thresholds that sometimes may remove not only technical noises but also biological variations.

6.6.2. Taxonomic context

One of the arguments for the biological origin of IGP in foraminifera is the fact that its level is closely related to the taxonomic affinities of species and their habitats. As far as the taxonomy is concerned, we found much lower rate of IGP in monothalamous foraminifera than in Globothalamea. At first glance, this finding may appear unexpected, as the monothalamiids are generally known to have highly divergent sequences (Pawlowski *et al.* 2003). Indeed, the previous study (Weber & Pawlowski 2014), did not report any particularly low intragenomic divergence in monothalamiids, averaging 2-3% in *Micrometula* and even exceeding 4% in one specimen of *Conqueria laevis*.

These differences between present and previous results could be simply explained by the effect of how the IGP was calculated: either deduced from the number of variable sites, or calculated with pairwise distances. In the previous study, the authors considered that all positions that mutated at least twice across all clone sequences were polymorphisms. Then they calculated the “worst scenario” where all mutations occur in the same sequence and deduce the percentage in function of the length of the sequence. Doing that they did not take into account the number of sequences used for analysis. This is probably not an issue in the case of Sanger sequencing of cloned amplicons, however it become very problematic in the case of HTS studies. Indeed, we tried to calculate the percentage of maximum differences, as calculated with Sanger sequencing dataset, in two species that were used in both studies (*Psammophaga magnetica* and *Leptohalysis scotti*). We took into account the sites that mutated at least in two ISUs, remembering that ISUs present with less than 10 reads were already discarded from the dataset. We inferred percentage of maximum difference ranging from 4% for the less sequenced specimen (*L. scotti* – about 30’000 reads) to 62% for the most sequenced specimen (*P. magnetica* – about 160’000 reads). The increase in the divergence seems to be correlated with the

sequencing depth (R^2 = 0.96 on the four specimens of *P. magnetica*, data not shown). This huge difference in calculation makes direct comparison very difficult.

Another explanation of the observed differences between both studies could be that the fragment analysed here was about twice smaller than in Weber & Pawlowski (2014), and that some IGP events were present in regions that were not sequenced here. For example, the *Micrometula* species shows a higher variability in the helix 43e and 45e than in 37f and 41f (about 5% and 7% of maximum intragenomic divergence against 4% and 2% for 37f and 41f).

Finally, the low IGP level in analysed monothalamids in the present study may be the effect of selection of particular lineages representing either slowly evolving deep-sea species (see below) or cultured freshwater foraminifera. In the later case, the low IGP level could be a characteristic feature of this group, resulting of the evolutionary bottleneck when the marine species adapted to the freshwater habitats. However, it could also be due to the bottleneck and possible clonal reproduction when the species were isolated and cultivated. Indeed, the IGP of marine *Allogromia latticolaris*, coming from culture, also showed a low level of polymorphism. There is some evidence that cultivated eukaryotes show much lower level of intraspecific variations than the natural isolates, e.g. *Caulerpa taxifolia* (Jousson *et al.* 2000) and some fungal species (Simon & Weiß 2008). This could explain the fact that the IGP goes often unnoticed when the genomes of cultivated strains are sequenced.

6.6.3. Ecological context

The most striking finding of this study is very low level of IGP in deep-sea species, compared to those living in shallow-water. The low genetic variations in abyssal foraminifera were already demonstrated in the case of the three rotaliids species (Pawlowski *et al.* 2007). Two of them (*Oridorsalis* and *Epistominella*) are also analysed here, together with another rotaliid (*Nuttalides*) and three xenophophores. This last group is particularly interesting because it comprises unusually large protists, which build huge skeletons filled with long strings of multinucleated cytoplasm (Gooday *et al.* 2017). The absence of IGP in such large plasmodial structures is quite surprising, given thousands of nuclei present in every extracted fragment (Lecroq *et al.* 2009b). Interestingly, there is no variation between different fragments of the same specimen, but each specimen of the same species is genetically different. This is very different from what is observed in *Epistominella*,

where specimens from distant locations are genetically almost identical (Pawlowski *et al.* 2007).

Low IGP in the deep-sea species compared to shallow water ones could be explained by lower evolution rate in deep-sea habitats. Several studies showed that for some eukaryotic groups, the rates of evolution in the tropics are higher than elsewhere in the globe (Davies *et al.* 2004; Allen & Gillooly 2006; Gillman *et al.* 2010). This may be influenced by such factors as generation time, metabolic rates and UV radiation (Weller & Wu 2015; reviewed in Dowle *et al.* 2013). In the case of deep-sea habitats, the low genetic variability could be explained by the lower density of organisms (Pernice *et al.* 2015), the wide dispersal of species and clonal reproduction. Concerted nuclear divisions in large-sized xenophyophores could explain the genetic homogeneity of sparsely distributed specimens. However, further studies of closely related coastal and deep-sea species will be necessary to show whether the observed difference in IGP level is due to ecological conditions rather than taxonomy.

6.6.4. Implications for the metabarcoding surveys

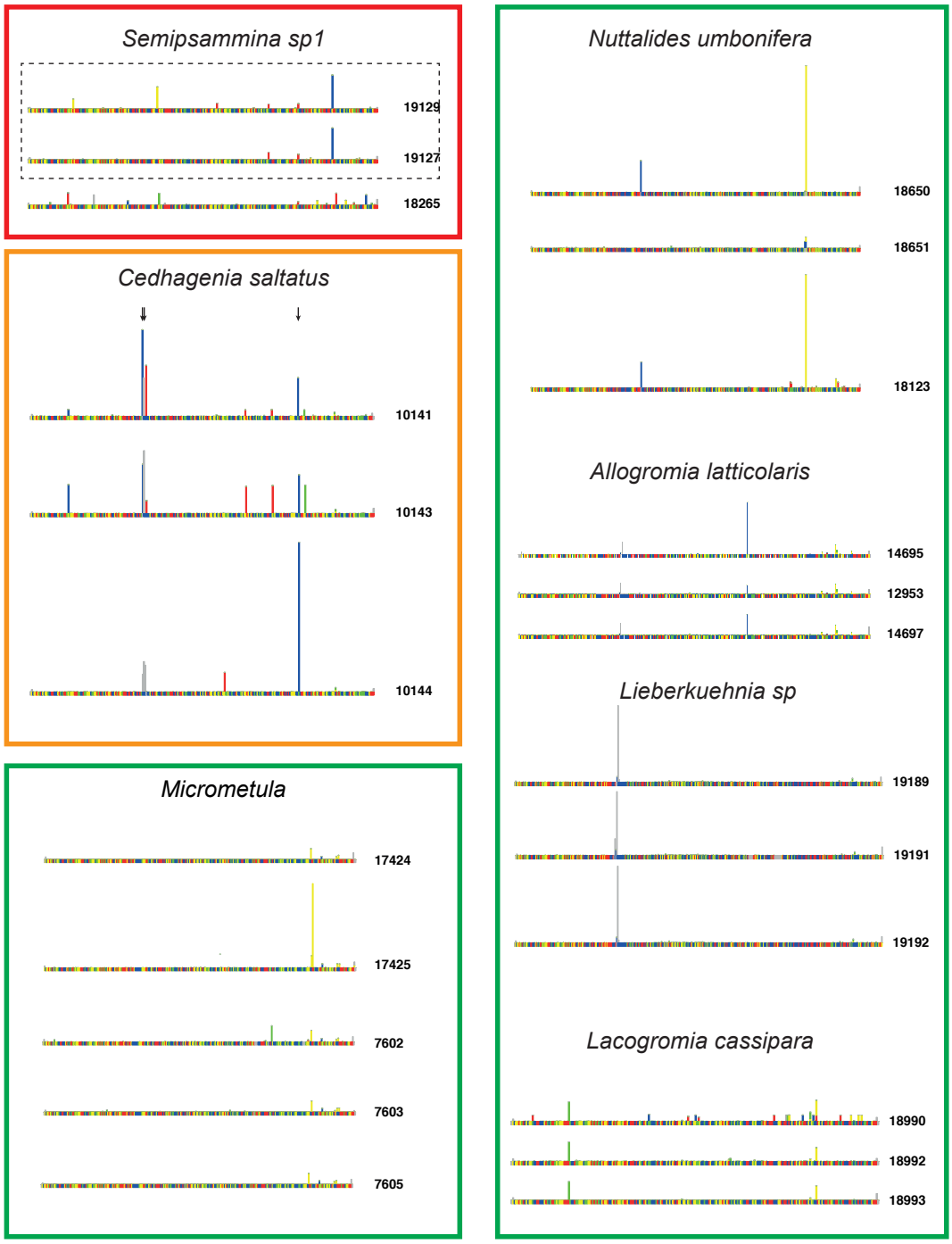
Obviously, the IGP have a great importance in the analysis of metabarcoding data. First, the IGP has to be assessed to avoid the overestimation of species richness. If all haplotypes are considered as different species, then of course the number of species is artificially inflated. As shown by our study, the commonly used clustering methods (Swarm, UPARSE) overestimate the real number of species from 3 to 25 times.

It is difficult to say whether the high level of IGP is specific to foraminifera or it concerns also other groups of protists. In general, with exception of few groups (e.g. choanoflagellates, Nitsche & Arndt (2015), nothing is known about the IGP level in most of protists taxa present in metabarcoding data. Up to our knowledge, no tests of IGP level are conducted prior to metabarcoding studies, such as the analysis of IGP level in genus *Reticulomyxa* that has been used as a threshold in HTS analysis diversity of deep-sea foraminifera (Lecroq *et al.* 2011). Perhaps, the interpretation of metabarcoding data could be different if the IGP was taken into account. For example, one could ask whether the high richness of apicomplexan in rain forest (Mahé *et al.* 2017), is not overestimated knowing the high level of genetic polymorphism in this group (Mercereau-Puijalon *et al.* 2002; Gardner *et al.* 2002; Rooney 2004).

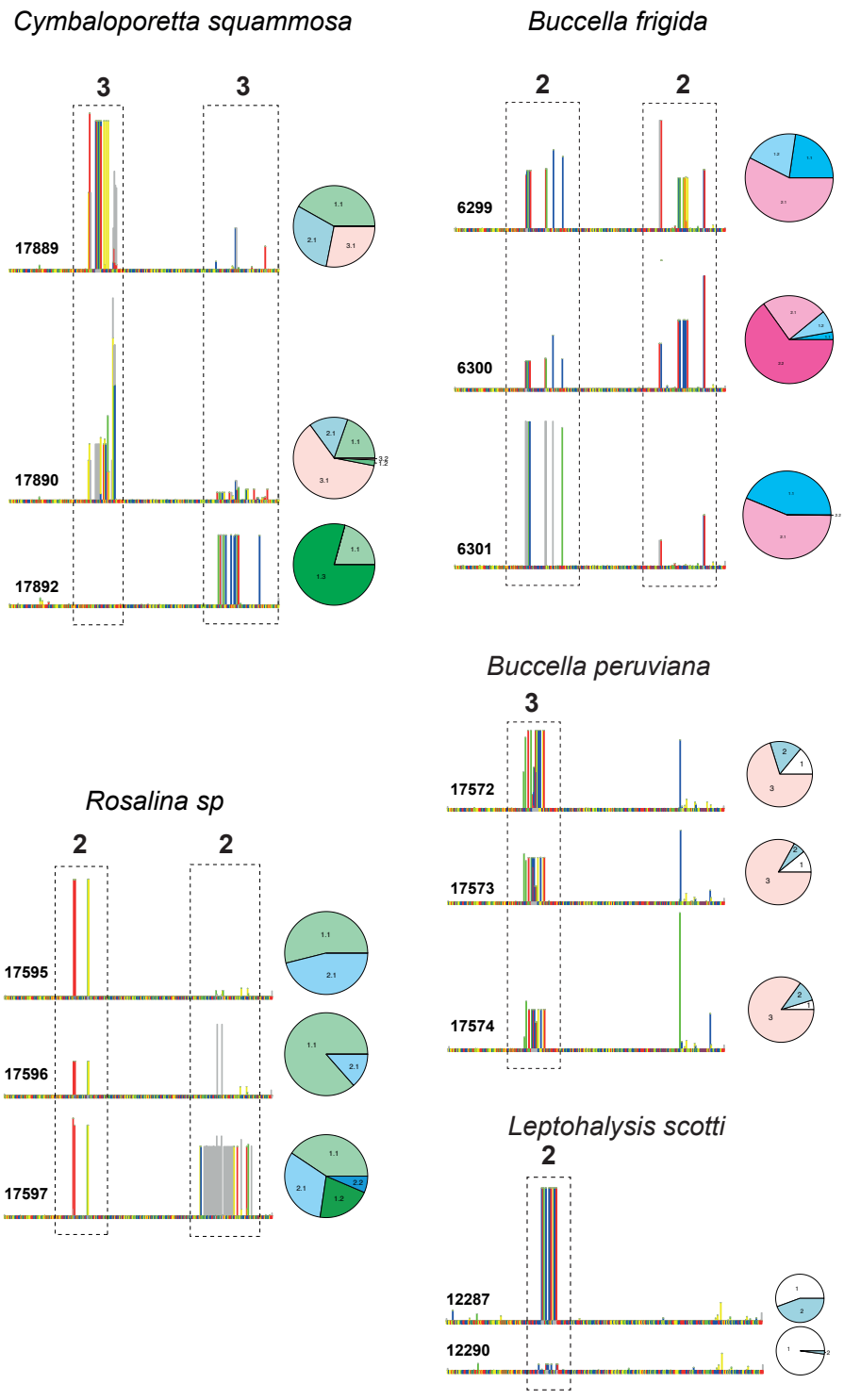
In addition, the single-cell HTS approach could help interpreting the environmental surveys by providing population-level information about encountered species. As shown by our study, screening of single cells polymorphisms allows detecting cryptic species and facilitates their distinction. The ultra deep sequencing of *E.exigua* shows the presence of a small intraspecific variation that has been ignored by previous Sanger sequencing-based studies (Pawlowski *et al.* 2007; Lecroq *et al.* 2009a) and would probably pass unnoticed in environmental survey. The analysis of single-cell HTS data could also provide insight into the occurrence and distribution of haplotypes associated with a given species. Commonly studied in metazoan and plants, the presence of haplotypes is often ignored in protistological research. Numerous haplotypes detected in *Ammonia* spp. show that these common shallow-water foraminifera are genetically complex organisms, originating probably as a result of hybridization between different populations. Compared to Sanger sequencing, commonly used in DNA barcoding, the HTS approach applied to single specimens offers new perspectives to explore the origin and geographic distribution of species. By combining the data on large number of specimens and large number of copies of amplified genes, the single-cell HTS contributes to a global view on species genetic variation, which is essential to our understanding of their biology and ecology.

6.7. Supplementary data

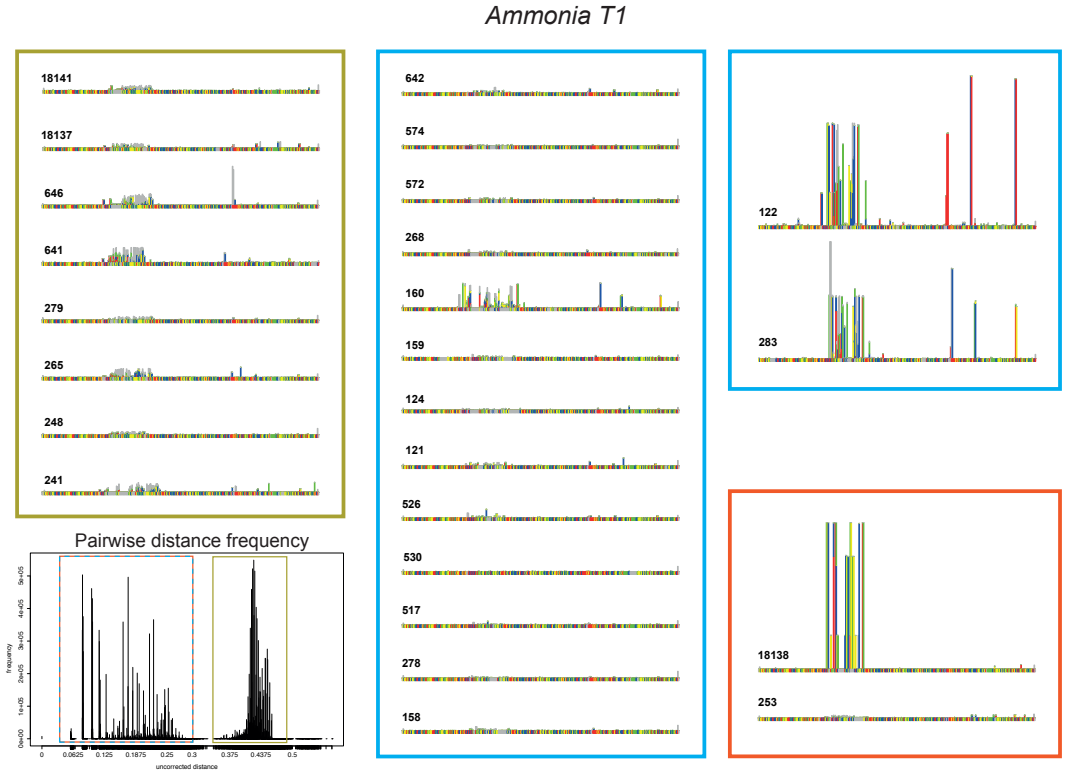
FigureS 6.1 SNP pattern for 7 different species. The consensus sequence for each specimen represents the most abundant ISU. Coloured boxes correspond to the three groups described in the results section. For each specimen, the number of extraction is indicated next to the consensus sequence. Dotted box correspond to several extractions from the same specimen. Arrows in *C. saltatus* correspond to the mutation sites conserved among the three specimens.



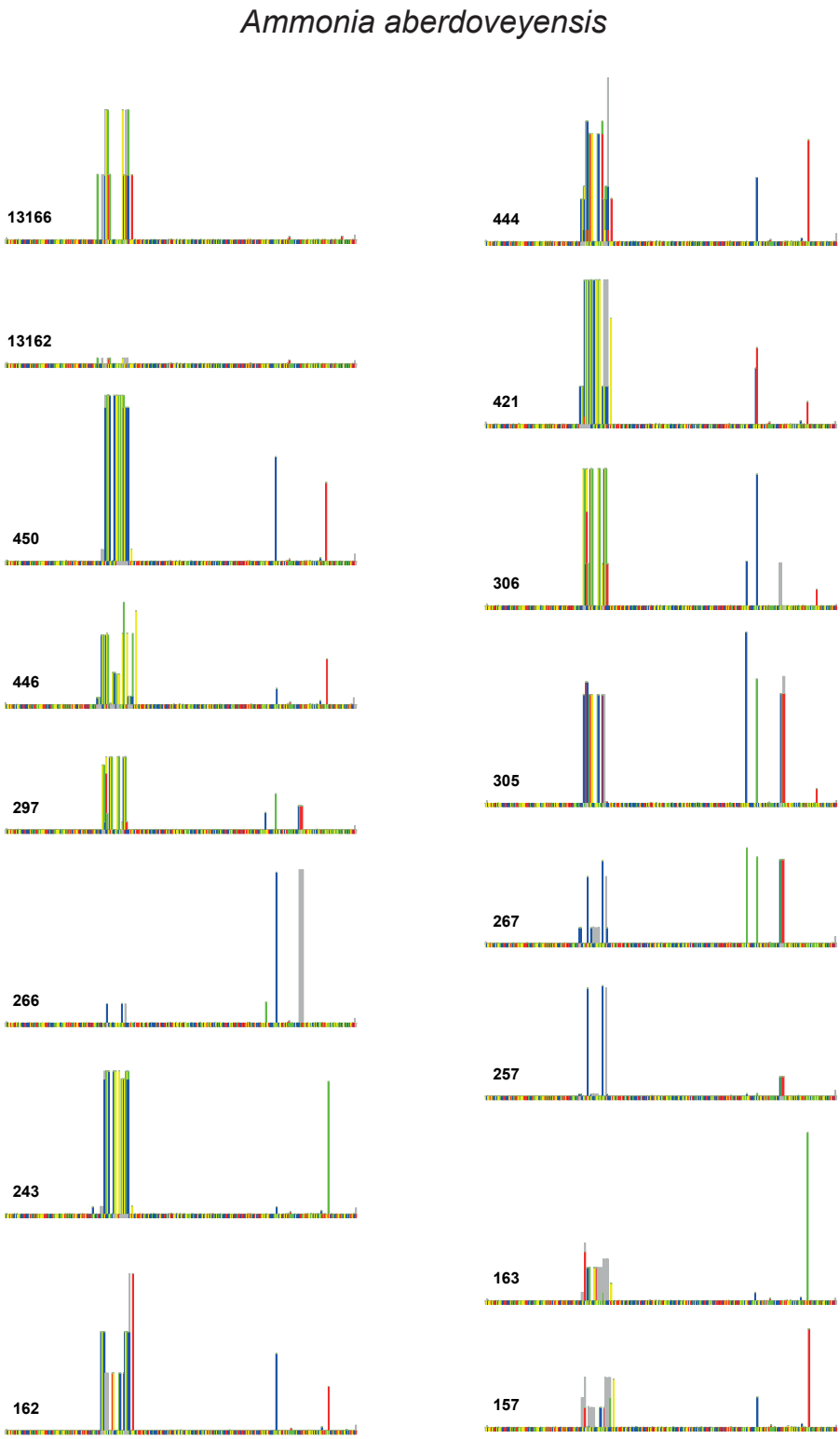
FigureS 6.2 ESP pattern of five species. The consensus sequence for each specimen represents the most abundant ISU. Pie charts show the distribution of each ESP within one specimen. The number in top of the dashed box represents the number of ESP found for this region. For each specimen, the number of extraction is indicated just above the consensus sequence.



FigureS 6.3 SNP pattern for the species *Ammonia T1*. Coloured boxes correspond to the most abundant ESP of the specimen. The frequency of the pairwise distance matrix is ploted in the bottom left corner.

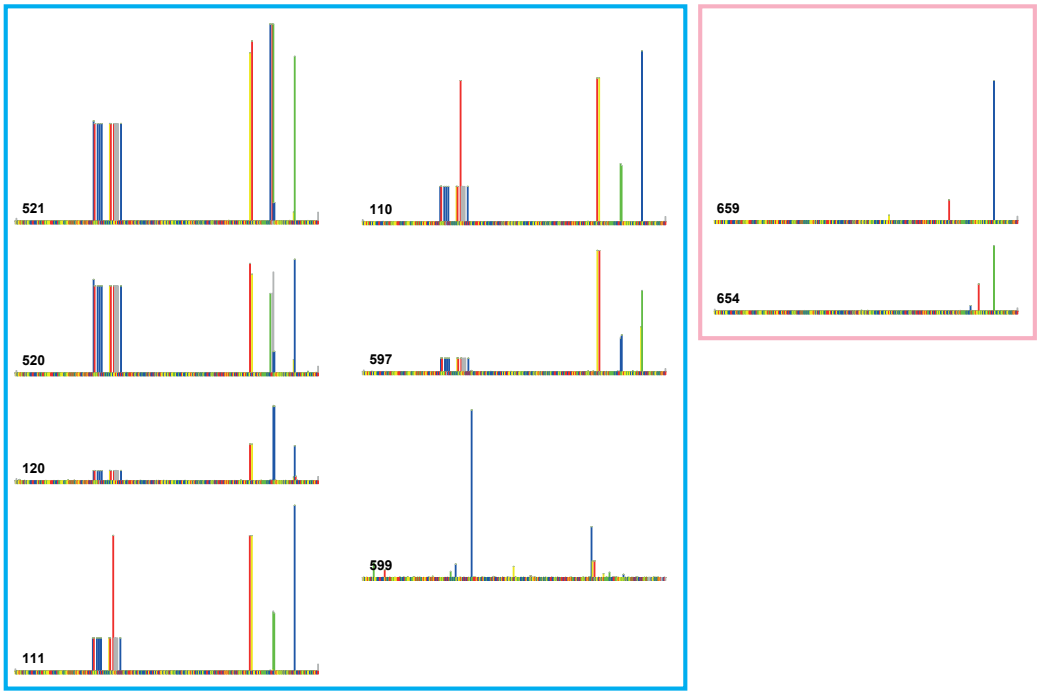


FigureS 6.4 SNP pattern for the species *Ammonia aberdoveyensis*.



FigureS 6.5 SNP pattern for the species *Ammonia aoteana*. Coloured boxes correspond to the most abundant hyplotype of the specimen.

Ammonia aoteana



TableS 6.1 All specimens used in this study with their illumina run code, spcies identification, DNA extraction isolate number, sampling location and date and the number of good sequences obtained.

Run	Species	Isolate	location	date	Seq
Poly	*Allogromia laticollaris*	14695	culture	2012	127955
Poly	*Allogromia laticollaris*	14697	culture	2012	75415
Poly	*Allogromia laticollaris*	12953	culture	2010	104972
Poly	*Ammonia aberdoveyensis*	243	Venice, Italy	1995	115257
Poly	*Ammonia aberdoveyensis*	257	Venice, Italy	1995	53433
Poly	*Ammonia aberdoveyensis*	450	Dovey Estuary, GBR	1997	88057
Poly	*Ammonia aberdoveyensis*	444	Dovey Estuary, GBR	1997	65221
Poly	*Ammonia aberdoveyensis*	446	Dovey Estuary, GBR	1997	93364
Poly	*Ammonia aberdoveyensis*	13162	Portugal-Estuary	2010	74349

Poly	*Ammonia aberdoveyensis*	13166	Portugal-Estuary	2010	137278
Poly	*Ammonia aberdoveyensis*	157	Cape Cod, USA	2000	51722
Poly	*Ammonia aberdoveyensis*	162	Cape Cod, USA	2000	109191
Poly	*Ammonia aberdoveyensis*	163	Cape Cod, USA	2000	87136
Poly	*Ammonia aberdoveyensis*	297	Camargue, France	1994	68368
Poly	*Ammonia aberdoveyensis*	305	Camargue, France	1994	92544
Poly	*Ammonia aberdoveyensis*	306	Camargue, France	1994	39938
Poly	*Ammonia aberdoveyensis*	266	Triest, Italy	1995	48558
Poly	*Ammonia aberdoveyensis*	267	Triest, Italy	1995	51778
Poly	*Ammonia aberdoveyensis*	421	Plymouth, GBR	1999	78805
Poly	*Ammonia aoteana*	520	La Ligua, Chile	1996	54014
Poly	*Ammonia aoteana*	521	La Ligua, Chile	1996	26945
Poly	*Ammonia aoteana*	110	Akaroa, NZ	2000	40968
Poly	*Ammonia aoteana*	111	Akaroa, NZ	2000	25382
Poly	*Ammonia aoteana*	120	Governors Bay, NZ	2000	49370
Poly	*Ammonia aoteana*	597	Governors Bay, NZ	2000	29808
Poly	*Ammonia aoteana*	599	Governors Bay, NZ	2000	70908
Poly	*Ammonia aoteana*	654	Burril lake, AU	2001	9969
Poly	*Ammonia aoteana*	659	Grays Point, AU	2001	1041
Poly	*Ammonia T1*	641	Cuba	2001	44222
Poly	*Ammonia T1*	642	Cuba	2001	45249
Poly	*Ammonia T1*	646	Cuba	2001	47468
Poly	*Ammonia T1*	124	Waitemata, NZ	2000	124039
Poly	*Ammonia T1*	18137	Persian Gulf	2015	55754
Poly	*Ammonia T1*	18138	Persian Gulf	2015	70583
Poly	*Ammonia T1*	18141	Persian Gulf	2015	86389
Poly	*Ammonia T1*	278	Camargue, France	1994	85477
Poly	*Ammonia T1*	279	Camargue, France	1994	72670
Poly	*Ammonia T1*	283	Camargue, France	1994	37597
Poly	*Ammonia T1*	265	Triest, Italy	1995	98660
Poly	*Ammonia T1*	268	Triest, Italy	1995	74292

Poly	*Ammonia T1*	241	Venice, Italy	1995	99240
Poly	*Ammonia T1*	248	Venice, Italy	1995	77130
Poly	*Ammonia T1*	253	Venice, Italy	1995	124335
Poly	*Ammonia T1*	517	La Ligua, Chile	1996	60453
Poly	*Ammonia T1*	526	La Ligua, Chile	1996	61895
Poly	*Ammonia T1*	530	La Ligua, Chile	1996	151172
Poly	*Ammonia T1*	572	Mok Baai, NL	1999	90308
Poly	*Ammonia T1*	574	Mok Baai, NL	1999	97769
Poly	*Ammonia T1*	158	Cape Cod, USA	2000	47988
Poly	*Ammonia T1*	159	Cape Cod, USA	2000	105678
Poly	*Ammonia T1*	160	Cape Cod, USA	2000	79927
Poly	*Ammonia T1*	121	Waitemata, NZ	2000	54157
Poly	*Ammonia T1*	122	Waitemata, NZ	2000	32162
Abyssline	*Aschemonella monile* -599	18272	CCZ, eastern Pacific	2015	6096
Abyssline	*Aschemonella monile* -599	18273	CCZ, eastern Pacific	2015	6070
Abyssline	*Aschemonella monile* -599	18274	CCZ, eastern Pacific	2015	7804
Abyssline	*Aschemonella monile* -599	18275	CCZ, eastern Pacific	2015	5894
Abyssline	*Aschemonella monile* -516	18277	CCZ, eastern Pacific	2015	4822
Abyssline	*Aschemonella monile* -310	18245	CCZ, eastern Pacific	2015	1884
Abyssline	*Aschemonella monile* -79	18238	CCZ, eastern Pacific	2015	1698
Abyssline	*Aschemonella aspera* -429	18263	CCZ, eastern Pacific	2015	8275
Abyssline	*Aschemonella aspera* -430	18232	CCZ, eastern Pacific	2015	5376
Abyssline	*Aschemonella aspera* -431	18264	CCZ, eastern Pacific	2015	7051
Poly	*Buccella frigida*	6299	Chile	2006	31269
Poly	*Buccella frigida*	6300	Chile	2006	28722
Poly	*Buccella frigida*	6301	Chile	2006	34958
Poly	*Buccella peruviana*	17572	Chile	2014	57798
Poly	*Buccella peruviana*	17573	Chile	2014	44333
Poly	*Buccella peruviana*	17574	Chile	2014	47937
Poly	*Cedhagenia saltatus*	10141	Black sea	2008	130711
Poly	*Cedhagenia saltatus*	10143	Black sea	2008	59660

Poly	*Cedhagenia saltatus*	10144	Black sea	2008	77378
Poly	*Cymbaloporetta squammosa*	17889	Japan	2014	77693
Poly	*Cymbaloporetta squammosa*	17890	Japan	2014	32503
Poly	*Cymbaloporetta squammosa*	17892	Japan	2014	29409
Poly	*Elphidium macellum*	5749	Chile	2005	77876
Poly	*Elphidium macellum*	6174	Chile	2006	80587
Poly	*Elphidium macellum*	5752	Chile	2005	108682
Poly	*Epistominella exigua*	5191	Antarctica	2005	16799
Poly	*Epistominella exigua*	5222	Antarctica	2005	57999
Poly	*Epistominella exigua*	3623	Antarctica	2002	58757
Poly	*Epistominella exigua*	6928	Hakuho-maru	2006	38167
Poly	*Epistominella exigua*	6929	Hakuho-maru	2006	70385
Abyssline	*Epistominella exigua*	18609	CCZ, eastern Pacific	2015	1737
Abyssline	*Epistominella exigua*	18611	CCZ, eastern Pacific	2015	1118
Abyssline	*Epistominella exigua*	18614	CCZ, eastern Pacific	2015	1357
Abyssline	*Epistominella exigua*	18616	CCZ, eastern Pacific	2015	19826
Abyssline	*Epistominella exigua*	18621	CCZ, eastern Pacific	2015	16561
Abyssline	*Epistominella exigua*	18622	CCZ, eastern Pacific	2015	3951
Poly	*Lacogromia cassipara*	18990	freshwater culture	2016	95086
Poly	*Lacogromia cassipara*	18992	freshwater culture	2016	189550
Poly	*Lacogromia cassipara*	18993	freshwater culture	2016	136816
Poly	*Leptohalysis scotti*	12287	Aarhus	2010	61552
Poly	*Leptohalysis scotti*	12290	Aarhus	2010	30722
Poly	*Lieberkühnia sp*	19189	freshwater culture	2016	78217
Poly	*Lieberkühnia sp*	19191	freshwater culture	2016	77454
Poly	*Lieberkühnia sp*	19192	freshwater culture	2016	46440
Poly	*Micrometula*	17424	Rothera, Antarctic	2013	81943
Poly	*Micrometula*	17425	Rothera, Antarctic	2013	52633
Poly	*Micrometula*	7602	Ushuaia	2007	74443
Poly	*Micrometula*	7603	Ushuaia	2007	63900
Poly	*Micrometula*	7605	Ushuaia	2007	123921

Poly	*Notorotalia finlayi*	16467	New Zealand	2012	39364
Poly	*Notorotalia finlayi*	17469	New Zealand	2013	36938
Poly	*Notorotalia finlayi*	17470	New Zealand	2013	43325
Abyssline	*Nuttalides umbonifera*	18650	CCZ, eastern Pacific	2015	14908
Abyssline	*Nuttalides umbonifera*	18651	CCZ, eastern Pacific	2015	16387
Abyssline	*Nuttalides umbonifera*	18123	CCZ, eastern Pacific	2013	5892
Abyssline	*Oridorsalis umbonatus*	18615	CCZ, eastern Pacific	2015	1123
Abyssline	*Oridorsalis umbonatus*	18617	CCZ, eastern Pacific	2015	15962
Abyssline	*Oridorsalis umbonatus*	18627	CCZ, eastern Pacific	2015	11120
Abyssline	*Oridorsalis umbonatus*	18629	CCZ, eastern Pacific	2015	5449
Abyssline	*Oridorsalis umbonatus*	18653	CCZ, eastern Pacific	2015	21265
Abyssline	*Oridorsalis umbonatus*	18656	CCZ, eastern Pacific	2015	2684
Abyssline	*Oridorsalis umbonatus*	18658	CCZ, eastern Pacific	2015	10992
Abyssline	*Oridorsalis umbonatus*	18659	CCZ, eastern Pacific	2015	6187
Abyssline	*Oridorsalis umbonatus*	17661	CCZ, eastern Pacific	2013	10085
Poly	*Planorbulinella sp.*	17893	Croatia	2014	73676
Poly	*Planorbulinella sp.*	17894	Croatia	2014	38845
Poly	*Planorbulinella sp.*	17895	Croatia	2014	59201
Poly	*Psammophaga magnetica*	8010	Antarctica	2007	117449
Poly	*Psammophaga magnetica*	8149	Antarctica	2007	160167
Poly	*Psammophaga magnetica*	8091	Antarctica	2007	39199
Poly	*Psammophaga magnetica*	3790	Antarctica	2003	118586
Poly	*Rosalina sp.*	17595	Chile	2014	48408
Poly	*Rosalina sp.*	17596	Chile	2014	49595
Poly	*Rosalina sp.*	17597	Chile	2014	39584
Abyssline	*Semipsammina sp1* - 456	18265	CCZ, eastern Pacific	2015	6006
Abyssline	*Semipsammina sp1* - 827	19127	CCZ, eastern Pacific	2015	8911
Abyssline	*Semipsammina sp1* - 827	19129	CCZ, eastern Pacific	2015	3189

TableS 6.2 Filtering process of the illumina runs used in this study

Statistics parameter	**Poly1**	**Poly2**	**Poly3**	**Abyss1**	**Abyss2**	**Abyss3**
Total number of reads	5262522	4696928	3471080	705508	1022173	906180
Reject ambiguous forward	0	0	0	0	0	0
Reject ambiguous reverse	0	0	0	0	0	0
Low mean quality forward	234461	217854	135614	22215	96976	25043
Low mean quality reverse	379625	286301	197652	37393	87801	37670
Low mean quality contig	6	4	21	0	0	0
Low base quality contig	497380	413806	341997	13582	60320	20964
Not enough matching contig	4976	4452	2902	20437	8300	9635
No primers forward	253611	247872	198671	38556	29271	49871
No primers reverse	201977	200906	155549	30933	15920	42302
Mismatch found in primers	77406	64394	48442	224151	294526	252263
Insufficient sequence length (dimers)	183	135	176	0	0	0
Total number of good reads	3612897	3261204	2390056	318241	429059	468432
Number of ISU		28718			5036	
Total number of samples		108			249	
Number of samples used		99			31	

CHAPTER 7

ENVIRONMENTAL MONITORING: INFERRING THE DIATOM INDEX FROM NEXT-GENERATION SEQUENCING DATA

JOANA AMORIM VISCO, LAURE APOTHÉLOZ-PERRET-GENTIL, ARIELLE CORDONNIER, PHILIPPE ESLING, LOIC PILLET, JAN PAWLOWSKI

Published in Environmental Science & Technology, **49**, 7597–7605, 2015

7.1. Project description

During the FISH experiments, diatoms sometimes gave strange positive results both in freshwater and marine sediments with foraminifera specific probes. One of the things that I did to investigate these results was to isolate those diatoms morphospecies, extract their DNA and try to amplify foraminifera's DNA from them. Diatoms never amplified, suggesting that the FISH results with diatom were false positive, but I develop certain skills for manipulation them under the microscope. At this time, Joana Visco, master student in the lab, started collaboration with the Service of Water Ecology of Geneva (Arielle Cordonier) and INRA Thonon (Agnès Bouchez and Frédéric Rimet) to investigate the potential of using diatoms molecular data for bioindication. I partly supervised her work trying to develop the diatom barcode database without cultivation by using single-cell approach. The diatoms were first isolated on scanning electron microscopy (SEM) stubs, photographed and then picked up again for molecular barcoding. The success rate was very low because most of specimens were lost during all the manipulation. Moreover, the diatoms were not always identifiable on SEM pictures because of their orientation and presence of cytoplasm that obscured details of frustules. The few specimens that succeed were common species already barcoded so this approach was abandoned. Finally for the published study, Joana did all the lab work and I performed all the molecular analysis presented in the paper.

7.2. Abstract

Diatoms are widely used as bio-indicators for the assessment of water quality in rivers and streams. Classically, the diatom biotic indices are based on the relative abundance of morphologically identified species weighted by their autoecological value. Obtaining such indices is time-consuming, costly and requires excellent taxonomic expertise, which is not always available. Here we tested the possibility to overcome these limitations by using a high-throughput sequencing (HTS) approach to identify and quantify diatoms found in environmental DNA and RNA samples. We analysed 27 river sites in the Geneva area (Switzerland), in order to compare the values of the Swiss Diatom Index (DI-CH) computed either by microscopic quantification of diatom species or directly from HTS data. Despite gaps in the reference database and variations in relative abundance of analysed species, the diatom index shows a significant correlation between morphological and molecular data indicating similar biological quality status for the majority of sites. This proof-of-concept study demonstrates the validity of HTS approach for identification and quantification of diatoms in environmental samples, opening new avenues towards the routine application of genetic tools for bioassessment and biomonitoring of aquatic ecosystems.

7.3. Introduction

Diatoms are phototrophic protists common in all aquatic ecosystems and widely used as bio-indicators of environmental conditions, particularly in rivers and streams (Stevenson *et al.* 2010; Rimet 2012). The applications of diatoms as bio-indicators range from routine monitoring of water quality to the assessment of industrial pollution impact (Belore *et al.* 2002; Lobo & Callegaro 2003; Poulíčková *et al.* 2004; Martin & Reyes Fernandez 2012). Because diatoms are highly sensitive to environmental conditions and grow rapidly, they respond quickly to changes in chemical, physical or biological factors. Hence, analysing the composition of their communities provides an easy method to detect environmental changes due to natural or anthropogenic causes.

Various biotic indices have been developed to assess environmental impact using diatoms (Kelly *et al.* 2009). Most of these indices are based on the relative frequency

of species weighted by their autoecological value and eventually other index-specific factors. In Europe, the Water Framework Directive (Directive 2000/60/EC) recommends using diatoms to assess water quality, but the computation of diatom indices vary from one country to another (Rimet 2012). In Switzerland, the Swiss Diatom Index (DI-CH) was proposed in order to characterise the biological status of rivers and streams using the frequencies and distributions of more than 400 diatom species and morphological varieties (Hürlimann & Niederhauser 2007). The DI-CH classifies watercourses into 5 categories, corresponding to *very good, good, average, poor* and *bad* degree of pollution, as established by the Swiss Federal Council in the Waters Protection Ordinance (Swiss Federal Council 1998).

The DI-CH is calculated as follows

$$DI - CH = \frac{\sum_{i=1}^{n} D_i G_i H_i}{\sum_{i=1}^{n} G_i H_i}$$

Where D_i is the factor based on the autoecological value for taxon i, G_i is the weighting factor for taxon i, H_i is the relative frequency of taxon i in a studied sample (number of valves found for the taxon i divided by the total number of valves counted) and n is the total number of taxa found in a sample.

The main limitation of all other diatom indices is related to the species identification being based on morphology. Indeed, diatoms constitute one of the most specious groups of protists with a number of species estimated to nearly 200,000 (Mann & Droop 1996). However, most freshwater diatoms are small (usually < 50 µm) and their microscopic identification requires special sample preparation methods and expert taxonomic knowledge. The size, shape and design of diatom valves are the main features used for taxonomic identification of diatom species. Yet, intra-specific variability can be very high and some morphological characters can become indistinct as a result of size reduction during the life cycle. In some cases, the morphological differences between species are so subtle that even trained taxonomists may come to different conclusions (Mann *et al.* 2010).

Over the past decade, molecular barcoding has become widely recognized as an efficient tool for species identification. This approach is based on the assumption that a short DNA sequence (DNA barcode) contains enough information to distinguish species. The main advantage of using DNA barcodes in applied studies is that

standardization and automation of the protocols is easier than in the traditional morphology-based approach. Several diatom barcoding studies have been performed based mainly on the analysis of five genes: COI (Evans *et al.* 2007; Evans & Mann 2009), the *rbc*L gene (Hamsher *et al.* 2011; MacGillivary & Kaczmarska 2011), the ITS region (Moniz & Kaczmarska 2009, 2010), the V4 region of the 18S rDNA (Zimmermann *et al.* 2011; Luddington *et al.* 2012), and the D2/D3 region of the LSU rRNA gene (Hamsher *et al.* 2011). Although there is no consensus on the ideal diatom DNA barcode, it has been proposed that some highly discriminating barcodes (ITS, COI) are more suitable for taxonomic studies, while those that are less variable but more universal (18S, *rbc*L) are more appropriate for applied studies (Mann *et al.* 2010).

Recent developments of high-throughput sequencing (HTS) technologies offer the possibility to use molecular barcoding for fast and reliable diversity surveys based on environmental samples. HTS-based environmental monitoring has been proposed as a time and cost-effective alternative to the traditional morphology-based approaches (Baird & Hajibabaei 2012; Taberlet *et al.* 2012; Bohmann *et al.* 2014). Several experimental studies have been conducted on HTS-based inventories of freshwater benthic macroinvertebrates (Hajibabaei *et al.* 2011, 2012; Carew *et al.* 2013). Previous studies focusing specifically on diatoms completed their taxonomic reference database, evaluated different DNA barcodes, and compared the composition of diatom communities inferred from microscopic and HTS data (Kermarrec *et al.* 2013, 2014, Zimmermann *et al.* 2014, 2015). One of these studies also briefly compared the diatom indices computed from morphological and molecular data (Kermarrec *et al.* 2014), although this aspect has still not been thoroughly examined up to now.

Here, we test the accuracy of water quality assessment through the HTS-based diatom biotic index. To do so, we analyse the diatom communities in 27 watercourses of the Geneva basin. We use the hypervariable region V4 of 18S rDNA as diatom DNA barcode and Illumina® Miseq platform for high-throughput sequencing. We taxonomically assign HTS reads using a reference database and phylogenetic analyses in order to find the best match between morphological and molecular data. We compute the DI-CH values for each site using the relative abundance of sequences found for each taxon and compare them with the values inferred from microscopic study.

7.4. Materials and methods

7.4.1. Sampling.

The samples were collected in 2013-14 as part of a routine bioassessment campaign performed by the Service of Water Ecology (SECOE) of the Department of Environment, Transport and Agriculture in Geneva, Switzerland (Cordonier *et al.* 2010). The biofilm containing epilithic diatoms was collected in 27 sites located in shallow waterways of the Geneva basin following the directives established by the Swiss Federal Office for the Environment (Hürlimann & Niederhauser 2007) (TableS 7.1). Between three to five stones were selected at each sampling site. The periphyton taken by scratching the stones with diatom-scraping devices was resuspended with freshwater taken from the river and then transferred to sampling bottles. Each sample was homogenized and divided into two subsamples, one for morphological analysis by the SECOE and the other for molecular analysis. Morphological samples were preserved in a concentrated (37%) formaldehyde solution, while molecular samples were kept cold (ca. 0°C) during sampling (max. 4 hours). Upon arrival to the laboratory, 1 ml of homogenized periphyton suspension was transferred to 1.5 ml tubes and centrifuged at 8000g for 10 minutes. Supernatant was discarded and pellets stored at -80°C until DNA/RNA extractions.

7.4.2. Morphological analysis.

Sample preparation, species identification, counting and DI-CH calculations were performed as recommended by the Swiss Federal Office for the Environment (Hürlimann & Niederhauser 2007). Periphyton suspensions were sorted and undesirable material was discarded. A decarbonation step using hydrochloric acid was performed, followed by the elimination of organic material by calcination combined with a treatment with hydrogen peroxide. Diatoms were then washed and mounted in Naphrax. Diatoms slides were observed by using an Olympus light microscope with Nomarski differential interference contrast optics at a magnification of 1000x. Species identification was performed with the bibliographic support of The Flora of Diatoms (Krammer & Lange-Bertalot 1986), Diatoms of Europe (Lange-Bertalot 2001), Iconographia Diatomologica (Lange-Bertalot & Metzeltin 1996; Reichardt 1999), and Diatomeen im Süsswasser-Benthos von Mitteleuropa (Hofmann *et al.* 2011).

7.4.3. DNA/RNA extraction.

DNA and RNA were extracted with PowerBiofilm® DNA and RNA isolation kits (MO BIO Laboratories Inc.) following the manufacturer instructions. RNA was purified from carried-over DNA molecules with TURBO DNase™ kit Ambion® (Life Technologies) and cDNA obtained by reverse transcription using SuperScript® III Reverse Transcriptase kit (Invitrogen™). A total of 27 DNA and 27 cDNA (RNA) samples were obtained for this study.

For the extraction of cultured diatoms, pelleted cells were prepared by centrifuging 1ml of fresh diatoms cultures at 8000 g for 10 minutes. The extractions were then performed with DNeasy® Plant Mini Kit (Qiagen) or PowerBiofilm® DNA isolation (MO BIO).

7.4.4. Reference Database.

We built a reference database of the V4 region composed of 460 unique diatom sequences. First, we downloaded from the GenBank database all sequences corresponding to the species and genera found in the morphological analyses of Geneva samples and also those commonly found in Switzerland (Hürlimann & Niederhauser 2007). The alignment was performed with the Seaview program (Gouy *et al.* 2010). Sequences were analysed by Maximum Likelihood (ML) phylogenetic inference and those showing incorrect identification were discarded. A total of 298 unique sequences from GenBank were kept.

To extend our reference database we sequenced 10 diatom species obtained from culture collections: *Fragilaria pinnata* and *Nitzschia ovalis* from the CCAP (Culture Collection of Algae and Protozoa, SAMS Research Services Ltd, Scottish Marine Institute, Oban, UK, http://www.ccap.ac.uk), *Achnanthidium minutissimum, Achnanthidium pyrenaicum, Achnanthidium straubianum, Amphora pediculus, Cocconeis placentula, Encyonema silesiacum, Nitzschia palea* and *Sellaphora seminulum* from the TCC (Thonon Culture Collection, INRA-UMR Carrtel, Thonon-les-Bains, France, http://www6.inra.fr/carrtel-collection). We also added 152 Sanger sequences from other eDNA analyses of Geneva watercourses.

7.4.5. PCR amplification, cloning and Sanger sequencing.

To complete the reference database and to test the specificity of PCR primers, the diatom cultures and environmental samples cited above were examined. The hypervariable region V4 of the 18S rRNA gene was amplified using primers modified after Zimmermann *et al.* (2011). PCR amplifications were performed in a total volume of 25µl using Taq DNA Polymerase by Roche Applied Science. PCR regime included an initial denaturation at 94°C for 2 min, then 35 cycles of denaturation at 94°C for 45 s, annealing at 50°C for 45 s, elongation at 72°C for 1 min and a final elongation at 72°C for 10 min. PCR amplicons were purified with High Pure PCR Product Purification kit (Roche Applied Science) and cloned using TOPO® TA Cloning® kit for sequencing (Invitrogen™). Sequence reactions were performed with BigDye® Terminator (Applied Biosystems), and sequences were obtained by Sanger sequencing on ABI PRISM 3130XL Genetic Analyser System (Applied Biosystems/Hitachi).

7.4.6. PCR amplification for next-generation sequencing.

PCR were performed on DNA and RNA (cDNA) isolated from periphyton samples using unique combinations of forward and reverse tagged primers. Individual tags are composed of 8 nucleotides attached at each primers 5'- extremities. A total of 20 different forward and reverse tagged primers were designed to enable multiplexing of all PCR products in a unique sequencing library. PCRs were performed as described above. Purified PCR products were quantified by fluorometric method using QuBit HS dsDNA kit (Invitrogen). Concentrations were then calculated and normalized for all samples. Approximately 50ng of amplicons of each DNA and RNA samples from the SECOE 2013 (DIATOM 2013) and 2014 (DIATOM 2014) campaigns were pooled. An amount of 100ng of pooled amplicons was used for Illumina library preparation.

7.4.7. Illumina library preparation and sequencing.

Indexed paired-end libraries of pooled amplicons for consecutive cluster generation and DNA sequencing were constructed using Illumina TruSeq® Nano DNA Sample Preparation Kit – Low Throughput. Libraries were prepared following the manufacturer instructions. The fragment sizes of each library were verified by loading 3µl of the final product in a 1.5% agarose gel with 1x SYBR®Safe (Invitrogen) and

quantified by fluorometric method using QuBit HS dsDNA kit (Invitrogen). MiSeq Reagent Nano kit v2, 500 cycles with nano (2 tiles) flow cells were used to run libraries on MiSeq System. Two 250 cycles were used for an expected output of 500Mb and an expected number of 1 million reads per library.

7.4.8. HTS data analysis.

Operational Taxonomic Units (OTUs) were obtained and assigned following the method described in Pawlowski *et al.* (2014a) by using the diatoms reference database described above. Raw FASTQ reads were quality-filtered with extremely stringent parameters keeping only high-quality reads. Then, paired-end reads were assembled by aligning them into a contiguous sequence with highest similarity. In case of mismatching bases, we kept in the final contig the closest base from the read 5'- extremity, based on the fact that the probability of miscalls increases towards the 3'- extremity. These sequences were then de-multiplexed (assigned to their corresponding sample) depending on the tagged primers found at each end. De-replication of the dataset obtained after assembly was necessary in order to obtain unique sequences, called Independent Sequence Units (ISUs). An abundance threshold of 10 was used for the minimum number of replicates found for each ISU, and this abundance was recorded for further analyses. Subsequently, ISUs were assigned by performing a pairwise Needleman-Wunsch global alignment against our entire reference database. For the ISUs that were not assigned at the end of this procedure, we relied on a BLAST filtering procedure. We removed the ISUs that did not match any Bacillariophyceae sequences in the NCBI database with at least 99% coverage and 97% identity.

7.4.9. Phylogenetic analyses.

The taxonomic assignment of OTUs was checked by phylogenetic analyses. A tree was built with all the sequences from the database and the OTUs from the HTS analysis. The most abundant ISU was used as the representative sequence for each OTU. The ML phylogeny was constructed using RAxML (Stamatakis 2014), with GTR + G as model of evolution and 1000 replicates for the bootstrap analysis. The OTUs were assigned to the reference morphospecies if they formed a clade supported by bootstrap values > 60.

7.5. Results

7.5.1. HTS data statistics.

For DIATOM 2013, we obtained 1,176,424 reads from Illumina sequencing (TableS 7.2). The filtering process rejected 169,841 reads with low mean quality, 61,508 reads with low base quality, 2,205 reads with not enough matching bases in the contig region and 177,325 reads with errors or mismatches in the primers. Hence, a total of 765,545 reads remained after filtering and were available for further analysis. For DIATOM 2014, we obtained 1,055,387 reads. The filtering process rejected 296,799 reads with low mean quality, 17,095 reads with low base quality, 152,394 reads with not enough matching bases in the contig region, 247,694 reads with errors or mismatches in the primers and 23,222 with insufficient sequence lengths. Hence, a total of 318,183 good reads remained for further analysis.

7.5.2. Morphological data and DI-CH calculation.

For each sampling site, about 400 valves were observed and identified with light microscopy at SECOE. Morphospecies were counted and the relative abundance of each taxon was calculated for each site (TableS 7.3). A total of 96 species was found by morphological identification. The number of taxa per site varied from 5 (AMB) to 37 (HEB). One species (*Amphora pediculus*) was found at every site and represented the most abundant taxon counted for all sites together. The values of DI-CH were calculated using the formula presented previously. The DI-CH values varied from 3.64 (NAM) to 7.98 (AMB). Highest DI-CH values were obtained for sites with larger numbers of diatoms with high autoecological values, such as *Nitzschia amphibia, Sellaphora seminulum, Eolimna minima, Gomphonema micropus, Gomphonema parvulum, Eolimna subminuscula, Navicula veneta* and *Nitzschia acicularis*.

7.5.3. Taxonomic assignment of HTS data.

Analysis of the HTS data grouped the reads into 242 OTU for the DIATOM 2013 and 103 for the DIATOM 2014 runs. In order to assign those OTUs to morphological taxa, a ML tree with all OTUs and our reference database was built. After phylogenetic analysis we removed 128 OTUs for the DIATOM 2013 run and 60 OTUs for the DIATOM 2014 run because they could not be univocally assigned to any

morphological clade. In total, 144 OTUs remained and were assigned to 30 taxa. Twenty-three of these taxa corresponded to the morphospecies found in microscopic analyses, while seven matched to species in the reference database that were not evidently found with the morphology-based approach.

Among the 23 assigned species (Figure 7.1A), 15 were confidently identified, i.e. they formed well-supported clades (BV > 60) including reference sequences assigned to a single morphospecies. One species (*Encyonema* spp.) was a special case since the only GenBank reference sequence of the clade was not identified beyond the genus level. Five species formed clades with reference sequences assigned to two different species of the same genus. These species were *Amphora pediculus, Achnanthes minutissima, Cocconeis placentula/pediculus, Mayamea atomus* and *Fistulifera saprophila*.

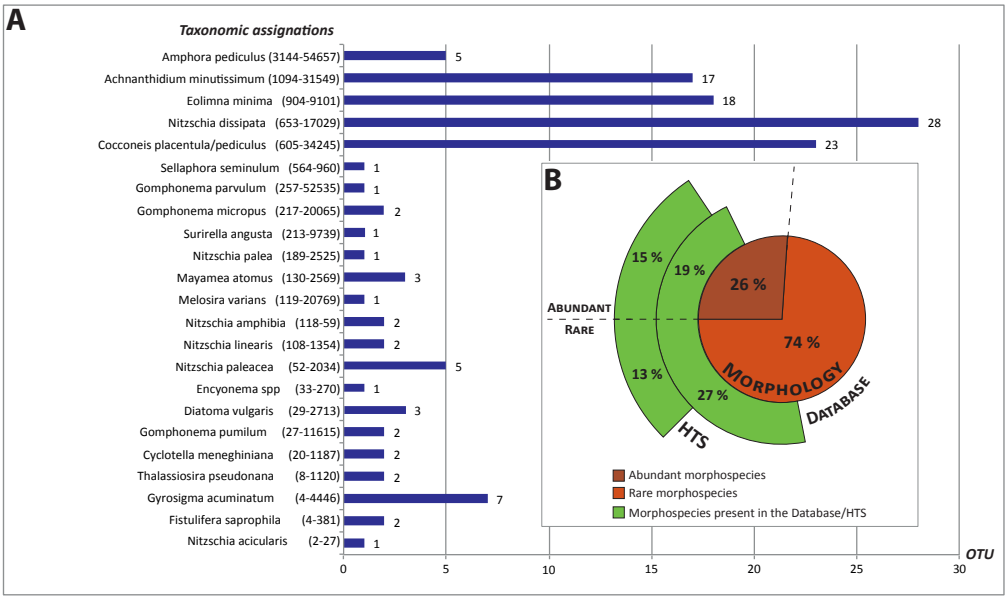


Figure 7.1 A. Taxonomic assignments in common with morphospecies sorted by the number of counts in the morphologic analysis (in parenthesis). The bar plot represent the number of OTU in each taxonomic assignation. B. Pie chart of abundant (dark red) and rare (light red) morphospecies found in morphologic analysis. Arcs in green represent the morphospecies present in the database (internal one) and in the HTS assignations (external one). Each arc is divided between abundant and rare species by a dashed line.

Two assignments were particularly problematic. The OTUs assigned to *Cyclotella meneghiniana* formed a well-supported clade (BV 78) with 8 other *Cyclotella* species, half of which were marine species. We assigned these OTUs to *C. meneghiniana*

because it was the only species present in the morphological list with an autoecological value. In the second case, the two OTUs assigned to the morphospecies *Thalassiosira pseudonana* formed a well-supported clade (BV 88) with 13 other *Thalassiosira* species and with the species *Stephanodiscus minutulus*. As both *S. minutulus* and *T.pseudonana* have the same autoecological value, we kept them together using the name of *T. pseudonana* as in morphological analyses.

In total, the number of morphospecies recognised in the HTS data amount to only 28% of all those identified in this study microscopically. However, it should be noted that the GenBank database only covers 46% of the morphospecies found in microscopic analyses (Figure 7.1B). The difference between these two percentages is accounted for by morphospecies (i.e. genus *Navicula*) that could not be identified unambiguously due to the lack of resolution of the V4 region. However, it is important to notice that most species not found in HTS were rare (below 100 counts in the morphologic analysis), as shown by the Figure 7.1B. The list of the morphospecies with their count in the morphologic analysis and their presence in the database and in the HTS assignation are reported in the TableS 7.4.

7.5.4. Abundance of assigned species.

As the calculation of diatom indices includes the relative abundance of species, we analysed the variations in morphological counts and the number of reads inferred from DNA and RNA data for each assigned species. As can be seen in Supplementary Material (TableS 7.5 and FigureS 7.1), the relative abundance of species per site varies considerably depending on the type of data. In particular, the proportion of a species in DNA samples is often lower than in morphological counts and RNA samples. We checked whether this could be a consequence of the high abundance of undetermined sequences in some samples, but the re-analysis of data with assigned OTUs only changed the proportions between DNA, RNA and morphological abundances only in few cases.

The correlation between the number of reads and individuals for the most ubiquitous and abundant species is significant for both DNA and RNA of *A. pediculus* and DNA of *A.minutissima* (Figure 7.2). The relative abundance of some species (*A. pediculus, E. minima*) is higher in morphocounts than in HTS data. However, among the assigned morphospecies, there are very few sites where the species was found in microscopic preparations but not in the HTS data. This deviation is more obvious in

less common taxa, with species such as *Nitzschia amphibia* being found almost exclusively in morphological analyses, while some species (e.g., *Gyrosigma acuminatum*) or genera (e.g. *Gomphonema*) are overrepresented in HTS data (FigureS 7.1).

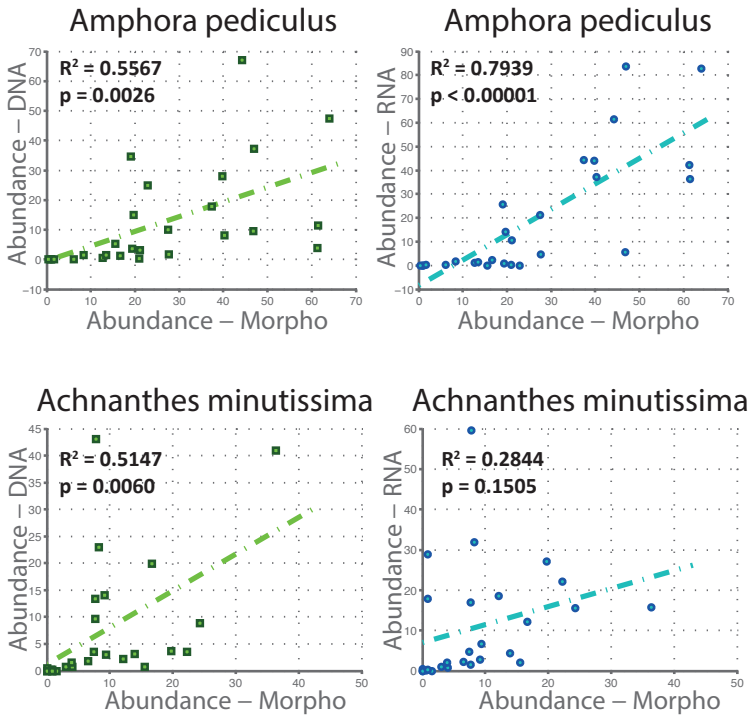


Figure 7.2 Relationships between the relative abundance of the two most abundant species *Amphora pediculus* (upper) and *Achnanthes minutissima* (lower). This information is displayed separately for DNA (left) and RNA (right) where each point shows the relationship between the relative abundance found in morphological (x-axis) or molecular (y-axis) counts. The dotted lines represent the results of model II regression with a least squares fitting for the relative abundances of all samples. The R^2 and p-value are indicated for each regression axis.

7.5.5. Diatom index.

The HTS DI-CH index was calculated with the 23 taxa, for which the D and G values were available. When those values were different for a variety or subspecies of the same species, the values of the most abundant and frequent taxa were retained. All the DI-CH values for morphology, DNA and RNA per sites are presented in TableS 7.6.

The variations in diatom indices inferred from morphological and molecular (DNA/RNA) data for 27 sites are illustrated in Figure 7.3. For the majority of sites (25 out of 27) the deviation between morphological and at least one of the molecular indices (DNA or RNA) was less than 1 unit and the biological quality status inferred from the two types of data was identical. For 17 sites (63%), the morphological index indicated the same level of water quality as at least one type of molecular data. Both DNA and RNA data were congruent with the morphological index in 7 out of 27 sites. When considered separately, the same level was indicated in 10 and 12 sites for DNA and RNA, respectively. The values of the morphological index exceeded those inferred from DNA and RNA in 16 sites (20 in the case of RNA). As we can see, the correlation between morphological and molecular indices is significant for DNA (Figure 7.4A) with R^2=0.59 and p-value = 0.0013 and becomes strongly supported in the case of RNA (Figure 7.4B) with R^2=0.85 and p-value < 0.0001.

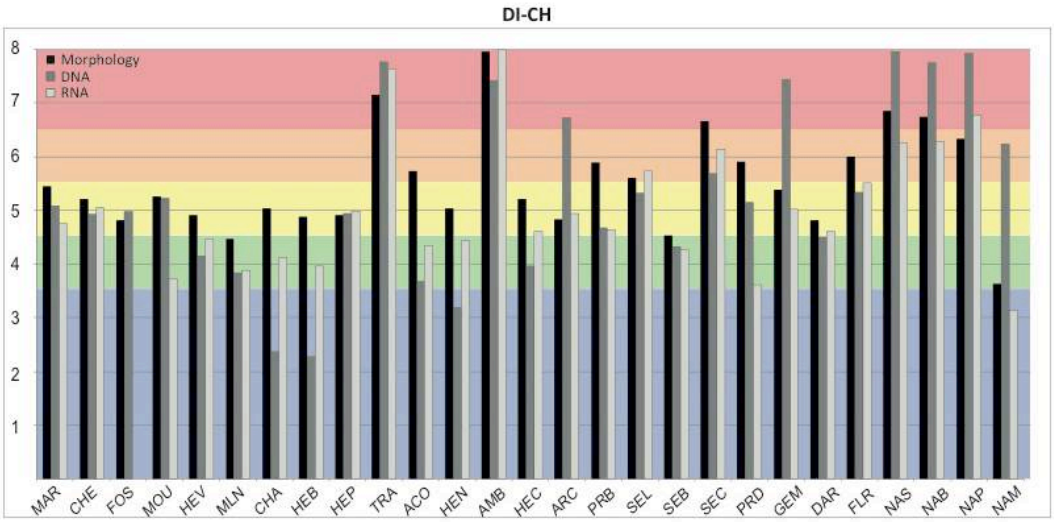


Figure 7.3 DI-CH values for morphologic analysis (black), DNA (dark grey) and RNA (light grey) per sites. Colours represent the threshold for water quality given by the DI-CH index.

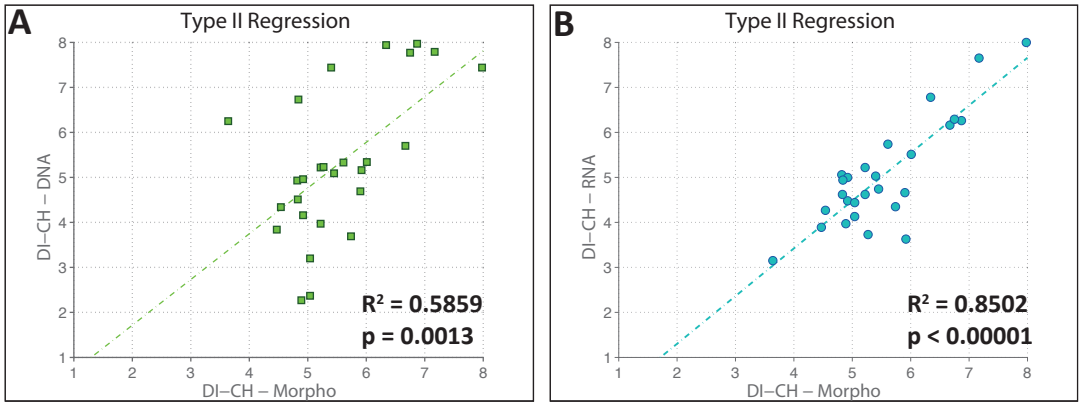


Figure 7.4 Relationships between the DI-CH inferred from morphological and DNA (A) or RNA (B) abundances per sites. Each point shows the relationship between the DI-CH found in morphological (x-axis) or molecular (y-axis) counts over all sites. The dotted lines represent the results of model II regression with a least squares fitting for the relative abundances of all samples. The R^2 and p-value are indicated for each regression axis.

7.6. Discussion

By exhibiting the strong similarity between the DI-CH values inferred from microscopic and HTS analyses of diatom communities, our proof-of-concept study clearly demonstrates the usefulness of HTS diatom data to evaluate water conditions. Our results confirm the previously reported similarity between values of the Specific Pollution Sensitivity biotic index obtained by microscopy and by HTS (pyrosequencing) analysis of SSU and *rbc*L barcodes (Kermarrec *et al.* 2014). Both studies fully support the growing evidence that HTS environmental studies have the potential to become new tools for the assessment of aquatic ecosystems health, based on analysis of benthic macroinvertebrates (Hajibabaei *et al.* 2011, 2012), diatoms (Kermarrec *et al.* 2013; Zimmermann *et al.* 2014), and other protists (Pawlowski *et al.* 2014a).

The congruence between diatom indices inferred either from morphological or HTS data is remarkable. The correlation is especially strong for RNA (Figure 7.4B), likely because it provides a better depiction of the living diatom community composition. The DNA, on the other hand, can be preserved in water for a certain period of time and even carried over long distances (Deiner & Altermatt 2014). Interestingly, the correlation between HTS and morphology in species relative abundances have

limited impact on the correlation between indices. This could be due to the fact that the index is calculated as the sum of a set of species with their respective weighting factors, which tends to reduce the effect of variations for individual species. Noticeably, the index correlates better in the sites with lower species richness, which might be related to the reduction of technical or biological biases in low complexity samples.

Although the results of our study are promising, there is still a wide potential to reduce the divergences between molecular and morphological results by addressing the current limitations of HTS data analysis. We discuss here the three major causes of these divergences: (1) database incompleteness, (2) inconsistencies between molecular and morphological taxonomy, and (3) biases in the quantitative analysis of HTS data.

7.6.1. The incompleteness of databases.

Gaps in reference databases are commonly believed to be the main hindrance to assigning taxonomy to environmental sequences. In fact, the diatom database is probably more exhaustive than that of any other groups of protists, especially those that cannot be cultivated (Pawlowski *et al.* 2012). The proportion of genetically characterized species in our study (46%) is slightly lower than in other studies targeting well-studied temperate regions (53-78%) but remains higher than those conducted in tropical regions (30-38%) (Kermarrec *et al.* 2014). The development of comprehensive databases, like that of Zimmermann *et al.* (2014) which provided molecular (V4, *rbc*L) and morphological (LM, SEM) data for 70 cultured diatom strains, is an important step towards filling the gaps in diatom inventories. However, establishing cultures of diatom species for every eco-region could be extremely time-consuming and might not always be successful. An alternative approach could be based on single-cell PCR followed or preceded by LM or SEM study (Lang & Kaczmarska 2011). The success rate of these methods is still very low, but further developments in the field of single-cell genomics might rapidly improve their efficiency.

It should be noted that, although completing the database is important, it does not imply that the sequencing of all morphospecies is necessary. In our study, we assigned species according to very stringent criteria by removing all uncertain cases. Once the reference database is completed for common species such as *Achnanthes*

lanceolata, and the identification of *Navicula* species is improved by using more rapidly evolving marker, the correlation between HTS and morphological indices might become even stronger. In fact, the vast majority of species currently missing from the database are rare with less than 100 specimens counted in all samples. Their relative importance in the computation of diatom indices depends on the autoecological value associated with each species. However, it might be sufficient to correctly assign all common species and those rare species with high autoecological value to obtain a perfect match.

7.6.2. Molecular vs morphological taxonomy.

Another potential source of conflict lies in the divergence between the morphological and molecular (phylogenetic) determination of diatom species. On the one hand, almost all morphospecies are represented by several genetically distinctive types. On the other hand, some morphospecies are subdivided into subspecies or morphological varieties, each with their own specific autoecological values. In the first case, the cryptic diversity may constitute a considerable advantage for biomonitoring, particularly if the cryptic species are associated with some specific ecological conditions. The second case is more problematic because the sub-specific taxa are generally uncharacterized genetically.

In this study, we combined all subspecies and morphotypes belonging to the same species because it was impossible to distinguish them genetically. We also combined two species of *Cocconeis*, to avoid a possible misidentification of numerous phylotypes forming the clade of *C. placentula*, among which *C. pediculus* branches. In our approach we followed the principle that the species can be grouped if they share the same ecologies and morphologies (DeNicola 2000) and if they form a clade in phylogenetic analysis. Grouping at generic level (Rimet & Bouchez 2012) may be useful, as in the case of *Encyonema*, but it is not necessary and may even be inappropriate in the case of polyphyletic genera.

Taxonomic resolution largely depends on the choice of the DNA barcode. Until now, only the chloroplastic *rbc*L and nuclear ribosomal 18S V4 region have been used in HTS diatom studies. Here, we chose the V4 region because its amplification from eDNA samples is easier and its size better fits the sequencing length of Illumina Miseq. It has been shown that the taxonomic resolution of V4 (and 18S in general) is lower than *rbc*L (Kermarrec *et al.* 2013). However, the inter-species variation of a

given barcode may change between genera, and its efficiency will depend on the taxonomic composition of diatom community. For example, in our study, the resolution of V4 was too low to unambiguously assign *Navicula* species, but it was sufficient to distinguish most of the species of *Nitzschia* and *Gomphonema*. Ideally, as both V4 and *rbc*L barcodes are complementary they should be used together in HTS analyses.

7.6.3. Relative abundance.

Undoubtedly, the quantitative analysis of HTS data presents the greatest challenge in efforts to alleviate biases in the calculation of diatom indices. Indeed, numerous HTS environmental surveys exhibited discrepancies between the number of sequences assigned to a given species and the number of specimens of the same species in microscopic preparations (Nolte *et al.* 2010; Stoeck *et al.* 2014) or even mock communities (Amend *et al.* 2010). This lack of correlation between the abundance of reads and individuals could be explained either by technical biases introduced during DNA extraction, PCR amplification or sequencing (Pawlowski *et al.* 2014b), or by biological factors such as the variations of rRNA gene copies (Weber & Pawlowski 2013), which may depend on genome size (Prokopowich *et al.* 2003), number of nuclei (Heyse *et al.* 2010), or differences in cell size (Godhe *et al.* 2008).

Our study shows that molecular and morphological counts are well correlated in some species, but differ significantly in others (Figure 7.2). These variations seem taxon-specific and could be explained by variation in the numbers of rRNA gene copies in different diatom species. However, the ground-truth biological data necessary to test such a hypothesis are not available for diatoms. In fact, the correlation between molecular and morphological abundance data was previously observed in the HTS study of changes in foraminiferal communities associated with the environmental impact of fish-farming (Pawlowski *et al.* 2014a), as well as in the study of the seasonal abundance in some species of ciliates and chrysophytes (Medinger *et al.* 2010). As the match between microscopic and molecular abundances concerns mainly the abundant species, this could explain why the impact of abundance variations on the final computation of the diatom index is relatively moderate.

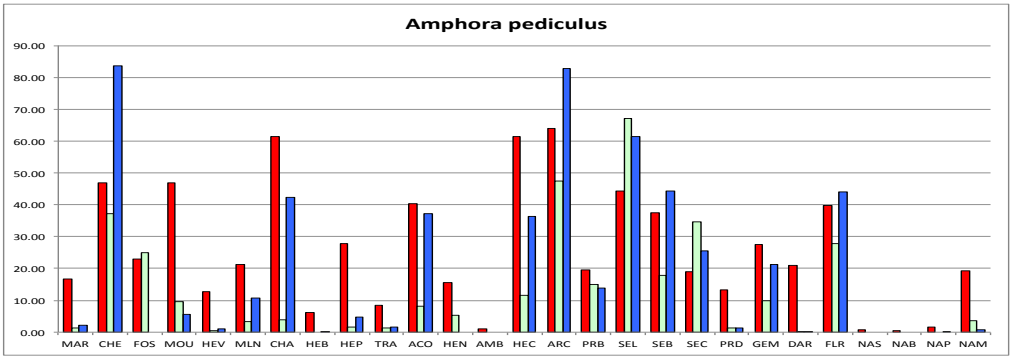
7.6.4. Future perspectives.

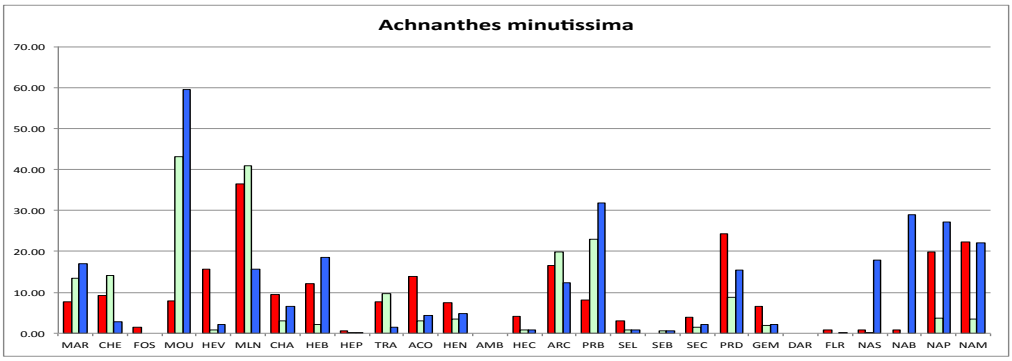
The further development of HTS-based surveys of diatom diversity will require substantial efforts by diatom taxonomists and biologists to complete the DNA barcoding reference database and to determine the range of genetic and morphological variation in diatom species. Better knowledge of diatom genomes, especially the quantification of nuclear and chloroplast genes copies, will help improving the estimation of species abundance from molecular data. Additional HTS studies of diatom communities in different ecological settings are also needed in order to optimize the molecular protocols and improve the accuracy of HTS data analysis.

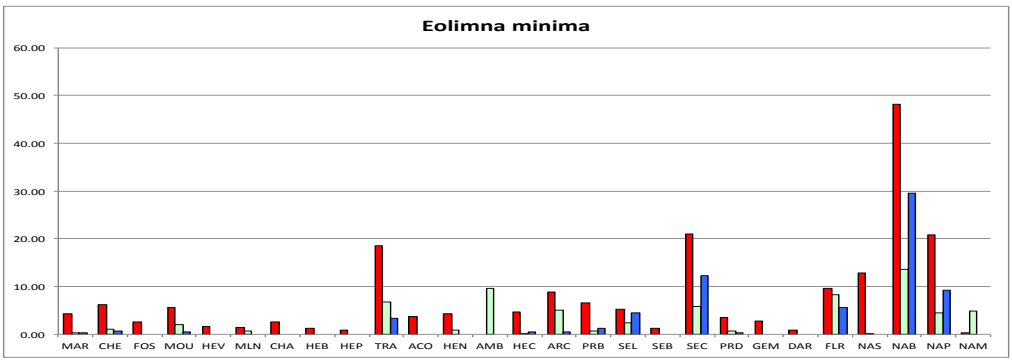
All these efforts are worthwhile considering the tremendous benefits that the routine application of HTS approaches would bring to diatom-based monitoring. First, the use of DNA barcodes will allow standardization of species identification, which will help overcoming the recurrent problems of misidentification and will facilitate the comparison of species inventories. Second, the molecular approach will provide more accurate real-time assessment of living communities, especially if RNA is analysed rather than DNA. Third, the use of HTS technology coupled with the automation of molecular protocols will considerably reduce the time for sample processing, which will, in turn, allow an increase in the number of monitored sites. Finally, given the rapidly diminishing costs of HTS technologies, the application of these new tools will allow important savings.

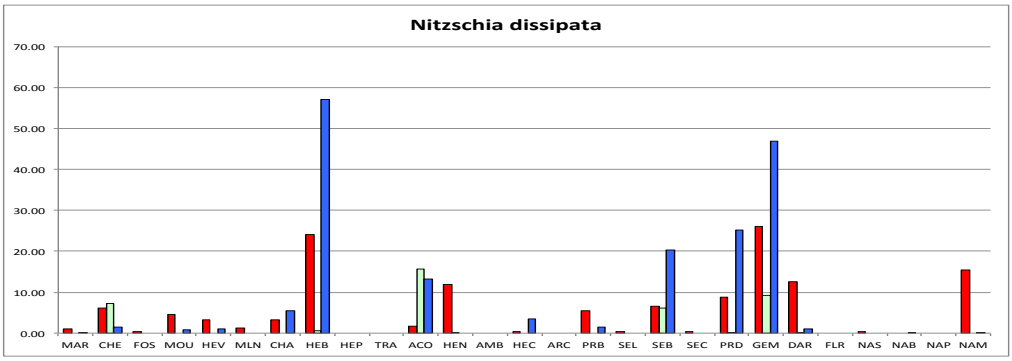
7.7. Supplementary data

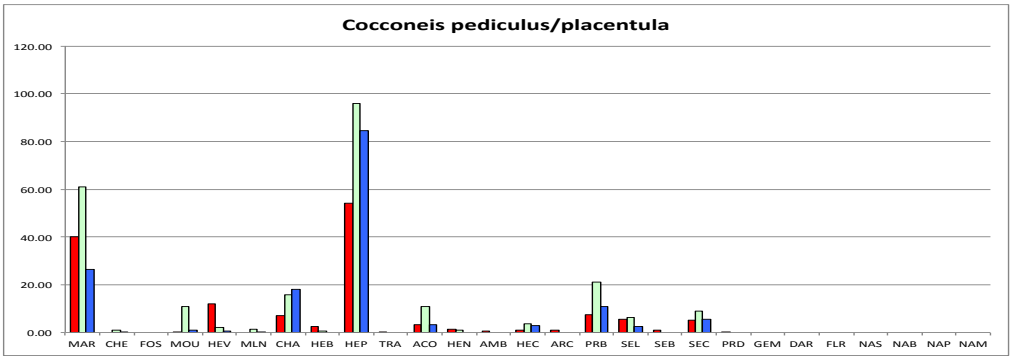
FigureS 7.1 Relative abundance of 23 assigned taxa inferred for morphology (red), DNA (light green) and RNA (blue).

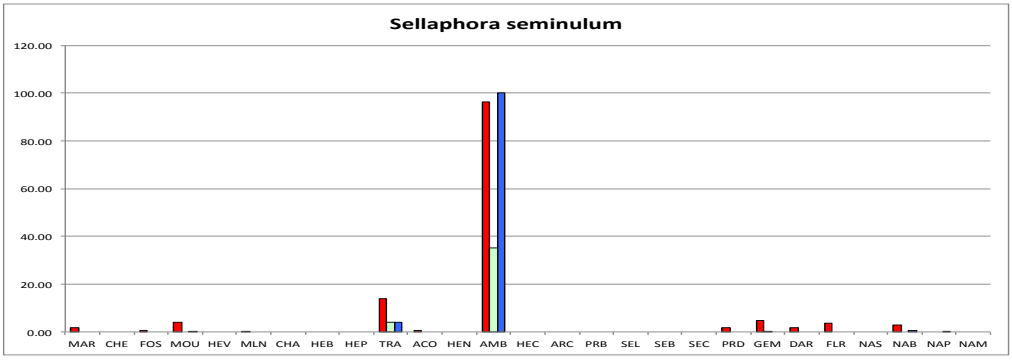


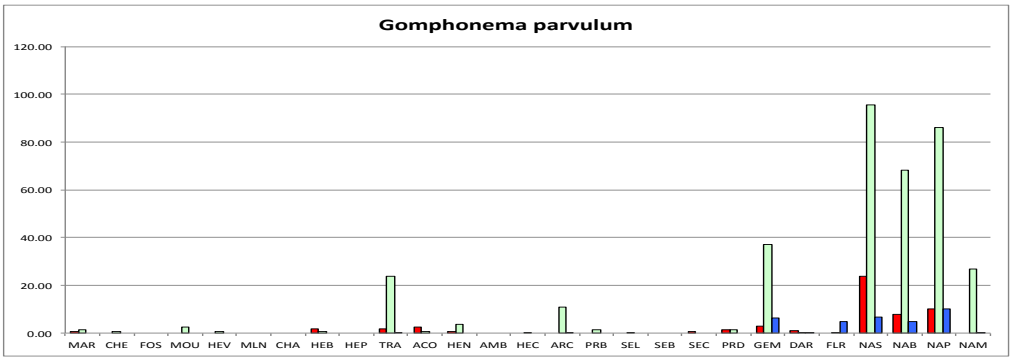


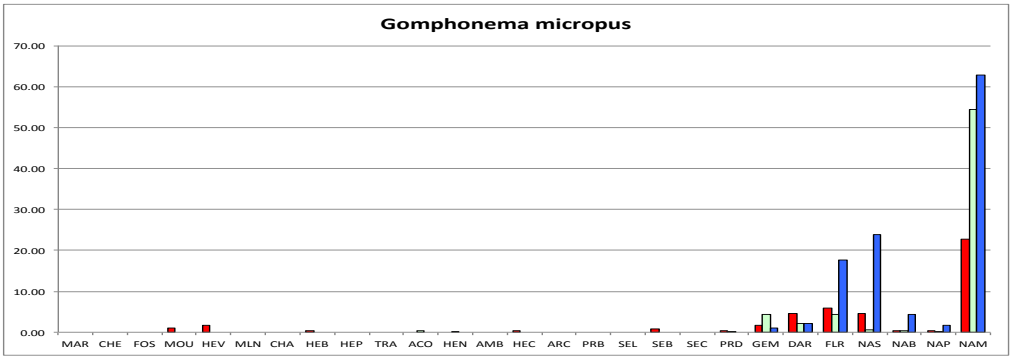


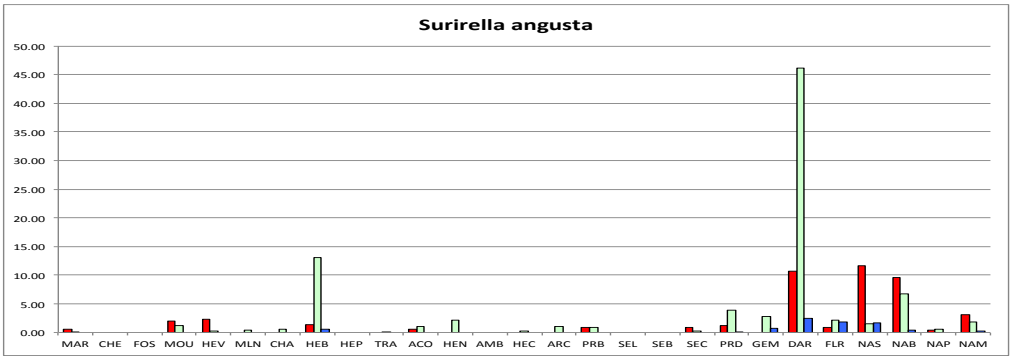


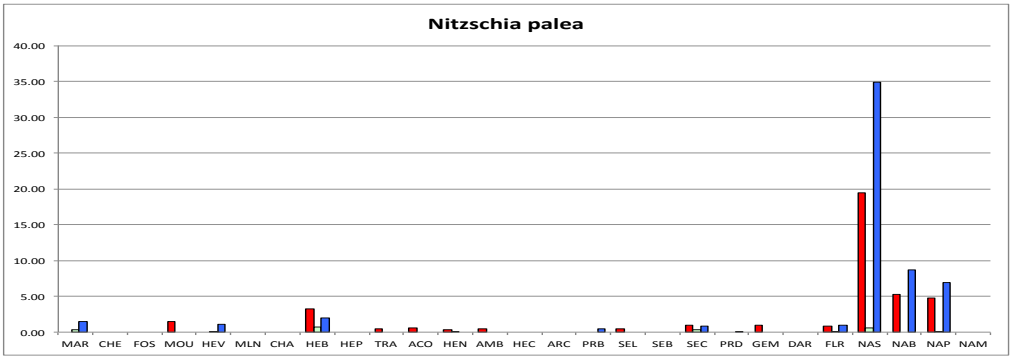

Cocconeis pediculus/placentula
120.00
100.00
80.00
60.00
40.00
20.00
0.00
MAR CHE FOS MOU HEV MLN CHA HEB HEP TRA ACO HEN AMB HEC ARC PRB SEL SEB SEC PRD GEM DAR FLR NAS NAB NAP NAM

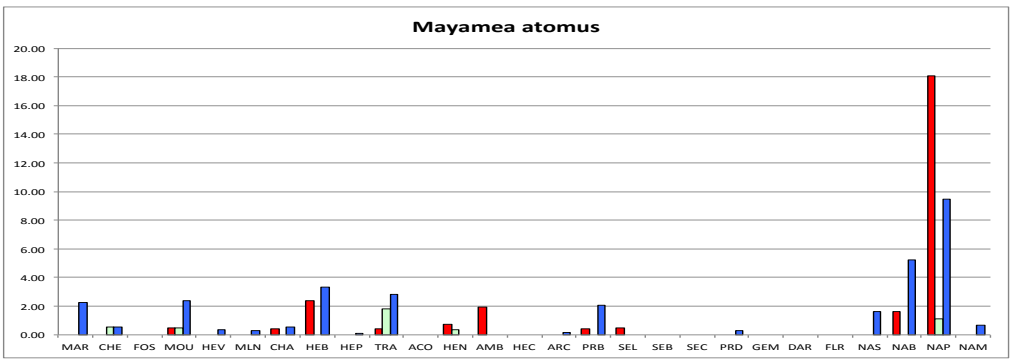

Sellaphora seminulum
120.00
100.00
80.00
60.00
40.00
20.00
0.00
MAR CHE FOS MOU HEV MLN CHA HEB HEP TRA ACO HEN AMB HEC ARC PRB SEL SEB SEC PRD GEM DAR FLR NAS NAB NAP NAM

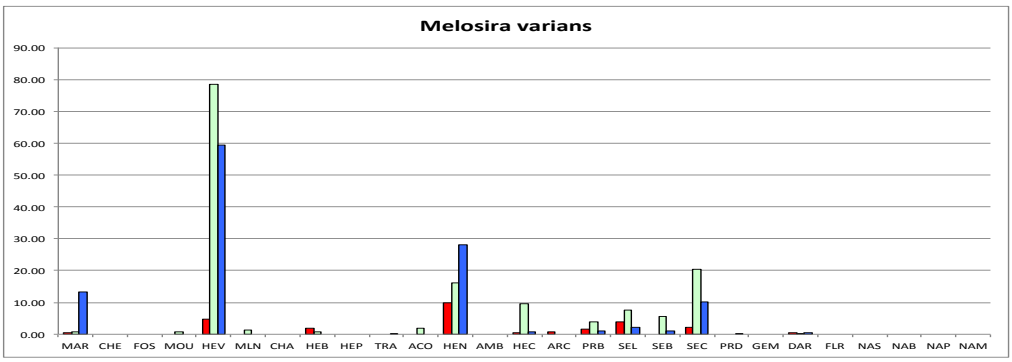

Gomphonema parvulum
120.00
100.00
80.00
60.00
40.00
20.00
0.00
MAR CHE FOS MOU HEV MLN CHA HEB HEP TRA ACO HEN AMB HEC ARC PRB SEL SEB SEC PRD GEM DAR FLR NAS NAB NAP NAM

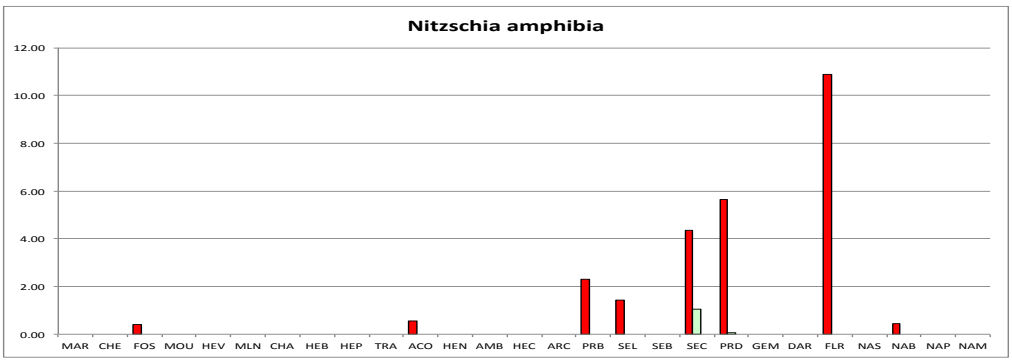

Gomphonema micropus
70.00
60.00
50.00
40.00
30.00
20.00
10.00
0.00
MAR CHE FOS MOU HEV MLN CHA HEB HEP TRA ACO HEN AMB HEC ARC PRB SEL SEB SEC PRD GEM DAR FLR NAS NAB NAP NAM

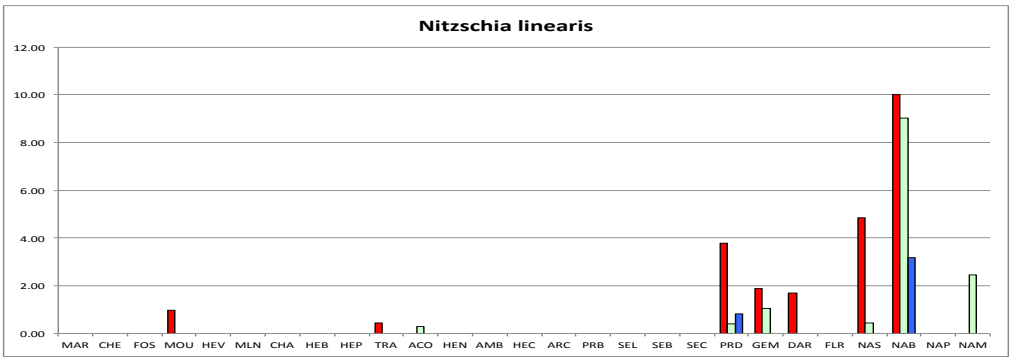

Surirella angusta
50.00
45.00
40.00
35.00
30.00
25.00
20.00
15.00
10.00
5.00
0.00
MAR CHE FOS MOU HEV MLN CHA HEB HEP TRA ACO HEN AMB HEC ARC PRB SEL SEB SEC PRD GEM DAR FLR NAS NAB NAP NAM

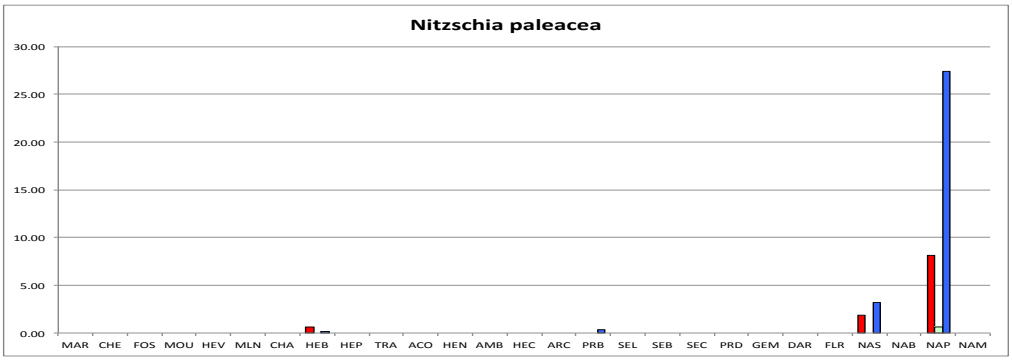

Nitzschia palea
40.00
35.00
30.00
25.00
20.00
15.00
10.00
5.00
0.00
MAR CHE FOS MOU HEV MLN CHA HEB HEP TRA ACO HEN AMB HEC ARC PRB SEL SEB SEC PRD GEM DAR FLR NAS NAB NAP NAM

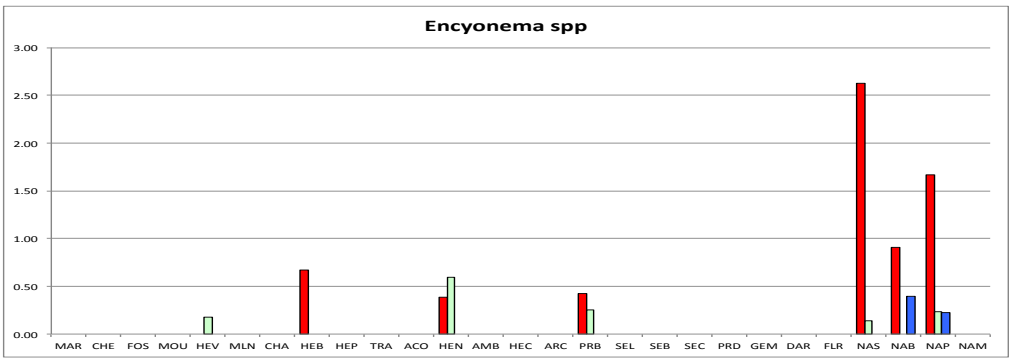

Mayamea atomus
20.00
18.00
16.00
14.00
12.00
10.00
8.00
6.00
4.00
2.00
0.00
MAR CHE FOS MOU HEV MLN CHA HEB HEP TRA ACO HEN AMB HEC ARC PRB SEL SEB SEC PRD GEM DAR FLR NAS NAB NAP NAM

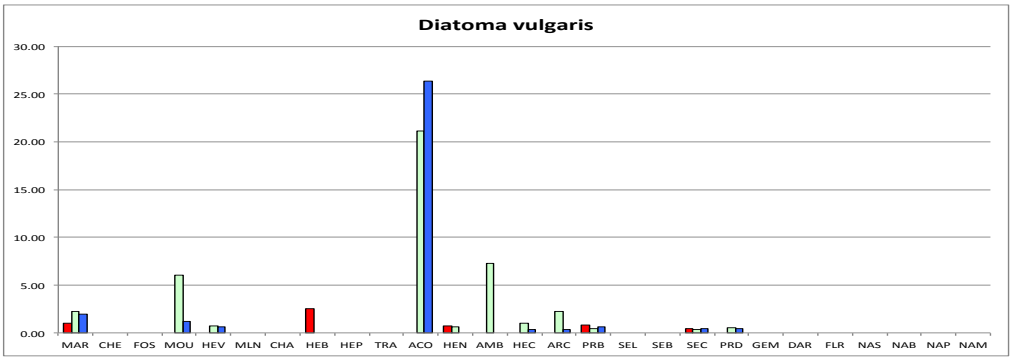

Melosira varians
90.00
80.00
70.00
60.00
50.00
40.00
30.00
20.00
10.00
0.00
MAR CHE FOS MOU HEV MLN CHA HEB HEP TRA ACO HEN AMB HEC ARC PRB SEL SEB SEC PRD GEM DAR FLR NAS NAB NAP NAM

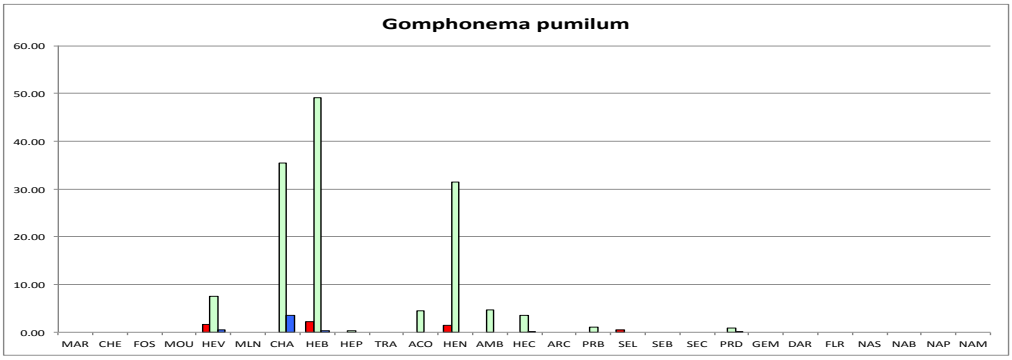

Nitzschia amphibia
12.00
10.00
8.00
6.00
4.00
2.00
0.00
MAR CHE FOS MOU HEV MLN CHA HEB HEP TRA ACO HEN AMB HEC ARC PRB SEL SEB SEC PRD GEM DAR FLR NAS NAB NAP NAM

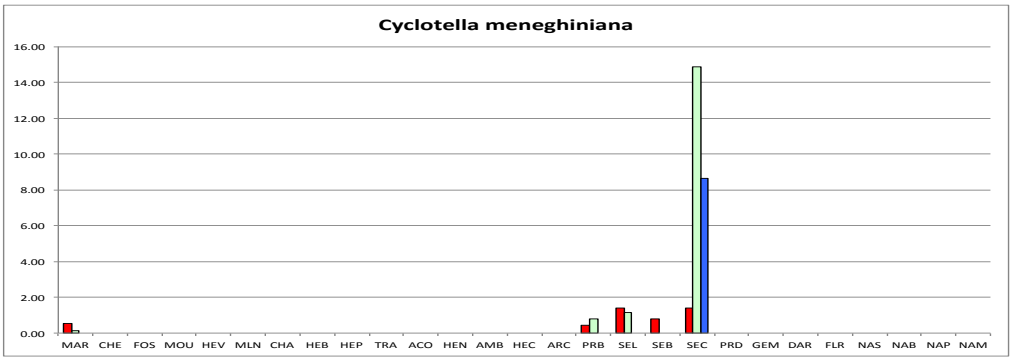

Nitzschia linearis
12.00
10.00
8.00
6.00
4.00
2.00
0.00
MAR CHE FOS MOU HEV MLN CHA HEB HEP TRA ACO HEN AMB HEC ARC PRB SEL SEB SEC PRD GEM DAR FLR NAS NAB NAP NAM

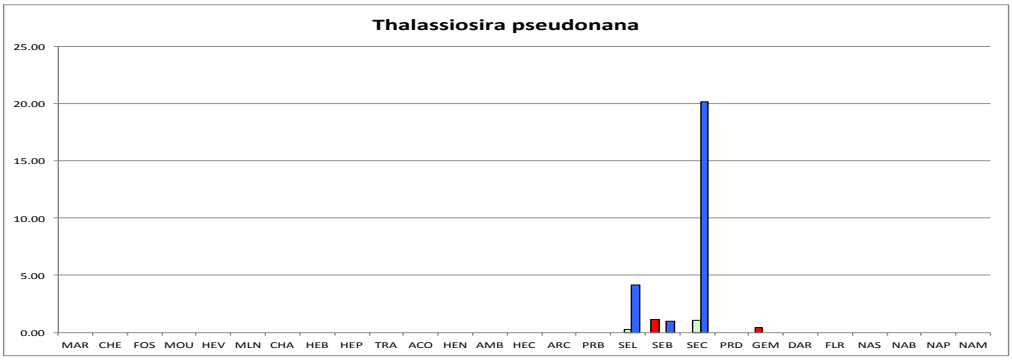

Nitzschia paleacea
30.00
25.00
20.00
15.00
10.00
5.00
0.00
MAR CHE FOS MOU HEV MLN CHA HEB HEP TRA ACO HEN AMB HEC ARC PRB SEL SEB SEC PRD GEM DAR FLR NAS NAB NAP NAM

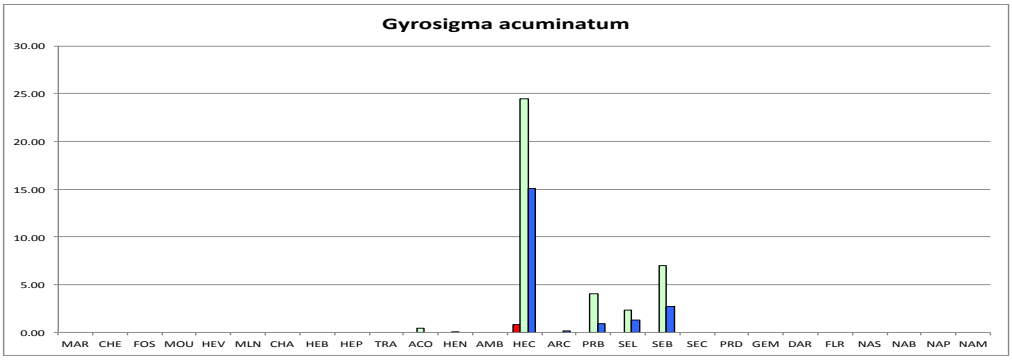

Encyonema spp
3.00
2.50
2.00
1.50
1.00
0.50
0.00
MAR CHE FOS MOU HEV MLN CHA HEB HEP TRA ACO HEN AMB HEC ARC PRB SEL SEB SEC PRD GEM DAR FLR NAS NAB NAP NAM

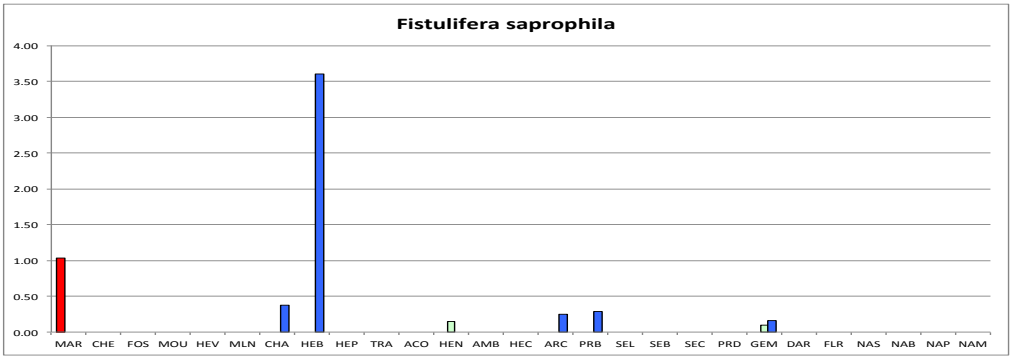

Diatoma vulgaris
30.00
25.00
20.00
15.00
10.00
5.00
0.00
MAR CHE FOS MOU HEV MLN CHA HEB HEP TRA ACO HEN AMB HEC ARC PRB SEL SEB SEC PRD GEM DAR FLR NAS NAB NAP NAM

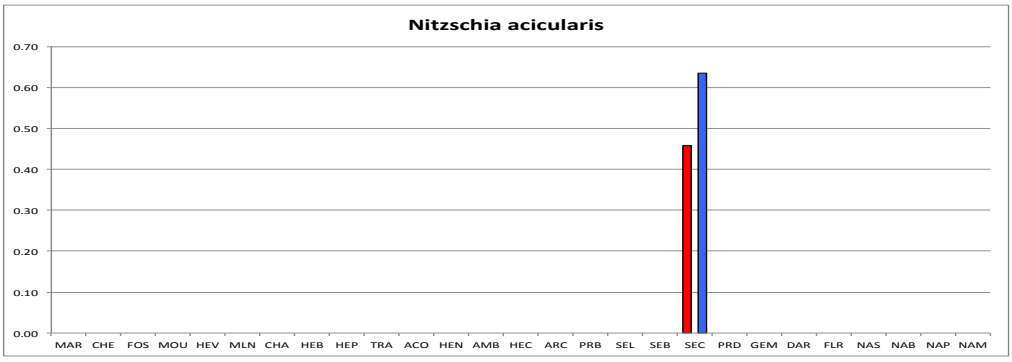

Gomphonema pumilum
60.00
50.00
40.00
30.00
20.00
10.00
0.00
MAR CHE FOS MOU HEV MLN CHA HEB HEP TRA ACO HEN AMB HEC ARC PRB SEL SEB SEC PRD GEM DAR FLR NAS NAB NAP NAM


Cyclotella meneghiniana
16.00
14.00
12.00
10.00
8.00
6.00
4.00
2.00
0.00
MAR CHE FOS MOU HEV MLN CHA HEB HEP TRA ACO HEN AMB HEC ARC PRB SEL SEB SEC PRD GEM DAR FLR NAS NAB NAP NAM


Thalassiosira pseudonana
25.00
20.00
15.00
10.00
5.00
0.00
MAR CHE FOS MOU HEV MLN CHA HEB HEP TRA ACO HEN AMB HEC ARC PRB SEL SEB SEC PRD GEM DAR FLR NAS NAB NAP NAM


Gyrosigma acuminatum
30.00
25.00
20.00
15.00
10.00
5.00
0.00
MAR CHE FOS MOU HEV MLN CHA HEB HEP TRA ACO HEN AMB HEC ARC PRB SEL SEB SEC PRD GEM DAR FLR NAS NAB NAP NAM


Fistulifera saprophila
4.00
3.50
3.00
2.50
2.00
1.50
1.00
0.50
0.00
MAR CHE FOS MOU HEV MLN CHA HEB HEP TRA ACO HEN AMB HEC ARC PRB SEL SEB SEC PRD GEM DAR FLR NAS NAB NAP NAM


Nitzschia acicularis
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
MAR CHE FOS MOU HEV MLN CHA HEB HEP TRA ACO HEN AMB HEC ARC PRB SEL SEB SEC PRD GEM DAR FLR NAS NAB NAP NAM

TableS 7.1 Site locations, geographic references and sampling dates performed along Geneva basin (Switzerland) in collaboration with SECOE-DETA and used for the study.

Number	Site	Location	Latitude	Longitude	Date
1	MAR	Marnot embouchure	2'508'599.45	1'127'400.27	10.09.13
2	CHE	Cherre amont chemin Armand-Dufaux	2'505'159.56	1'124'378.54	10.09.13
3	FOS	Fossaz amont chemin du Milieu	2'504'142.89	1'123'293.06	10.09.13
4	MOU	Moulin aval route d'Hermance	2'507'794.10	1'127'699.52	10.09.13
5	HEV	Hermance les Verrières	2'511'419.00	1'124'653.00	10.09.13
6	MLN	Moulanais amont chemin de la Montagne	2'504'066.57	1'117'522.85	10.09.13
7	CHA	Chamburaz embouchure	2'508'379.43	1'128'490.27	10.09.13
8	HEB	Hermance embouchure	2'507'959.43	1'128'740.28	23.09.13
9	HEP	Hermance Pont de Bouringe	2'508'224.44	1'128'395.28	23.09.13
10	TRA	Traînant Traînant	2'502'434.43	1'118'530.18	23.09.13
11	ACO	Aisy Côte d'or	2'506'449.85	1'124'702.40	23.09.13
12	HEN	Hermance Pont Neuf	2'507'789.47	1'125'310.25	23.09.13
13	AMB	Aisy embouchure	2'505'859.40	1'125'125.24	23.09.13
14	HEC	Hermance Pont de Crévy	2'507'694.48	1'126'490.26	23.09.13
15	ARC	Aisy route de Covéry	2'507'964.46	1'123'435.24	24.09.13
16	PRB	Paradis embouchure	2'507'214.42	1'120'065.17	24.09.13
17	SEL	Seymaz pont Ladame	2'505'401.54	1'118'537.50	24.09.13
18	SEB	Seymaz embouchure	2'502'969.43	1'115'070.26	24.09.13
19	SEC	Seymaz pont de Choulex/Montagnys	2'506'909.42	1'119'900.17	24.09.13
20	PRD	Paradis Les Doillets	2'510'164.53	1'120'155.14	24.09.13
21	GEM	Grebattes embouchure	2'496'099.34	1'117'410.18	13.03.14
22	DAR	Maison carrée	2'493'299.29	1'117'420.15	13.03.14
23	FLR	Mont fleuri	2'494'079.16	1'118'458.75	13.03.14
24	NAS	Nant d'avril Satigny	2'492'059.31	1'118'700.15	11.03.14
25	NAB	Nant d'avril Bourdigny	2'492'589.30	1'119'305.15	11.03.14
26	NAP	Nant d'avril Peney	2'492'114.29	1'118'010.15	11.03.14
27	NAM	Nant de la maille	2'493'984.13	1'122'427.65	11.03.14

TableS 7.2 Showing the filtering process on libraries DIATOM 2013 and DIATOM 2014

Statistics parameter	DIATOM 2013	DIATOM 2014
Total number of reads	1176424	1055387
Reject ambiguous forward	0	0
Reject ambiguous reverse	0	0
Low mean quality forward	52295	41746
Low mean quality reverse	117546	255053
Low mean quality contig	0	0
Low base quality contig	61508	17095
Not enough matching contig	2205	152394
No primers forward	55701	52065
Error in primers forward	4297	3075
No primers reverse	45934	35218
Error in primers reverse	4288	3659
Mismatch found in primers	67105	153677
Insufficient sequence length (dimers)	0	23222
Total number of good reads	765545	318183

TableS 7.3 Relative abundance and DI-CH values of morphological data per site location.

Species	D	G	MAR	CHE	FOS	MOU	HEV	MLN	CHA	HEB	HEP	TRA	ACO	HEN	AMB	HEC	ARC	PRB	SEL	SEB	SEC	PRD	GEM	DAR	FLR	NAS	NAB	NAP	NAM
Achnanthidium inconspicuum (Oestrup) Lange-Bertalot	1.0	4.0	0.0	0.0	0.0	3.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.3	7.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Achnanthidium minutissimum (Kützing) Czarnecki var. minutissimum	3.0	0.5	7.7	9.2	1.6	3.9	15.5	36.4	9.4	12.1	0.7	7.7	14.0	7.5	0.0	2.8	9.5	8.2	3.1	0.0	3.9	24.3	6.5	0.0	0.9	0.8	0.9	19.8	22.3
Achnanthidium straubianum (Lange-Bertalot) Lange-Bertalot	2.5	1.0	0.0	0.0	0.0	2.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Adlafia minuscula var. minuscula	4.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.4	0.0	2.3	0.0
Amphora aequalis Krammer			0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Amphora copulata (Kützing) Schoemann & Archibald			0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.3	2.6	0.0	0.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Amphora inariensis Krammer	3.5	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.3	0.9	0.0	0.0	0.4	0.0	0.0	0.0	0.0	0.0	0.4	2.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Amphora indistincta Levkov			0.5	0.0	0.0	0.0	0.0	0.0	0.9	0.0	0.0	0.0	3.0	0.8	0.0	0.0	2.0	3.4	3.1	3.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Amphora lange-bertalotii Levkov & Metzeltin var. lange-bertalotii			0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.4	0.0	0.0	0.0	0.0	0.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Amphora montana Krasske			0.0	0.0	0.0	1.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Amphora pediculus (Kütz.) Grün.	5.0	0.5	16.7	46.9	22.9	46.8	12.7	21.2	61.3	6.1	27.7	8.3	40.3	15.5	1.0	61.4	64.1	19.6	44.2	37.5	19.0	13.4	27.6	21.0	39.8	0.8	0.5	1.7	19.4
Caloneis bacillum (Grunow) Cleve	3.0	2.0	0.0	2.6	0.8	0.0	0.0	2.6	0.0	0.0	0.0	0.9	1.1	0.8	0.0	0.0	0.0	0.8	0.0	2.2	0.0	0.8	0.5	3.1	10.9	2.4	1.6	1.9	0.0
Cocconeis pediculus Ehrenberg	5.5	2.0	0.0	0.0	0.0	0.0	0.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.8	1.4	0.8	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Cocconeis placentula var. euglypta (Ehrenberg) van Heurck	5.0	1.0	40.1	0.0	0.0	0.2	11.2	0.0	7.2	2.5	44.0	0.4	3.2	1.3	0.0	0.8	0.9	6.8	4.0	0.4	4.8	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Cocconeis placentula var. pseudolineata Geitler	5.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	10.2	0.0	0.0	0.0	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Craticula minusculoides (Hustedt) L.B. 2001			0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Cyclotella meneghiniana Kützing	6.0	1.0	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.4	1.4	0.8	1.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Denticula tenuis Kützing	1.0	4.0	0.5	0.0	0.0	0.0	0.0	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.8	0.0
Diadesmis contenta (Grunow ex van Heurck) Mann 1990	4.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.7	0.0	0.0	0.0	0.0	0.0
Diatoma vulgaris Bory	4.0	2.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	2.5	0.0	0.0	0.0	0.8	0.0	0.0	0.0	0.8	0.0	0.0	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Diploneis elliptica (Kützing) Cleve	3.5	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Diploneis minuta Petersen			0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.8	0.0	0.4	0.0	3.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Diploneis oblongella (Naegeli) Cleve-Euler	3.5	4.0	0.0	0.0	0.0	0.0	0.0	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Encyonema minutum (Hilse) D.G.Mann (Hilse) D.G. Mann	2.5	2.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.2	0.0	1.7	0.0
Encyonema ventricosum (Agardh) Grunow	2.5	2.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.7	0.0	0.0	0.0	0.4	0.0	0.0	0.0	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.4	0.9	0.0	0.0
Eolimna minima (Grunow) L.B.	8.0	4.0	4.4	6.1	2.5	5.6	1.6	1.5	2.6	1.2	0.9	18.6	3.8	4.2	0.0	4.7	8.9	6.5	5.2	1.2	21.1	3.6	2.8	0.8	9.6	12.9	48.1	20.8	0.4
Eolimnia subminuscula Manguin	7.0	4.0	1.0	0.0	0.0	0.0	5.6	0.0	0.4	6.6	0.0	0.0	0.0	1.1	0.0	0.0	0.0	5.7	0.0	0.0	0.0	0.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Fistulifera saprophila (Lange-Bertalot & Bonik) Lange-Bertalot	7.0	2.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Fragilaria capucina var. rumpens (Kützing) L.B.	2.0	2.0	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Fragilaria ulna (Nitzsch) Lange-B.	4.0	1.0	0.0	0.0	0.0	0.0	2.4	0.0	0.0	0.3	0.0	0.0	0.0	0.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Frustulia vulgaris (Thwaites) De Toni.	4.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Gomphonema acuminatum Ehrenberg var. acuminatum			0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Gomphonema angustatum (Kützing) Rabenhorst	3.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Gomphonema micropumilum Reichardt	2.0	4.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Gomphonema micropus Kützing	3.0	1.0	0.0	0.0	0.0	1.0	1.6	0.0	0.0	0.3	0.0	0.0	0.0	0.0	0.0	0.4	0.0	0.0	0.0	0.8	0.0	0.4	1.6	4.6	6.0	4.6	0.5	0.4	22.7
Gomphonema minusculum Krasske	2.0	4.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Gomphonema minutum (Agardh) Agardh	2.5	2.0	0.0	0.0	0.0	0.0	0.8	0.0	0.9	1.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Gomphonema olivaceum (Horn.) Bréb.	3.0	0.5	2.1	1.3	0.0	2.9	1.6	0.4	3.1	4.5	0.0	0.0	3.2	1.1	0.0	0.0	0.0	0.0	0.0	0.8	0.0	0.0	0.9	0.0	0.0	0.0	0.5	0.4	4.5
Gomphonema parvulum (Kütz.) Kütz	8.0	4.0	0.5	0.0	0.0	0.0	0.0	0.0	0.0	1.7	0.0	0.9	2.4	0.8	0.0	0.0	0.0	0.0	0.0	0.0	0.5	1.3	2.8	0.8	0.0	23.8	7.7	10.0	0.0
Gomphonema parvulum var exilissimum Grunow	3.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Gomphonema pumilum (Grunow) Lange-bertalot & Reichardt	2.0	4.0	0.0	0.0	0.0	0.0	1.6	0.0	0.0	2.2	0.0	0.0	0.0	0.4	0.0	0.0	0.0	0.0	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Species	D	G	MAR	CHE	FOS	MOU	HEV	MLN	CHA	HEB	HEP	TRA	ACO	HEN	AMB	HEC	ARC	PRB	SEL	SEB	SEC	PRD	GEM	DAR	FLR	NAS	NAB	NAP	NAM
Gomphonema pumilum var. rigidum Reichardt et Lange-Bertalot	2.0	4.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Gomphonema tergestinum Fricke	3.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Gyrosigma acuminatum (Kütz.) Rabenhorst	4.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Gyrosigma nodiferum (Grunow) Reimer	4.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.5	1.2	0.0	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Gyrosigma obtusatum (Sullivant & Wormley) C.S. Boyer			0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Gyrosigma scalproïdes (Rabenhorst) Cleve			0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Mayamea atomus var. permitis (Hustedt) L.B.	6.0	1.0	0.0	0.0	0.0	0.5	0.0	0.0	0.4	2.4	0.0	0.4	0.0	0.8	1.9	0.0	0.0	0.4	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.6	18.1	0.0
Melosira varians J.G.Agardh	4.5	2.0	0.5	0.0	0.0	0.0	4.8	0.0	0.0	2.0	0.0	0.0	0.0	9.8	0.0	0.4	0.9	1.7	3.8	0.0	2.3	0.0	0.0	0.4	0.0	0.0	0.0	0.0	0.0
Meridion circulare (Greville) Agardh	3.5	2.0	0.0	0.0	0.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.5	2.7	0.4	0.0	0.0	2.2
Navicula antonii Lange-B.	5.0	1.0	0.5	0.0	0.0	0.0	0.8	0.0	0.0	1.3	0.0	0.0	0.0	3.8	0.0	1.7	2.4	4.4	2.6	1.8	0.9	3.6	0.0	1.9	0.0	0.0	0.0	0.0	0.4
Navicula capitatoradiata Germain	4.0	1.0	0.0	0.0	0.0	0.0	1.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Navicula contenta Grunow	4.0	1.0	0.0	0.0	53.3	1.0	0.0	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Navicula cryptocephala (Artengruppe) Kützing	4.0	1.0	0.0	0.0	0.0	0.0	0.8	0.0	0.0	0.0	0.0	0.0	0.0	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Navicula cryptotenella Lange-B.	4.0	0.5	1.0	0.9	0.4	1.5	3.2	0.4	2.0	3.9	0.4	0.4	3.8	2.7	0.0	2.5	0.4	9.3	0.0	7.1	0.9	0.4	0.0	1.5	0.4	0.0	0.0	0.4	0.4
Navicula cryptotenelloïdes Lange-Bertalot	4.0	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Navicula gregaria Donkin	5.5	1.0	1.5	0.0	0.0	1.5	2.4	0.0	0.4	1.2	0.0	0.0	3.2	6.1	0.0	1.3	0.0	2.3	0.0	2.5	0.0	1.9	0.9	23.9	0.0	3.8	3.2	0.0	0.0
Navicula lanceolata C.Agardh Ehrenberg	4.5	1.0	0.5	0.0	0.0	3.7	4.0	0.0	0.0	0.7	0.0	0.0	0.0	3.6	0.0	0.8	0.0	0.0	0.5	0.0	0.0	0.0	0.0	2.9	0.0	0.0	0.9	0.0	0.0
Navicula lenzii Hustedt	4.5	1.0	0.0	0.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.6	0.0	0.0	0.4	0.0	0.4	0.0	2.2	0.9	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Navicula oligotraphenta L.B. & Hofmann			0.0	0.0	0.0	0.5	0.0	0.0	0.0	0.0	0.0	0.0	1.1	0.0	0.0	0.0	0.0	0.8	0.0	0.0	0.0	1.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Navicula reichardtiana L.B. 1989	4.0	1.0	3.1	0.0	0.0	0.0	3.2	0.0	0.0	3.2	0.9	0.0	0.0	2.7	0.0	1.7	0.0	2.5	0.0	0.0	0.0	2.5	2.1	0.8	0.0	0.8	0.0	0.0	3.3
Navicula simulata Manguin			0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Navicula sublucidula Hustedt	4.0	2.0	0.0	2.0	0.8	0.5	0.0	0.0	0.0	0.0	0.0	0.0	1.1	0.0	0.0	0.0	0.0	0.0	0.0	2.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Navicula tripunctata (Müller) Bory	4.0	1.0	0.0	2.0	0.8	0.0	7.2	0.0	0.0	2.7	0.0	0.0	2.2	9.0	0.0	3.2	0.0	3.2	3.8	8.2	3.9	5.5	0.0	4.8	0.0	0.0	0.0	0.0	0.0
Navicula veneta Kützing	8.0	4.0	0.0	0.0	0.0	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.5	1.3	0.0	0.4	0.4	0.0	0.5	0.4	0.9	0.4	0.0	0.0	0.0	1.4	3.4	0.4	0.0
Navicula vilaplanii (L.B. & Sabater) L.B. & Sabater nov. Stat.			0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Nitzschia acicularis (Kütz.) W. Smith	7.0	4.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Nitzschia amphibia Grunow	7.0	8.0	0.0	0.0	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.5	0.0	0.0	0.0	0.0	2.3	1.4	0.0	4.4	5.7	0.0	0.0	10.9	0.0	0.5	0.0	0.0
Nitzschia angustatula Lange-Bertalot	4.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Nitzschia capitellata Hustedt	7.5	4.0	0.0	0.0	0.0	0.0	0.8	0.0	0.0	0.0	0.0	0.0	1.1	0.0	0.0	0.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Nitzschia communis Rabenhorst	8.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.8	1.4	2.5	0.0
Nitzschia constricta (Kützing) Ralfs	5.0	1.0	0.0	0.0	2.0	0.0	0.0	0.0	0.0	0.3	0.0	0.0	0.0	0.0	0.0	0.4	0.0	0.0	0.0	0.0	0.0	1.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Nitzschia debilis Arnott			0.0	0.0	0.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Nitzschia dissipata (Kütz.) Grün.	3.5	1.0	1.0	6.1	0.4	4.6	3.2	1.3	3.3	24.1	0.0	0.0	1.6	11.9	0.0	0.4	0.0	5.5	0.5	6.5	0.5	8.8	26.2	12.6	0.0	0.4	0.0	0.0	15.5
Nitzschia fonticola Grunow	3.5	1.0	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Nitzschia inconspicua Grunow	5.5	1.0	0.0	0.0	0.0	0.0	0.8	0.0	0.4	1.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Nitzschia intermedia Hantzsch	4.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Nitzschia lacuum Lange-Bertalot	4.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.8	0.0	1.3	0.0	0.0	0.0	1.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Nitzschia linearis (J.G.Agardh) W.Smith	4.5	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.8	1.9	1.7	0.0	4.8	10.0	0.0	0.0
Nitzschia palea (Kütz.) W.Smith	8.0	1.0	0.0	0.0	0.0	1.5	0.0	0.0	0.0	3.2	0.0	0.4	0.5	0.4	0.5	0.0	0.0	0.0	0.5	0.0	0.9	0.0	0.9	0.0	0.9	19.4	5.2	4.8	0.0
Nitzschia paleacea Grunow	7.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.8	0.0	8.1	0.0
Nitzschia perminuta (Grunow) M.Peragallo	3.5	1.0	0.0	0.0	0.0	0.0	0.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Nitzschia pusilla Grunow	5.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.8	2.3	0.4	0.0	1.4	0.5	1.9	0.0

Species	D	G	MAR	CHE	FOS	MOU	HEV	MLN	CHA	HEB	HEP	TRA	ACO	HEN	AMB	HEC	ARC	PRB	SEL	SEB	SEC	PRD	GEM	DAR	FLR	NAS	NAB	NAP	NAM
Nitzschia recta Hantzsch	3.5	2.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.7	0.0	0.0	0.5	0.0	0.0	0.8	0.0	0.0	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Nitzschia sociabilis Hustedt	3.5	2.0	0.0	0.0	0.0	0.0	0.0	0.0	0.9	0.0	0.0	0.0	0.5	1.9	0.0	0.4	0.0	0.0	0.0	4.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Nitzschia sublinearis Hustedt	3.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Nitzschia supralitorea (Lange-Bertalot)	6.5	1.0	0.0	0.0	0.0	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Planothidium frequentissimum (Lange-Bertalot) Lange-Bertalot	6.0	1.0	0.0	4.6	3.5	1.5	0.0	13.7	0.4	1.9	1.3	37.3	3.2	1.3	0.0	0.0	0.0	1.7	2.8	0.0	6.4	6.5	8.9	1.7	11.3	0.6	0.0	1.2	3.5
Planothidium lanceolatum (Brebisson) Lange-Bertalot	4.0	1.0	2.6	10.7	0.4	1.0	5.6	15.4	0.9	0.3	0.4	9.0	3.8	0.8	0.0	0.0	0.9	1.3	0.0	0.0	1.4	0.8	7.9	0.8	1.3	2.6	0.5	2.5	1.6
Platessa conspicua (A.Meyer) Lange-Bertalot	4.0	1.0	0.5	0.0	1.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.5	7.5	0.0	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Psammothidium grischunum (Wutrich) Bukhtiyarova & Round	1.0	2.0	0.0	0.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.8
Psammothidium lauenburgianum (Hustedt) Bukhtiyarova & Round	4.5	1.0	0.0	3.3	4.3	1.5	0.0	1.7	0.4	0.0	0.4	0.0	1.1	0.0	0.0	0.0	0.0	0.4	0.5	1.2	4.6	0.0	0.0	0.4	0.9	0.0	0.0	0.0	0.0
Reimeria sinuata (Greg.) Kociolek et Stoermer	3.5	1.0	6.2	2.2	0.0	0.5	1.6	3.2	4.6	1.0	6.5	0.0	1.1	0.4	0.0	0.8	1.6	1.9	0.0	1.8	2.8	0.0	0.0	0.0	0.0	0.4	0.0	0.0	0.0
Reimeria uniseriata Sala Guerrero & Ferrario			2.1	0.4	0.0	0.0	0.8	0.0	0.0	1.2	4.4	0.0	0.0	1.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Roicosphenia abbreviata (Agardt) Lange-B.	4.5	1.0	1.0	0.0	0.0	3.4	0.8	0.0	0.4	2.2	0.0	0.0	0.5	4.0	0.0	3.2	0.4	2.5	8.5	1.6	9.6	1.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Sellaphora seminulum (Grunow) D.G.Mann	8.0	4.0	1.8	0.0	0.8	3.9	0.0	0.4	0.0	0.0	0.0	14.1	0.5	0.0	96.2	0.0	0.0	0.0	0.0	0.0	0.0	1.7	4.7	1.9	3.6	0.0	2.7	0.0	0.0
Simonsenia delognei (Grunow) L.B.	5.0	1.0	0.0	0.0	1.8	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.8	0.0	0.0	0.0	2.4	0.0	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Surirella angusta Kützing	4.5	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	7.9	4.1	0.4	0.4
Surirella brebissonii kützingii Krammer et L.Bertalot	4.5	2.0	0.5	0.0	0.0	2.0	2.4	0.0	0.0	1.3	0.0	0.0	0.5	0.0	0.0	0.0	0.0	0.8	0.0	0.0	0.9	1.3	0.0	10.7	0.9	3.8	5.5	0.0	2.7
Surirella terricola L.B. & Alles			0.0	0.0	0.0	0.0	0.0	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Thalassiosira pseudonana Hasle & Heimdal	4.5	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
DI-CH value			5.45	5.22	4.82	5.27	4.92	4.47	5.04	4.89	4.89	7.17	5.74	5.04	7.98	5.22	4.84	5.90	5.61	4.54	6.67	5.92	5.40	4.83	6.01	6.87	6.75	6.34	3.64

TableS 7.4 List and counting of species found during the morphological analysis of the two campaigns and their presence in the database (DN) and in the molecular assignation (NGS).

Morphospecies	count	DB	NGS
Amphora pediculus (Kütz.) Grün.	3144	X	X
Achnanthidium minutissimum (Kützing) Czarnecki var. minutissimum	1040	X	X
Eolimna minima (Grunow) L.B.	904	X	X
Nitzschia dissipata (Kütz.) Grün.	653	X	X
Sellaphora seminulum (Grunow) D.G.Mann	564	X	X
Cocconeis placentula var. euglypta (Ehrenberg) van Heurck	538	X	X
Planothidium frequentissimum (Lange-Bertalot) Lange-Bertalot	525		
Planothidium lanceolatum (Brebisson) Lange-Bertalot	318		
Navicula contenta Grunow	278		
Navicula gregaria Donkin	261	X	
Navicula tripunctata (Müller) Bory	256	X	
Gomphonema parvulum (Kütz.) Kütz	253	X	X
Gomphonema micropus Kützing	217	X	X
Navicula cryptotenella Lange-B.	203	X	
Nitzschia palea (Kütz.) W.Smith	189	X	X
Roicosphenia abbreviata (Agardt) Lange-B.	181		
Reimeria sinuata (Greg.) Kociolek et Stoermer	161	X	
Caloneis bacillum (Grunow) Cleve	153		
Surirella brebissonii kützingii Krammer et L.Bertalot	152	X	X
Mayamea atomus var. permitis (Hustedt) L.B.	130	X	X
Gomphonema olivaceum (Horn.) Bréb.	126		
Navicula antonii Lange-B.	124		
Navicula reichardtiana L.B. 1989	123		
Melosira varians J.G.Agardh	119	X	X
Nitzschia amphibia Grunow	118	X	X
Nitzschia linearis (J.G.Agardh) W.Smith	108	X	X

Eolimnia subminuscula Manguin	96	X	
Psammothidium lauenburgianum (Hustedt) Bukhtiyarova & Round	95		
Amphora indistincta Levkov	75		
Navicula lanceolata C.Agardh Ehrenberg	74	X	
Surirella angusta Kützing	61	X	X
Achnanthidium inconspicuum (Oestrup) Lange-Bertalot	54	X	X
Platessa conspicua (A.Meyer) Lange-Bertalot	52		
Nitzschia paleacea Grunow	52	X	X
Cocconeis placentula var. pseudolineata Geitler	49	X	X
Navicula veneta Kützing	49	X	
Reimeria uniseriata Sala Guerrero & Ferrario	45		
Nitzschia sociabilis Hustedt	40		
Simonsenia delognei (Grunow) L.B.	38		
Meridion circulare (Greville) Agardh	36		
Nitzschia pusilla Grunow	36	X	
Navicula lenzii Hustedt	31		
Navicula sublucidula Hustedt	31		
Diatoma vulgaris Bory	29	X	X
Diploneis minuta Petersen	25		
Navicula cryptotenelloïdes Lange-Bertalot	24		
Nitzschia communis Rabenhorst	22	X	
Nitzschia constricta (Kützing) Ralfs	22		
Amphora copulata (Kützing) Schoemann & Archibald	21		
Amphora inariensis Krammer	21		
Gomphonema pumilum (Grunow) Lange-bertalot & Reichardt	21	X	X
Cyclotella meneghiniana Kützing	20	X	X
Encyonema minutum (Hilse) D.G.Mann (Hilse) D.G. Mann	19	X	X

Adlafia minuscula var. minuscula	18		
Cocconeis pediculus Ehrenberg	18	X	X
Nitzschia inconspicua Grunow	17	X	
Navicula oligotraphenta L.B. & Hofmann	16		
Nitzschia lacuum Lange-Bertalot	16		
Gomphonema minutum (Agardh) Agardh	15		
Encyonema ventricosum (Agardh) Grunow	14		X
Gomphonema acuminatum Ehrenberg var. acuminatum	14	X	
Achnanthidium straubianum (Lange-Bertalot) Lange-Bertalot	12	X	
Fragilaria ulna (Nitzsch) Lange-B.	12		
Nitzschia recta Hantzsch	12		
Gyrosigma nodiferum (Grunow) Reimer	10		
Nitzschia capitellata Hustedt	10		
Psammothidium grischunum (Wutrich) Bukhtiyarova & Round	9		
Denticula tenuis Kützing	8		
Diadesmis contenta (Grunow ex van Heurck) Mann 1990	8		
Gomphonema tergestinum Fricke	8		
Thalassiosira pseudonana Hasle & Heimdal	8	X	X
Amphora lange-bertalotii Levkov & Metzeltin var. lange-bertalotii	6		
Amphora montana Krasske	6		
Fragilaria capucina var. rumpens (Kützing) L.B.	6	X	
Gomphonema pumilum var. rigidum Reichardt et Lange-Bertalot	6	X	X
Gyrosigma scalproïdes (Rabenhorst) Cleve	6		
Navicula cryptocephala (Artengruppe) Kützing	6	X	
Diploneis oblongella (Naegeli) Cleve-Euler	4		
Fistulifera saprophila (Lange-Bertalot & Bonik) Lange-Bertalot	4	X	X
Gomphonema angustatum (Kützing) Rabenhorst	4	X	

Gomphonema parvulum var exilissimum Grunow	4	X	X
Gyrosigma acuminatum (Kütz.) Rabenhorst	4	X	X
Navicula capitatoradiata Germain	4	X	
Nitzschia debilis Arnott	4		
Surirella terricola L.B. & Alles	4		
Amphora aequalis Krammer	2		
Craticula minusculoides (Hustedt) L.B. 2001	2		
Diploneis elliptica (Kützing) Cleve	2		
Frustulia vulgaris (Thwaites) De Toni.	2	X	
Gomphonema micropumilum Reichardt	2		
Gomphonema minusculum Krasske	2		
Gyrosigma obtusatum (Sullivant & Wormley) C.S. Boyer	2		
Navicula simulata Manguin	2		
Navicula vilaplanii (L.B. & Sabater) L.B. & Sabater nov. Stat.	2		
Nitzschia acicularis (Kütz.) W. Smith	2	X	X
Nitzschia angustatula Lange-Bertalot	2		
Nitzschia fonticola Grunow	2	X	
Nitzschia intermedia Hantzsch	2		
Nitzschia perminuta (Grunow) M.Peragallo	2		
Nitzschia sublinearis Hustedt	2		
Nitzschia supralitorea (Lange-Bertalot)	2	X	

TableS 7.5 Relative abundance of morphologic, DNA and RNA data per sites.

	MAR			CHE			FOS			MOU			HEV		
Species	**Mor**	**DNA**	**RNA**	**Mor**	**DNA**	**RNA**	**Mor**	**DNA**	**RNA**	**Mor**	**DNA**	**RNA**	**Mor**	**DNA**	**RNA**
Achnanthidium minutissimum	7.7	13.37	16.98	9.21	14.01	2.80	1.57			7.80	43.05	59.57	15.54	0.74	2.06
Amphora pediculus	16.7	1.32	2.17	46.93	37.32	83.52	22.94	25.00		46.83	9.50	5.48	12.75	0.48	1.00
Cocconeis placentula/pediculus	40.1	60.90	26.41		0.99	0.17				0.24	10.88	0.88	11.95	2.00	0.70
Cyclotella meneghiniana	0.5	0.13													
Diatoma vulgaris	1.0	2.24	1.93								6.02	1.21		0.68	0.65
Encyonema														0.18	
Eolimna minima	4.4	0.33	0.39	6.14	1.14	0.63	2.55			5.61	1.95	0.46	1.59		
Fistulifera saprophila	1.0														
Gomphonema micropus										0.98			1.59		
Gomphonema parvulum	0.5	1.31			0.65						2.48			0.56	
Gomphonema pumilum													1.59	7.48	0.53
Gyrosigma acuminatum															
Mayamea atomus			2.27		0.57	0.55				0.49	0.46	2.38			0.33
Melosira varians	0.5	0.68	13.25								0.85		4.78	78.37	59.39
Nitzschia acicularis															
Nitzschia amphibia							0.39								
Nitzschia dissipata	1.0		0.25	6.14	7.16	1.58	0.39			4.63		0.77	3.19		0.97
Nitzschia linearis										0.98					
Nitzschia palea		0.28	1.45							1.46				0.11	1.09
Nitzschia paleacea															
Sellaphora seminulum	1.8						0.78			3.90		0.23			
Surirella angusta	0.5	0.15								1.95	1.28		2.39	0.26	
Thalassiosira pseudonana															

	MLN			CHA			HEB			HEP			TRA	
Species	**Mor**	**DNA**	**RNA**	**Mor**	**DNA**	**RNA**	**Mor**	**DNA**	**RNA**	**Mor**	**DNA**	**RNA**	**Mor**	**DNA**	**RNA**
Achnanthidium minutissimum	36.40	40.95	15.65	9.41	2.99	6.68	12.12	2.20	18.62	0.65	0.17	0.16	7.68	9.63	1.43
Amphora pediculus	21.20	3.18	10.64	61.27	3.95	42.26	6.06		0.19	27.67	1.67	4.71	8.32	1.39	1.56
Cocconeis placentula/pediculus		1.23	0.41	7.22	15.83	18.10	2.53	0.58		54.25	95.88	84.40	0.43		
Cyclotella meneghiniana															
Diatoma vulgaris							2.53								
Encyonema							0.67								
Eolimna minima	1.50	0.72		2.63			1.18			0.87			18.55	6.69	3.42
Fistulifera saprophila						0.37			3.60						
Gomphonema micropus							0.34								
Gomphonema parvulum							1.68	0.54					1.71	23.63	0.42
Gomphonema pumilum					35.35	3.51	2.19	49.17	0.30		0.34				
Gyrosigma acuminatum															
Mayamea atomus			0.27	0.44		0.57	2.36		3.33			0.13	0.43	1.80	2.80
Melosira varians		1.21					2.02	0.62							0.20
Nitzschia acicularis															
Nitzschia amphibia															
Nitzschia dissipata	1.28			3.28		5.54	24.07	0.62	57.15						
Nitzschia linearis													0.43		
Nitzschia palea							3.20	0.71	1.94				0.43		
Nitzschia paleacea							0.67		0.18						
Sellaphora seminulum	0.43												14.07	4.15	4.16
Surirella angusta		0.42			0.53		1.35	14.17	0.51					0.15	
Thalassiosira pseudonana															

	ACO			HEN			AMB			HEC			ARC	
Species	**Mor**	**DNA**	**RNA**	**Mor**	**DNA**	**RNA**	**Mor**	**DNA**	**RNA**	**Mor**	**DNA**	**RNA**	**Mor**	**DNA**	**RNA**
Achnanthidium minutissimum	13.98	3.10	4.31	7.46	3.53	4.70				4.03	0.76	0.78	16.63	19.87	12.22
Amphora pediculus	40.32	8.21	37.28	15.49	5.22		0.96			61.44	11.53	36.32	64.08	47.32	82.74
Cocconeis placentula/pediculus	3.23	10.70	3.45	1.34	1.09		0.48			0.85	3.54	2.78	0.89		
Cyclotella meneghiniana															
Diatoma vulgaris		21.16	26.40	0.76	0.34			7.31			0.99	0.32		2.28	0.33
Encyonema				0.38	0.59										
Eolimna minima	3.76			4.21	0.90			9.63		4.66	0.20	0.51	8.87	5.04	0.56
Fistulifera saprophila					0.15										0.24
Gomphonema micropus										0.42					
Gomphonema parvulum	2.42	0.59		0.76	3.75						0.21			10.83	0.13
Gomphonema pumilum		4.43		1.53	31.52			4.65			3.54	0.15			
Gyrosigma acuminatum		0.47			0.09					0.85	24.42	15.06			0.11
Mayamea atomus				0.76	0.33		1.91								0.19
Melosira varians		1.93		9.75	16.09	28.07				0.42	9.61	0.65	0.89		
Nitzschia acicularis															
Nitzschia amphibia	0.54														
Nitzschia dissipata	1.61	15.70	13.25	11.85	0.21					0.42		3.53			
Nitzschia linearis		0.30													
Nitzschia palea	0.54			0.38	0.11		0.48								
Nitzschia paleacea															
Sellaphora seminulum	0.54						96.17	35.22	100.00						
Surirella angusta	0.54	1.12			2.31						0.25			1.02	
Thalassiosira pseudonana															

	PRB			SEL			SEB			SEC			PRD	
Species	**Mor**	**DNA**	**RNA**	**Mor**	**DNA**	**RNA**	**Mor**	**DNA**	**RNA**	**Mor**	**DNA**	**RNA**	**Mor**	**DNA**	**RNA**
Achnanthidium minutissimum	8.23	23.00	31.78	3.06	0.72	0.90		0.50	0.50	3.90	1.49	2.11	24.32	8.89	15.45
Amphora pediculus	19.62	15.05	13.98	44.24	67.09	61.34	37.45	17.73	44.24	19.04	34.52	25.65	13.42	1.44	1.46
Cocconeis placentula/pediculus	7.59	21.27	10.78	5.41	6.17	2.61	1.18			5.28	9.03	5.66	0.42		
Cyclotella meneghiniana	0.42	0.80		1.41	1.17		0.78			1.38	14.86	8.64			
Diatoma vulgaris	0.84	0.41	0.61							0.46	0.31	0.40		0.51	0.42
Encyonema	0.42	0.25													
Eolimna minima	6.54	0.73	1.26	5.18	2.49	4.45	1.18			21.10	5.88	12.19	3.56	0.73	0.31
Fistulifera saprophila			0.29												
Gomphonema micropus							0.78						0.42		
Gomphonema parvulum		1.26			0.14					0.46			1.26	1.52	
Gomphonema pumilum		1.09		0.47										0.91	0.13
Gyrosigma acuminatum		4.01	0.94		2.34	1.29		7.01	2.73						
Mayamea atomus	0.42		2.09	0.47											0.31
Melosira varians	1.69	4.01	1.15	3.76	7.52	2.32		5.69	0.98	2.29	20.41	10.17			0.24
Nitzschia acicularis										0.46		0.63			
Nitzschia amphibia	2.32			1.41						4.36	1.06		5.66	0.08	
Nitzschia dissipata	5.49		1.38	0.47			6.47	6.07	20.28	0.46			8.81	0.16	25.25
Nitzschia linearis													3.77	0.39	0.83
Nitzschia palea			0.40	0.47						0.92	0.33	0.85			0.13
Nitzschia paleacea			0.36												
Sellaphora seminulum													1.68		
Surirella angusta	0.84	0.97								0.92	0.24		1.26	4.59	0.09
Thalassiosira pseudonana					0.30	4.16	1.18		1.01		1.11	20.13			

	GEM			DAR			FLR			NAS			NAB		
Species	**Mor**	**DNA**	**RNA**	**Mor**	**DNA**	**RNA**	**Mor**	**DNA**	**RNA**	**Mor**	**DNA**	**RNA**	**Mor**	**DNA**	**RNA**
Achnanthidium minutissimum	6.54	1.82	2.18				0.89		0.22	0.81	0.14	17.82	0.91		28.85
Amphora pediculus	27.57	10.00	21.19	20.96	0.28	0.11	39.78	27.89	43.93	0.81			0.46		
Cocconeis placentula/pediculus															
Cyclotella meneghiniana															
Diatoma vulgaris															
Encyonema										2.63	0.14		0.91		0.40
Eolimna minima	2.80			0.84			9.56	8.28	5.72	12.93	0.22		48.06	13.60	29.47
Fistulifera saprophila		0.10	0.16												
Gomphonema micropus	1.64			4.61	2.15	2.12	6.00	4.33	17.69	4.65	0.64	23.76	0.46	0.44	4.24
Gomphonema parvulum	2.80	37.11	6.42	0.84	0.27	0.36		0.26	4.72	23.84	95.72	6.74	7.74	68.29	4.95
Gomphonema pumilum															
Gyrosigma acuminatum															
Mayamea atomus												1.63	1.59		5.23
Melosira varians				0.42	0.15	0.33									
Nitzschia acicularis															
Nitzschia amphibia							10.89						0.46		
Nitzschia dissipata	26.17	9.33	46.87	12.58	0.10	0.97				0.40					0.18
Nitzschia linearis	1.87	1.07		1.68						4.85	0.45		10.02	9.01	3.19
Nitzschia palea	0.93						0.89	0.11	0.95	19.39	0.61	34.93	5.24		8.70
Nitzschia paleacea										1.82		3.19			
Sellaphora seminulum	4.67	0.27		1.89			3.56						2.73		0.48
Surirella angusta		2.73	0.77	10.69	46.19	2.49	0.89	2.09	1.78	11.72	1.48	1.75	9.57	6.77	0.50
Thalassiosira pseudonana	0.47														

	NAP			NAM		
Species	**Mor**	**DNA**	**RNA**	**Mor**	**DNA**	**RNA**
Achnanthidium minutissimum	19.75	3.70	27.14	22.31	3.54	22.16
Amphora pediculus	1.66		0.20	19.37	3.58	0.80
Cocconeis placentula/pediculus						
Cyclotella meneghiniana						
Diatoma vulgaris						
Encyonema	1.66	0.23	0.23			
Eolimna minima	20.79	4.55	9.19	0.39	4.81	
Fistulifera saprophila						
Gomphonema micropus	0.42	0.15	1.60	22.70	54.44	62.81
Gomphonema parvulum	9.98	86.04	10.02		26.89	0.41
Gomphonema pumilum						
Gyrosigma acuminatum						
Mayamea atomus	18.09	1.09	9.48			0.69
Melosira varians						
Nitzschia acicularis						
Nitzschia amphibia						
Nitzschia dissipata				15.46		0.22
Nitzschia linearis					2.47	
Nitzschia palea	4.78	0.08	6.87			
Nitzschia paleacea	8.11	0.64	27.40			
Sellaphora seminulum			0.35			
Surirella angusta	0.42	0.60		3.13	1.86	0.28
Thalassiosira pseudonana						

TableS 7.6 DI-CH values for morphologic, DNA and RNA data per sites.

DI-CH	MAR	CHE	FOS	MOU	HEV	MLN	CHA	HEB	HEP	TRA	ACO	HEN	AMB	HEC	ARC	PRB	SEL	SEB	SEC	PRD	GEM	DAR	FLR	NAS	NAB	NAP	NAM
Mor	5.45	5.22	4.82	5.27	4.92	4.47	5.04	4.89	4.92	7.17	5.74	5.04	7.98	5.22	4.84	5.90	5.61	4.54	6.67	5.92	5.40	4.83	6.01	6.87	6.75	6.34	3.64
DNA	5.09	4.93	5.00	5.23	4.16	3.84	2.37	2.27	4.96	7.79	3.69	3.20	7.44	3.97	6.73	4.69	5.33	4.34	5.70	5.16	7.44	4.51	5.34	7.97	7.77	7.94	6.25
RNA	4.74	5.06		3.73	4.48	3.89	4.13	3.97	5.00	7.65	4.35	4.44	8.00	4.62	4.94	4.66	5.74	4.27	6.16	3.63	5.03	4.62	5.51	6.26	6.29	6.78	3.15

CHAPTER 8

TAXONOMY-FREE MOLECULAR DIATOM INDEX FOR HIGH-THROUGHPUT eDNA BIOMONITORING

LAURE APOTHÉLOZ-PERRET-GENTIL, ARIELLE CORDONNIER, FRANÇOIS STRAUB, JENNIFER ISELI, PHILIPPE ESLING, JAN PAWLOWSKI

Published in Molecular Ecology Resources April 2017

8.1. Project description

The diatoms project in collaboration with Arielle Cordonier (Geneva) and François Straub (La Chaux-de-Fonds) continued after Joana Visco finished her master and I took the whole project in charge. Given the difficulties encountered with completing the reference database, we develop an alternative method assigning diatoms sequences to particular ecological conditions, without the taxonomic assignation. Philippe Esling, postdoctoral fellow in the lab at this time, help a lot in the development of the method in term of conception and data processing. I have conducted the lab work and data analysis, using small programs that I have developed to help me analysing all HTS dataset.

8.2. Abstract

Current biodiversity assessment and biomonitoring are largely based on the morphological identification of selected bioindicator taxa. Recently, several attempts have been made to use eDNA metabarcoding as an alternative tool. However, until now, most applied metabarcoding studies have been based on the taxonomic assignment of sequences that provides reference to morphospecies ecology. Usually, only a small portion of metabarcoding data can be used due to a limited reference database and a lack of phylogenetic resolution. Here, we investigate the possibility to overcome these limitations by using a taxonomy-free approach that allows the computing of a molecular index directly from eDNA data without any reference to morphotaxonomy. As a case study, we use the benthic diatoms index, commonly used for monitoring the biological quality of rivers and streams. We analysed 87 epilithic samples from Swiss rivers, the ecological status of which was established based on the microscopic identification of diatom species. We compared the diatom index derived from eDNA data obtained with or without taxonomic assignment. Our taxonomy-free approach yields promising results by providing a correct assessment for 77% of examined sites. The main advantage of this method is that almost 95% of OTUs could be used for index calculation, compared to 35% in the case of the taxonomic assignment approach. Its main limitations are under-sampling and the need to calibrate the index based on the microscopic assessment of diatoms communities. However, once calibrated, the taxonomy-free molecular index can be easily standardized and applied in routine biomonitoring, as a complementary tool allowing fast and cost-effective assessment of the biological quality of watercourses.

8.3. Introduction

Various biotic indices are widely used for the assessment of water quality. Traditionally, the indices are calculated based on the diversity of selected bioindicator taxa identified morphologically (Borja & Dauer 2008; Poikane *et al.* 2011). Recently, several attempts have been made to use eDNA data to infer the community structure of bioindicator species (Baird & Hajibabaei 2012; Chariton *et al.* 2015). Several factors have been identified that may potentially impede the correct

assignment of sequences to morphospecies and therefore the calculation of accurate indices. In particular, the incompleteness of the genetic database, the lack of resolution of phylogenetic markers and cryptic diversity (Yu *et al.* 2012; Carew *et al.* 2013; Eiler *et al.* 2013) have been highlighted as major issues. To overcome these limitations, we examine here whether it is possible to infer a molecular index directly from eDNA data without referring to the morphotaxonomy.

As a case study, we chose benthic diatoms, which are widely used as bioindicators of rivers and streams because of their high sensitivity to environmental changes and well-established taxon-specific ecological tolerances and preferences (Stevenson *et al.* 2010). In 2000, the European Union published a directive, the Water Framework Directive (Directive 2000/60/EC), that commits all member states to evaluate the status of their water bodies and to achieve a good status for them by a set deadline, recommending diatoms as one of the ideal bioindicators for river assessment. Different biotic indices are used across the different countries (Kelly *et al.* 2008) In Switzerland, two biological indices are used to comply with the concomitant ecological objectives specified by the Swiss decree on water protection (Swiss Federal Council 1998), the IB-CH using macrozoobenthos and the DI-CH, using diatoms. The Swiss Diatom Index (DI-CH) is based on chemical parameters indicating anthropogenic pollution and classifies the water quality into 5 different ecological classes on a scale from 1 to 8 (1-3.5: very good; 3.5-4.5: good; 4.5-5.5: average; 5.5-6.5: bad; 6.5-8: very bad). The calculation follows the weighted average equation of Zelinka & Marvan (1961) and is defined as

$$\text{DI-CH} = \frac{\sum_{i=1}^{n} D_i \, G_i H_i}{\sum_{i=1}^{n} G_i H_i}$$

This equation involves an autecological value D and a weighting factor G, which are specific to each species. It also uses an additional parameter H, which corresponds to the relative frequency of a particular taxon in the sample.

Like other diatom indices (Kelly *et al.* 2001; Coste *et al.* 2009), the DI-CH requires a morphologic determination to the species level. This requirement is a major weakness of the currently used system. Indeed, diatoms are a highly diverse group of protists and the identification of their tiny frustules requires special sample preparation, high quality microscopes and in-depth taxonomic expertise. Inter-calibration exercises among specialists are organised to validate the robustness of

the indices. These time-consuming limiting factors contrast with the need for the fast routine assessment of water quality required by Water Framework Directive and the Swiss Federal Office for the Environment.

The development of high-throughput sequencing (HTS) technologies applied to diversity surveys of microbial eukaryotes communities provided a possibility to overcome some of these limitations (Pawlowski *et al.* 2016b). Several attempts have been made to use HTS eDNA metabarcoding as a tool for identifying diatom species either in mock communities (Kermarrec *et al.* 2013, 2014) or in environmental samples (Kermarrec *et al.* 2014; Zimmermann *et al.* 2014, 2015; Visco *et al.* 2015 (Chapter 7)). Some authors attempted to infer diatom indices from metabarcoding data (Kermarrec *et al.* 2014; Visco *et al.* 2015 (Chapter 7); Keck *et al.* 2016).

However, the results of these studies were not entirely satisfactory due to uncertainties concerning the correct assignment of sequences to morphospecies and various biases involved in qualitative and quantitative analyses of molecular data.

Here, we propose a taxonomy-free approach to calculate the Swiss Diatom Index values directly from sequence data. To test this new approach, we analyse 87 epilithic samples from Swiss rivers, mostly located in the Geneva basin, using the hypervariable region V4 of 18S rDNA as the diatom DNA barcode and the Illumina Miseq platform for sequencing. As illustrated in Figure 8.1, we calculate the DI-CH values inferred from molecular data with two methods. First, by phylogenetic assignment of OTUs to morphospecies (DI-MOLTAXASSIGN - pathway 2), as previously described in Visco *et al.* (2015 - Chapter 7). Second, by assigning OTUs directly to ecological classes (DI-MOLTAXFREE - pathway 3). Finally, we compare those values with the ones derived from traditional microscopic studies (DI-CH - pathway 1).

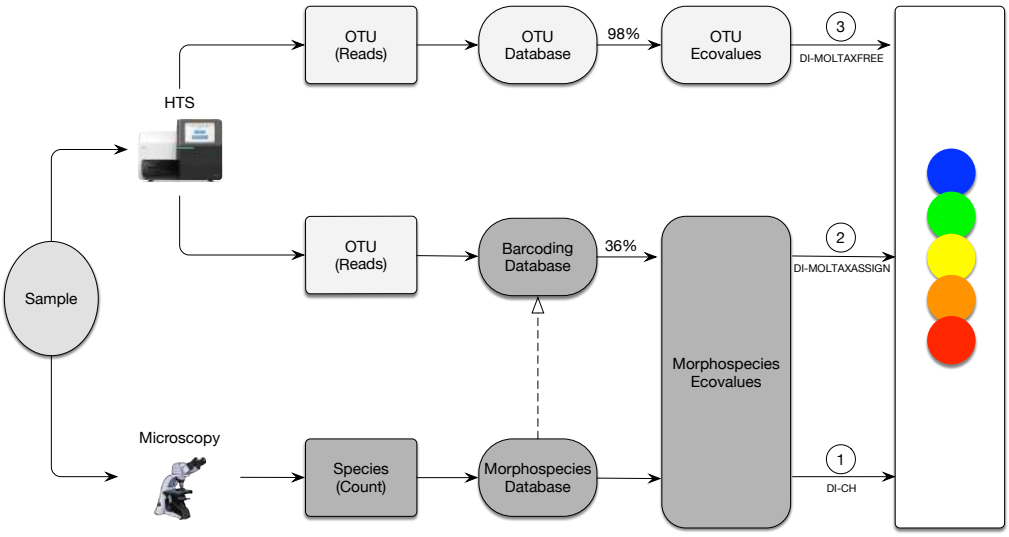


Figure 8.1 Workflow illustrating the different methods used in this paper.

8.4. Materials and methods

8.4.1. Sampling.

In total, 87 samples were collected during the 2013-2015 period in the Geneva and Neuchâtel cantons in Switzerland (TableS 8.1, FigureS 8.1). This number includes 27 samples already published in Visco *et al.* (2015 - Chapter 7). All the samples were collected as part of the monitoring program for water quality performed by the Service of Water Ecology (SECOE) of the Department of Environment, Transport and Agriculture of the Geneva canton and the Service of Energy and Environment of the Neuchâtel canton. The biofilm containing epilithic diatoms was collected following the directives established by the Swiss Federal Office for the Environment (Hürlimann & Niederhauser 2007). Each sample was divided into two subsamples for morphological and molecular analyses. Morphological samples were preserved with a final concentration of at least 4% of formaldehyde, while molecular samples were kept cold (ca. 0°C) during sampling. In the laboratory, about 1 ml of each sample suspension was centrifuged and pellets were stored at -80°C until further investigations.

8.4.2. Morphological analysis.

The preparation of diatoms slides for microscopic observation was performed as recommended by the Swiss Federal Office for the Environment (Hürlimann & Niederhauser 2007). About 500 valves per sample were counted and identified mainly with the bibliographic support of The Flora of Diatoms (Krammer & Lange-Bertalot 1986), Diatoms of Europe (Lange-Bertalot 2001) and Iconographia Diatomologica (Lange-Bertalot & Metzeltin 1996; Reichardt 1999), and Diatomeen im Süsswasser-Benthos von Mitteleuropa (Hofmann *et al.* 2011). In the case of the samples from Neuchâtel, after the 500 valves had been counted, the preparations were scanned for 20 minutes to find rare species. Finally the DI-CH values for each site were calculated following the equation described above.

8.4.3. Reference Database.

We chose the V4 region following the work of Zimmermann *et al.* (2011) and our previous study (Visco *et al.* 2015 - Chapter 7). Although alternative diatom barcodes, such as *rbc*L, seem to offer better taxonomic resolution, we favour the V4 region because its amplification from eDNA samples is easier and its size better fits the sequencing length of the Illumina Miseq platform. We built a reference database of the 18S V4 region of diatoms using online databases GenBank Release 212 and R-syst::diatom v5 (Rimet *et al.* 2016) and Sanger sequences from previous environmental studies in the Geneva basin . The region of interest was cut from downloaded sequences and aligned using the Seaview program (Gouy *et al.* 2010). The alignment was checked manually. Environmental sequences were screened using Uchime for chimeras (Edgar *et al.* 2011), which were then removed. The remaining sequences were analysed by Maximum Likelihood (ML) phylogenetic inference and those that did not branch in the clade corresponding to their morphological identification were discarded. After filtering, 1297 unique diatom sequences were kept, including 155 environmental sequences coming from the same geographical area as the study (TableS 8.2).

8.4.4. Molecular analysis.

DNA was extracted with the PowerBiofilm® DNA Isolation kit (MO BIO Laboratories Inc.) according to the manufacturer instructions. Three extraction replicates were performed for each sample. The hypervariable region V4 of the 18S rRNA gene of

diatoms was then enriched by PCR amplification using specific diatom primers modified after Zimmermann *et al.* (2011). Following previous studies, PCRs were performed as described in Visco *et al.* (2015 - Chapter 7), using unique combinations of forward and reverse primers tagged with individual tags composed of 8 nucleotides attached at each primers 5'-extremities (Esling *et al.* 2015). A total of 20 different forward and reverse tagged primers were designed to enable multiplexing of all PCR products in a unique sequencing library. The sequences of tags and primers are provided in TableS 8.3. Two PCR replicates were performed for each extraction and were then pooled for purification with High Pure PCR Cleanup Micro kit (Roche Diagnostics). In total, 6 PCR replicates were pooled for each sample. Purified PCR products were quantified with QuBit HS ds DNA kit (Invitrogen) and pooled in equimolar quantities. Two libraries were prepared (DIATOM03 for 2014 samples and DIATOM05 for 2015 samples, containing 24 and 36 samples respectively) using Illumina TruSeq® DNA PCR-Free Library Preparation Kit following the manufacturer's instructions. The libraries were then quantified with qPCR using KAPA Library Quantification Kit and sequenced on a MiSeq instrument using paired-end sequencing for 500 cycles with Nano kit v2.

8.4.5. HTS data analysis.

Quality filtering and assembly were performed according to the method described in Visco *et al*. (2015 – Chapter 7). The two runs from our previous study and the two from this study were combined and this complete dataset was de-replicated, i.e. the identical sequences were grouped together in order to obtain unique sequences, called Independent Sequence Units (ISUs). An abundance threshold of 10 was used for the minimum number of reads required for each ISU (Bokulich *et al.* 2013). We removed the ISUs that did not match any diatom sequences in the NCBI database with at least 99% coverage and 97% identity. ISUs were then grouped at 99% using complete-linkage clustering method. Finally, we removed chimeric sequences found with manual inspection of Uchime (Edgar *et al.* 2011) candidates.

8.4.6. Phylogenetic analyses.

Taxonomic assignment of the operational taxonomic units (OTUs) was checked by phylogenetic analyses. The most abundant ISUs were used as the representative sequence for each OTU and were aligned to the reference database. The Maximum Likelihood (ML) phylogeny was constructed using RAxML v.7.2.8 (Stamatakis 2014)

with GTR + G as model of evolution and 1000 replicates for the bootstrap analysis. The OTUs were then assigned to a morphospecies if they formed a clade supported by bootstrap values greater than 60, following our previous study (Visco *et al.* 2015 - Chapter 7) and that of Zimmermann *et al.* (2015). After the OTUs were assigned, DI-CHMOLTAXASSIGN scores were calculated based on the molecular data, using the D and G values given by the assigned species and the relative frequency of reads for the H factor.

8.4.7. Calculation of ecological values.

To calculate the autecological value D and the weighting factor G for each OTU, we rely on an approach similar to that used to create the DI-CH index itself (Hürlimann & Niederhauser 2007). For the calibration, the reference status for each site was given by the DI-CH values. For the calculation, only the OTUs with a relative frequency greater than 1% in at least one sample were kept. To find the autecological value D, the samples were grouped into 15 classes from 1 to 8 with a step of 0.5 according to their ecological status. For each OTU, the class with the highest 80^{th} percentile of relative frequencies was then kept as the D value. For the weighting factor G, the samples were grouped into 8 ecological classes. For each OTU, the distribution of 80% of its total abundance across the 8 classes was used to determine the weighting factor, using the following thresholds. 8: OTUs present in classes 1-3 and 7-8, corresponding to extreme ecological status. 4: OTUs present in 1 class only. 2: OTUs present in 2 classes. 1: OTUs present in 3 classes. 0.5: abundant OTUs present in a minimum of 4 classes or representing at least 3% in 3 classes. The workflow for this computation is summarised in FigureS 8.2. This calculation was first done with the complete dataset in order to compare the values given by the species assigned with the ones inferred from the DI-MOLTAXFREE approach.

8.4.8. Inference of the molecular index and cross-validation.

The molecular index was inferred from HTS data based either on those OTUs that could be assigned to morphospecies (DI-MOLTAXASSIGN) or all OTUs having a relative abundance of more than 1% in at least one sample of the dataset (DI-MOLTAXFREE). In the second case, the ecological values D and G were calculated as described above, while the H values were equal to the relative number of sequences (reads) for each OTU.

To evaluate the status of the taxonomy-free index (DI-MOLTAXFREE), two cross-validation tests were performed. In each case, the D and G values were recalculated without the tested samples. First, we used a leave-one-out cross-validation. To do so, one sample was removed from the dataset for the calculation of the value D and the factor G. Then, these D and G values were used to calculate the DI-MOLTAXFREE index of the removed sample. This process was repeated for each sample. Second, we performed a 25/75 cross-validation in which the D and G values were calculated for 65 sites and the evaluation of the index on the 22 remaining sample. The sites were randomly chosen and the validation was repeated for 1000 trials. The formula used to calculate the DI-MOLTAXFREE was the same as for the calculation of the morphological DI-CH presented in the introduction.

8.5. Results

8.5.1. HTS data.

The samples were sequenced in 4 independent Illumina runs. A total number of 2,206,456 good reads distributed across the 87 samples remained after filtering. The details for each run are described in TableS 8.4. The reads from all runs were dereplicated, resulting in 3079 ISUs. The ISUs were clustered into 663 OTUs. After chimera removal, a final number of 440 OTUs was used for further analyses. The distribution of these OTUs and the number of reads per site are detailed in TableS 8.5. The number of OTUs per site varied from 1 (FOS) to 77 (VXB) with a median value of 27 (TableS 8.6).

8.5.2. Morphological analysis.

Morphospecies were counted, and the relative abundance of each taxon was calculated for each site (TableS 8.7). A total of 269 morphospecies was identified across the 87 sites. The number of taxa per site varied from 5 (AMB) to 72 (PTH) with a median value of 24 (TableS 8.6). The ecological status values ranged between 1.61 (VXD) and 7.98 (AMB). The different ecological classes (very good, good, average, bad and very bad) were represented by 15, 26, 25, 12 and 9 sites respectively (TableS 8.8). These DI-CH values were used as references for the molecular analysis.

8.5.3. Taxonomic assignment.

We built a ML tree with our reference database and all OTUs (FigureS 8.3). After analysis, 152 OTUs (35%) were assigned to 43 morphospecies, of which 28 were found in the morphological analyses, while 15 matched to morphospecies not found microscopically in our samples. FigureS 8.4 shows the number of morphospecies recognised through morphological analysis, and in the genetic database and our HTS dataset after phylogenetic assignment. Almost 70% of the morphospecies (185/269) found in the morphological counts were not represented in the database, leaving 84 morphospecies that were represented in the database. However, among these only 28 species were assigned in the molecular dataset.

8.5.4. Ecological values comparison.

In this section, we compare the D and G values provided by the morphological database with those inferred from molecular data (DI-MOLTAXFREE). To do so, we selected 78 out of 152 taxonomically assigned OTUs that could be given the D and G values of the related morphospecies and represented more than 1% of the total number of sequences in at least one sample of the dataset. The selected OTUs were assigned to 23 different morphospecies. Their D and G values obtained from the morphotaxonomic database were compared to the values obtained by the taxonomy-free approach (Figure 8.2).

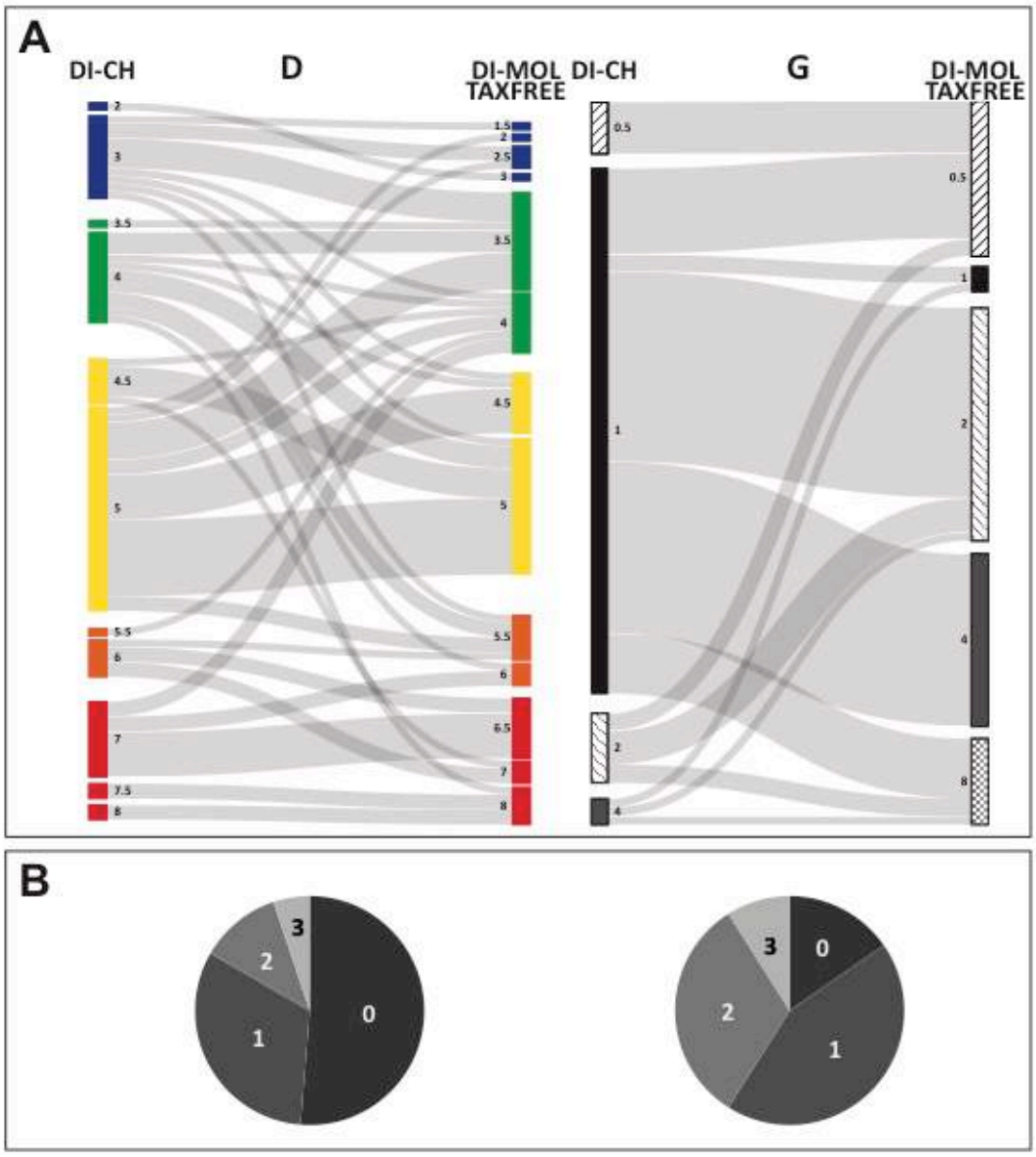


Figure 8.2 Comparison of DG values for 78 assigned OTUs. The figure is separated into two parts: D values on the left and G values on the right. For each value, the sankey diagram (A) represents the relationship between the values inferred from morphology (DI_CH), and those inferred by the molecular index (DI-MOLTAXFREE). The links represent the assigned OTUs. Pie charts (B) represent the proportion of assigned OTUs as a function of the number of classes that change between their two values. No class changes are indicated in black, one class changes in dark grey, two classes change in medium grey and 3 classes change in light grey. For the D value, the class are separated as follows: 1-3.5: very good; 3.5-4.5: good; 4.5-5.5: average; 5.5-6.5: bad; 6.5-8: very bad and the scale of the G value is 0.5, 1, 2, 4 and 8.

More than half of the 78 OTUs show a morphological and a molecular D value indicating the same ecological status and 15% of the OTUs show exactly the same G values. These numbers increase to 83% and 59% with a maximum of one change for the D value and G value, respectively. For both values, less than 10% show a drastic change of three categories difference. The D and G values are given for each assigned OTUs in TableS 8.9.

8.5.5. Relative abundance.

Besides the ecological values D and G, we also compared the relative abundance of each species based on microscopic counts of specimens found at a particular site to the relative abundance of the corresponding OTU represented by the number of HTS reads (sequences). In FigureS 8.5, we provide the results of this comparison for the 23 assigned morphospecies. In the majority of cases, we observed that the relative abundance of sequences is higher compared to the abundance of specimens (circles are located above the triangles). However, in few cases (e.g. *Sellaphora seminulum*) the opposite is observed. We calculated the correlation between the morphological and the molecular abundance for the four most abundant species. As shown in FigureS 8.6, three species (*Cocconeis placentula, Eolimna minima, Planothidium lanceolatum*) showed a strong correlation (R^2=0.79, 0.76 and 0.90 respectively, with p-values < 0.0001), whereas *Achnanthidium minutissimum* did not (R^2=0.41).

8.5.6. Diatom Index.

The molecular scores inferred using the taxonomic assignment (DI-MOLTAXASSIGN) and the taxonomy-free method (DI-MOLTAXFREE) were compared in order to examine the coverage of the HTS dataset by each of those two approaches. The range of the values calculated by the DI-MOLTAXASSIGN was 3.00-7.98 compared with 2.7-6.93 for the DI-MOLTAXFREE method. As illustrated in Figure 8.3, the taxonomic assignment method utilized 36% of the reads, whereas the taxonomy-free approach utilized 98% of the dataset. Similar proportions were found in the number of OTUs, with 38% and 85% of OTUs included in the taxonomic assignment and taxonomy-free approaches, respectively. For only one site (HEP), the number of OTUs used in the taxonomic assignment method was greater than in the taxonomy-free approach. This particular site shows a huge genetic diversity in *Cocconeis placentula* (17 different OTUs), although 6 of them were very rare and therefore were removed from the taxonomy-free analysis.

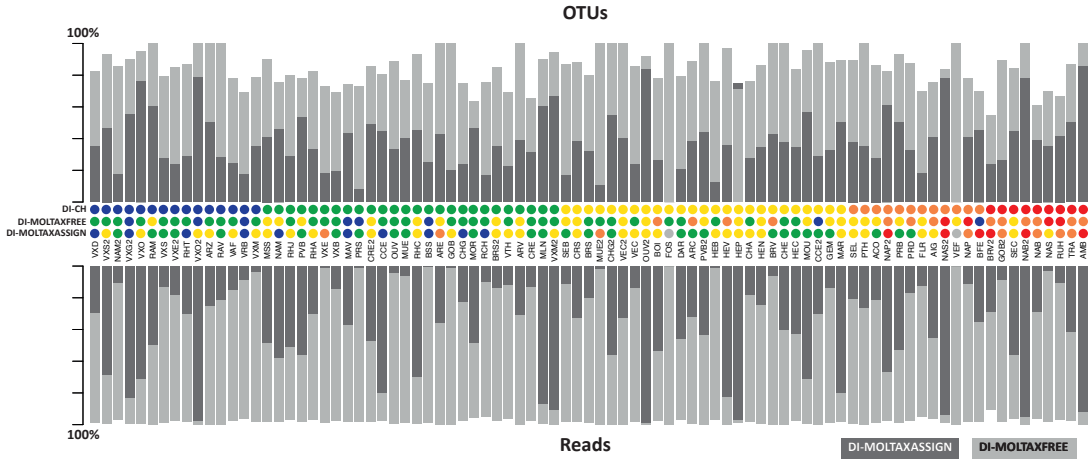


Figure 8.3 Percentage of the HTS dataset used by the taxonomic assignment (dark grey) and the molecular index (light grey) methods for each site. The OTUs are illustrated at the top and the reads at the bottom. In the middle, the coloured dots represent the ecological status given by the calculation of DI-CH values with Morphology (DI-CH), Molecular index (DI-MOLTAXFREE) or Taxonomic assignment (DI-MOLASSIGN). For the molecular index, the results of the leave-one-out cross-validation are used. The very good, good, average, bad and very bad statuses are represented with blue, green, yellow, orange and red colour, respectively.

The central part of Figure 8.3 indicates the ecological status inferred by each approach. The two molecular methods (taxonomic assignment and taxonomy-free) give the same ecological status for 45% (38/85) of the samples; 14 of them are congruent with the morphological evaluation. For 38% (33/87) of the samples, the DI-MOLTAXFREE gave the same class as the DI-CH compared with 30% (26/85) for the DI-MOLTAXASSIGN. For two sites (FOS and VEF), no sequences could be assigned and, therefore, no taxonomic assignment evaluation was possible.

The taxonomic assignment and taxonomy-free molecular indices are compared further in Figure 8.4, which shows the correlations of each index with the values of the morphological index (DI-CH) and indicates the difference compared to the values of DI-CH. The correlation between DI-MOLTAXASSIGN and the DI-CH (R^2 = 0.57 and p-value < 0.0001) is lower than the correlation between DI-MOLTAXFREE and the DI-CH (R^2 = 0.67 and p-value < 0.0001). The values of the indexes differ by less than 1 in 77% of the samples for the DI-MOLTAXFREE, compared to 52% for the DI-MOLTAXASSIGN. The proportion of sites correctly assessed with the DI-MOLTAXASSIGN increases to 88% for the most sampled sites belonging to the good and average classes, which are the best represented in our dataset. The under-

sampled classes show less good results, with 75%, 67%, and 46% of correctly assessed sites for the bad, very bad, and very good classes, respectively (FigureS 8.7).

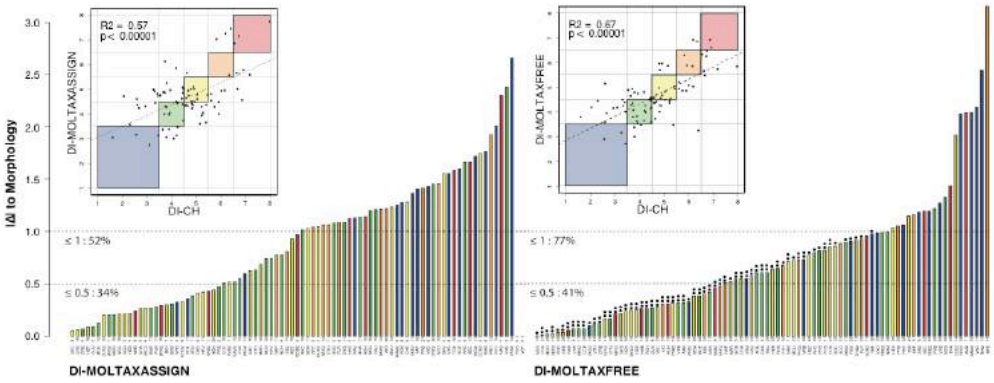


Figure 8.4 Comparison between the DI-CH values given by morphology and molecular methods (taxonomic assignment on the left and leave-one-out cross-validation on the right). For each method, two types of graphics are represented. The scatter plots show the relationships between the DI-CH inferred from morphological (x-axis) and the molecular methods (y-axis). Coloured boxes represent the ecological status given by the DI-CH (blue: very good, green: good, yellow: average, orange: poor, red: bad). The regression line for all samples is represented by dashed line and the R^2 and p-value are indicated for each graph. The bar plots show for each site the absolute difference between the DI-CH values given by the morphology and the molecular methods. Sites are coloured in function of their DI-CH value (blue: very good, green: good, yellow: average, orange: poor, red: bad). Above each site name, the number of OTUs used to calculate the index is indicated. Dashed lines are drawn at the 0.5 and 1 difference thresholds. Percentages of sites below these thresholds are indicated in the graphs. For each site, the black dots show the results of the 25/75 cross-validation. One dot is used if at least 70% of the replicates gave an absolute difference with the morphological DI-CH below 1 and two dots are used if this percentage is above 90%.

In the case of DI-MOLTAXFREE, the leave-one-out cross-validation test was used to better illustrate the comparison with DI-MOLTAXASSIGN (Figure 8.4). However, similar results were obtained using the 25/75 cross-validation tests (shown as stars in Figure 8.4 and illustrated in FigureS 8.8 and FigureS 8.9). The seven most problematic sites remain the same in the two cross-validation tests. In those cases, the difference compared to the DI-CH is greater than 1.5 for the leave-one-out analysis and for the 25/75 cross-validation, less than 6% of the trials show a

difference below 1. Four out of these seven sites belong to the very good quality class.

8.6. Discussion

8.6.1. Overcoming the taxonomic assignment issue.

The main objective of this study was to test whether the step of taxonomic assignment is necessary to calculate a molecular diatom index with eDNA data. Previous studies highlighted various biases introduced by this step but still kept it as an integral part of their analyses (Kermarrec *et al.* 2014; Zimmermann *et al.* 2015; Visco *et al.* 2015 - Chapter 7). The present study shows that the molecular index computed with (DI-MOLTAXASSIGN) or without (DI-MOLTAXFREE) taxonomic assignment is not significantly different. Moreover, we observe a higher correlation between morphological and molecular indices in the case of the taxonomy-free approach (Figure 8.4), suggesting that taxonomic assignment may not be essential for eDNA-based diatom monitoring.

Our results suggest that the main benefit of taxonomy-free approach lies in its much higher data coverage compared to the use of taxonomic assignment. The latter step considerably reduces the amount of available data due to the incompleteness of genetic reference databases, which comprise only 31% of the morphospecies identified in this study. This small number is reduced further to 10%, as 56 morphospecies present in genetic database (many belonging to the genus *Navicula*) could not be correctly assigned because of the lack of resolution of the 18S V4 marker. The selection of another marker (e.g. *rbc*L proposed by MacGillivary & Kaczmarska 2011; Kermarrec *et al.* 2013) could probably improve the phylogenetic assignment for some species. However, it is uncertain whether the global data coverage would be much better.

Even if all morphospecies were sequenced with a more highly resolving marker, the taxonomic assignment will still be compromised by the issue of cryptic genetic diversity. It is well known that, in common with many other protists, the majority of diatom morphospecies are represented by large numbers of OTUs that are not always monophyletic (Beszteri *et al.* 2007; Amato *et al.* 2007; Rimet *et al.* 2014; Van den Wyngaert *et al.* 2015; Rovira *et al.* 2015). For example, *Cocconeis placentula* is represented in our data by 17 OTUs. Although this species complex has been split

morphologically into several subspecies, their correspondence to numerous OTUs branching within the *C. placentula* clade is not well established. As a result, it is not possible to use different ecological values assigned to these subspecies and, conversely, to take advantage of ecological values assigned to *C. placentula* OTUs by the taxonomy-free approach. Regarding the practical application of the diatom index, the main problem with the taxonomic assignment approach is not so much the lack of correspondence between OTUs and morphospecies, but the difficulty of avoiding the errors introduced by the direct translation of ecological values associated with morphospecies to corresponding OTUs.

8.6.2. Accuracy of ecological values.

By overcoming the step of taxonomic assignment, our method provides an independent assessment of ecological values. These values have been estimated directly from the HTS data, using morphological analyses as a reference to establish the ecological status of each site. Since such estimations have never been attempted before, we examine the difference between these newly calculated values and those given by morphological observations. Although this comparison could only be done on a few assigned OTUs and a limited number of sites, the results shown in Figure 8.2 and FigureS 8.6 are promising.

In the case of the autecological D values, the same ecological status was obtained for most of the OTUs. On the contrary, the variations of the weighting factor G are wider, with most of the OTUs having G values more or less equally distributed between 0.5 and 8, while most morphospecies are characterized by a G value of 1 (Figure 8.2). As the G value reflects the occurrence of species/OTU across the sites, it is possible that these wider variations are related to the presence of extracellular DNA that can be dispersed over large distances (Deiner & Altermatt 2014). Alternatively, it is possible that the G values are affected by low amplification efficiency, which artificially reduces the range of occurrence, making the ecological tolerance of a given OTU appear narrower than in morphological surveys.

The accuracy of the DI-MOLTAXFREE also depends on the stability of D and G values during cross-validation. As illustrated in FigureS 8.10, the values of the weighting factor G are relatively stable, with 83% of 228 analysed OTUs changing less than one category. In the case of D values, the variations are greater, although they rarely exceed 2 points. These large variations can be an effect of under-

sampling, limiting the number of sites where an OTU occurs. This probably applies in the case of OTU 427, which is responsible for the highest difference between the DI-MOLTAXFREE and the DI-CH index found at the site BFE. Another possibility is that morphological misidentification leads to an erroneous assessment of some sites where an OTU is present. Such misidentifications can occur when the samples are processed routinely without a detailed scanning electron microscope examination of each specimen. To avoid such errors, it is necessary to stabilize the D and G values by increasing the number of sites and adapting the D and G values to the specificities of molecular data.

8.6.3. The issue of relative abundance.

The third factor that influences the molecular index is relative abundance. This is also examined here. It is widely accepted that different technical and biological biases impact the relative abundance of specimens and sequences, making impossible the use of quantitative data in HTS surveys (Elbrecht & Leese 2015). However, this was not confirmed by the present study, at least as far as the most abundant species are concerned (FigureS 8.6). The same tendency was observed in other protists, such as foraminifera, where the same species dominated morphological and molecular assemblages (Pawlowski *et al.* 2014a). We could speculate that this relatively good match between the numbers of specimens and sequences of abundant species is reinforced by the exponential character of PCR amplification. As shown in the case of *C. placentula* and *E. minima* (FigureS 8.6), when a species is very abundant in microscopic counts, it is often even more abundant in HTS reads. However, this is not always true. For example, at some sites the relative abundance of specimens of *Sellaphora seminulum* exceeds the abundance of reads (FigureS 8.7), suggesting that the PCR amplification may not be very efficient in this species.

In general, the importance of quantitative biases seems to be reduced in the case of small, single-cell organisms such as diatoms or foraminifera. However, the biomass of protistan cells can also vary considerably and the variability of rRNA copy numbers has been demonstrated in some diatoms (Alverson & Kolnick 2005; Godhe *et al.* 2008) and other protists (Gong *et al.* 2013; Weber & Pawlowski 2014). The taxonomy-free approach avoids this problem, because it does not involve the direct comparison of the relative abundance of specimens and sequences. Assuming that the PCR and other technical biases are the same across the samples for a given

OTU (as long as the experimental conditions remain unchanged), the impact of these biases on the accuracy of taxonomy-free molecular index will be less important and easier to control than in the case of taxonomic assignment approach. Nevertheless, the formulae on which current indices are based are not adapted specifically for quantitative HTS data; a special effort will be required to address this issue in future studies.

8.6.4. Limitations of taxonomy-free approach.

Although, as mentioned above, the taxonomy-free index has many advantages, it also has some important limitations that have to be overcome before the index can be used routinely. In view of our results, the most important factor causing incongruence between molecular and morphological indices is the lack of comprehensive sampling. As illustrated in Figure 8.4, the DI-MOLTAXFREE approach considerably reduced the number of incorrectly assigned sites compared to the DI_MOLTAXASSIGN method. Yet, there are still sites that differ significantly from their status according to the morphological DI-CH method, and remarkably, most of them belonging to the under-sampled classes of very bad, bad and very good water quality.

The effect of under-sampling is particularly dramatic in the case of very good (blue) sites, half of which lie outside the 1-point limit (FigureS 8.5). This can be explained by the fact that these very good water quality sites are not characterized by specific indicator species but rather by different species-rich communities (Whitton *et al.* 1991; Hürlimann & Niederhauser 2007), which might be difficult to reconstruct without an extensive sampling. Conversely, the lack of congruence observed in the case of the bad and very bad quality sites can be explained the fact that these sites are usually characterized by high abundances of a few indicator species (Hill *et al.* 2001; Stevenson *et al.* 2010). When these sites appear rarely in the dataset because of under-sampling, the absence of indicator species/OTUs in cross-validation studies may lead to the totally wrong assignment of a given site, as possibly happened in the case of sites AMB and BFE in our analyses (Figure 8.4).

These few examples highlight the importance of sampling effort to ensure the accuracy of ecological values associated with OTUs in the taxonomy-free approach. However, even the most extensive eDNA sampling will not be able to alleviate all limitations of using OTUs rather than morphospecies to evaluate the quality of the

environment. In particular, the metabarcoding data is unable to provide the kind of ecological information that is available through microscopic observations. For example, the list of OTUs and their relative frequencies says nothing about the physiological state of species, which can be measured by the proportion of teratological morphotypes in microscopic analyses (reviewed in Falasco *et al.* 2009). In general, the extensive knowledge of the taxonomy, biology and ecology of diatoms that can be derived from microscopic observations cannot be easily applied to the interpretion of molecular data. Therefore, the taxonomy-free index should be considered as a complementary tool rather than as a replacement for morphology-based studies.

8.6.5. Future challenges and perspectives.

Our study raises several questions concerning the applicability of taxonomy-free approach in routine biomonitoring. Some of these questions, concerning the geographic range of OTUs and their ecological preferences, can hardly be answered without extensive sampling. Therefore, to further test the taxonomy-free index, the most important challenge is to obtain data from a much broader geographic area and from more diverse habitats. As shown by our results, the assessment of water quality is relatively good in the case of sites of average and good ecological status that dominate in our sampling. On the contrary, the diatom communities of the very good and very bad quality sites are not yet sufficiently represented in our datasets and, therefore, the inferred ecological values are not accurate enough. This highlights the importance of having not only numerous sites but also sufficiently varied sampling habitats to cover the widest diversity possible.

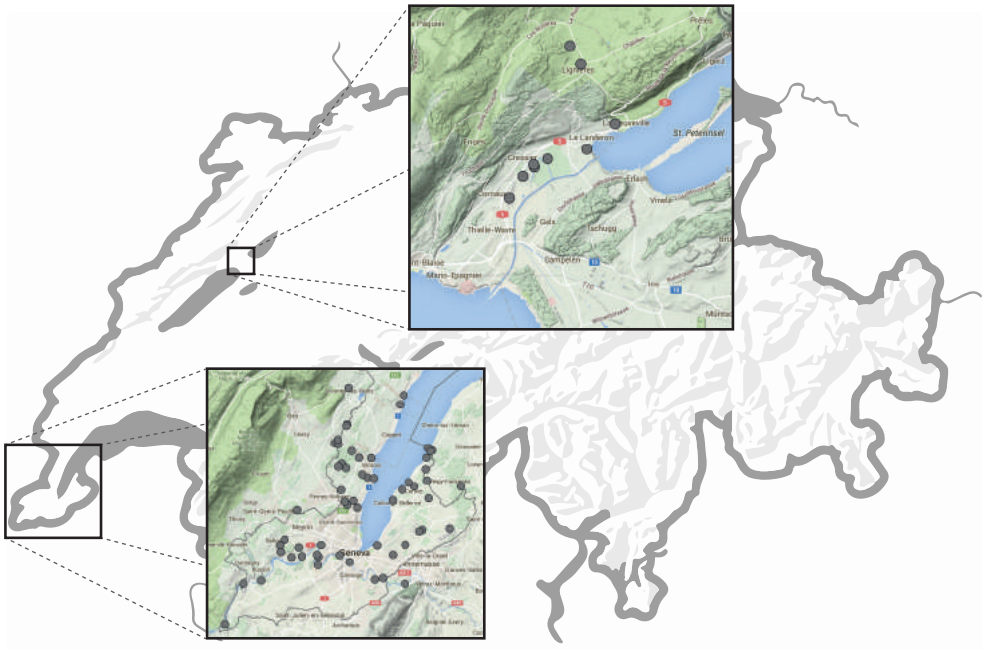
Another important challenge is the calibration of the taxonomy-free index. In the present study, we relied on a well-established diatom index that is routinely used to characterise water quality in Swiss rivers and streams. The Swiss index, and other diatom indices currently available, are based on decades of microscopic data collection that has provided comprehensive information about diatom species ecology and distribution. These morphological data are essential to calibrate the taxonomy-free index and ensure its accuracy and robustness. However, where morpho-taxonomic data are not available due to a lack of taxonomic expertise, other types of data, such as chemical parameters or macro-invertebrate surveys, could serve as alternative calibration options. The most readily available data are chemical

parameters. Yet, to be useful for diatom index calibration the chemical analyses have to be conducted over longer periods of time. Depending on the diversity and geographic ranges of diatom OTUs, calibration of the taxonomy-free index would be necessary for different habitats and geographic localities. However, once the index is properly calibrated, the ecological values for each OTU will be more stable and the values of diatom index will be more reliable.

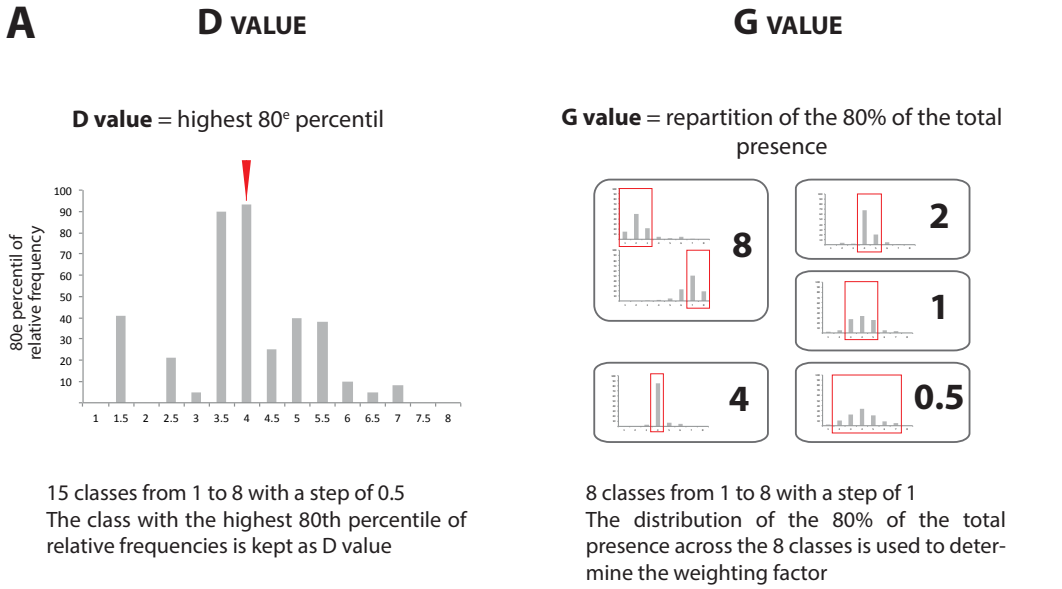
To conclude, our study demonstrates the great potential of the taxonomy-free molecular index for environmental biomonitoring. Although our work focuses on diatoms and the specific case of the Swiss diatom index, the taxonomy-free approach could easily be applied to other groups of single-cell bioindicators, such as ciliates (Lee *et al.* 2004; Chen *et al.* 2008; Jiang *et al.* 2011), and foraminifera (Schönfeld *et al.* 2012; Vidovic *et al.* 2014; Alve *et al.* 2016). New molecular indices could also be tested for microbial and meiofaunal taxa that are not currently used as bioindicators. The implementation of these new indices would help to extend the range of monitored sites and increase the frequency of monitoring. Once established, molecular indices could provide a fast, easily standardized and highly sensitive tool that complements the current morphology-based methods available for the water quality assessment.

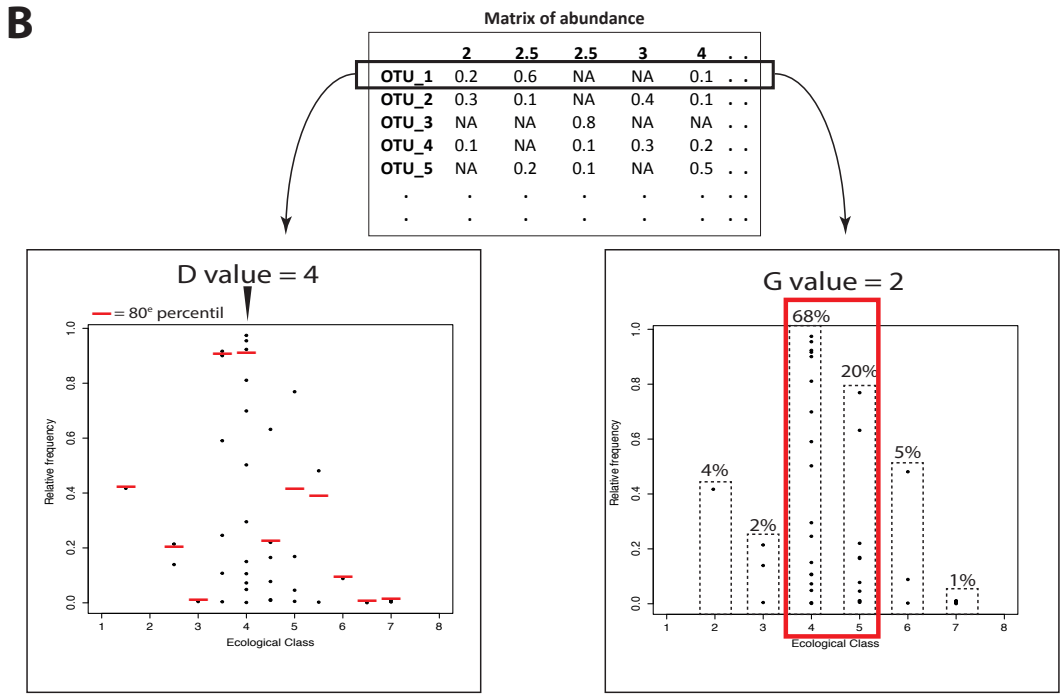
8.7. Supplementary data

FigureS 8.1 Map of sampling sites



FigureS 8.2 A. Schematic representation of the calculation of the D and G value for the molecular method. Only the OTUs present in at least 1% in one sample in the entire dataset are used. B. An example illustrating how D and G values are calculated.

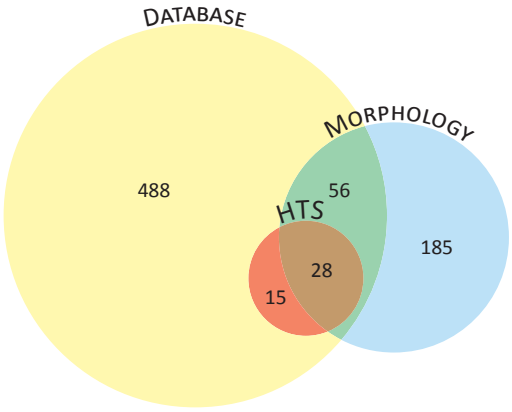




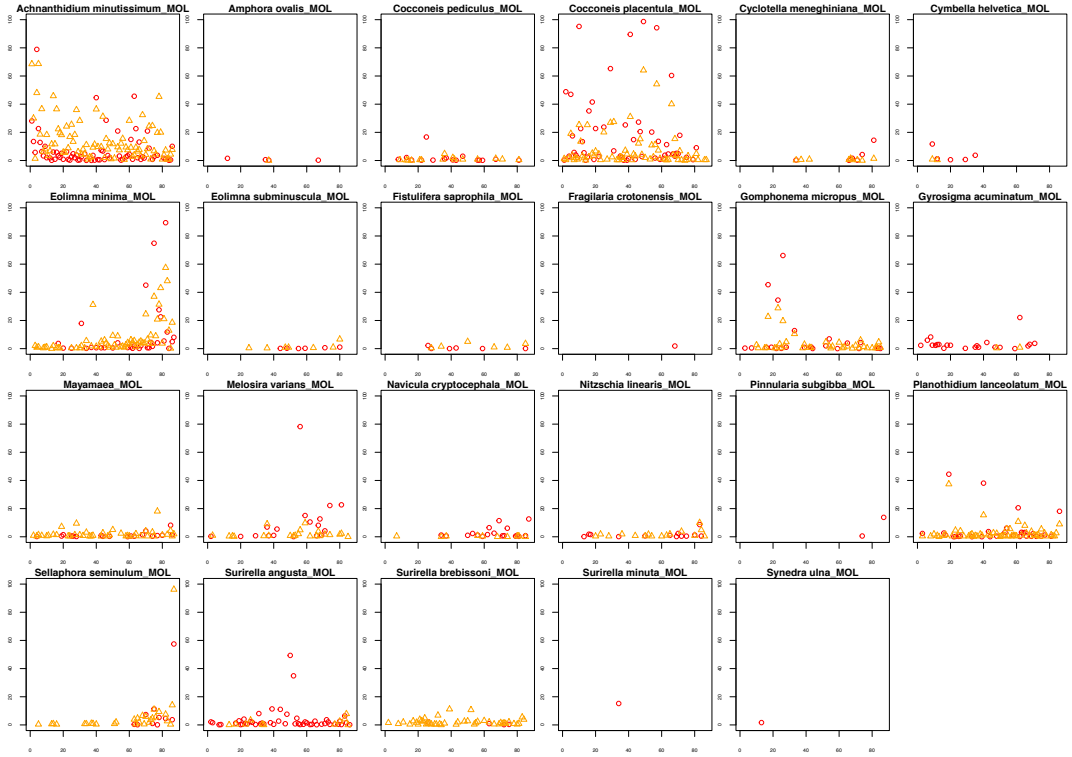
FigureS 8.3 RAxML tree with sequences from the database and the OTUs from the HTS analysis

Available on http://onlinelibrary.wiley.com/doi/10.1111/1755-0998.12668/full

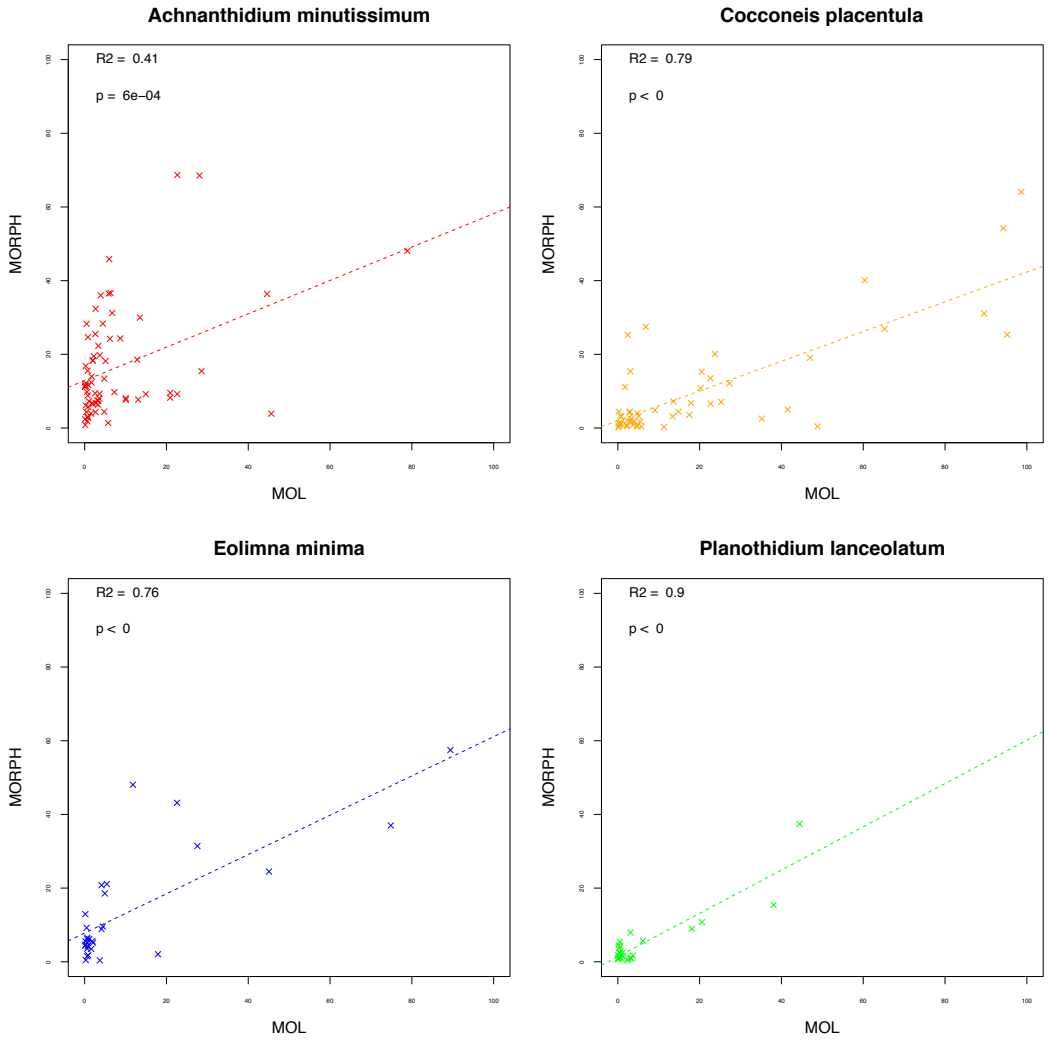
FigureS 8.4 Venn diagram of morphospecies represented in the database (yellow), morphological analysis (blue) and found in HTS dataset by taxonomic assignment method (red).



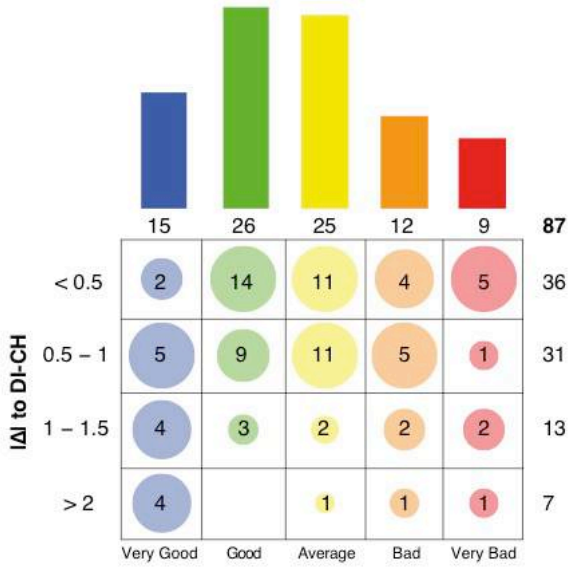
FigureS 8.5 Scatter plot of the relative frequency for all the assigned species. Counts (orange triangle) and reads (red circle) are normalized for each sample. For four species, no morphological data were available.



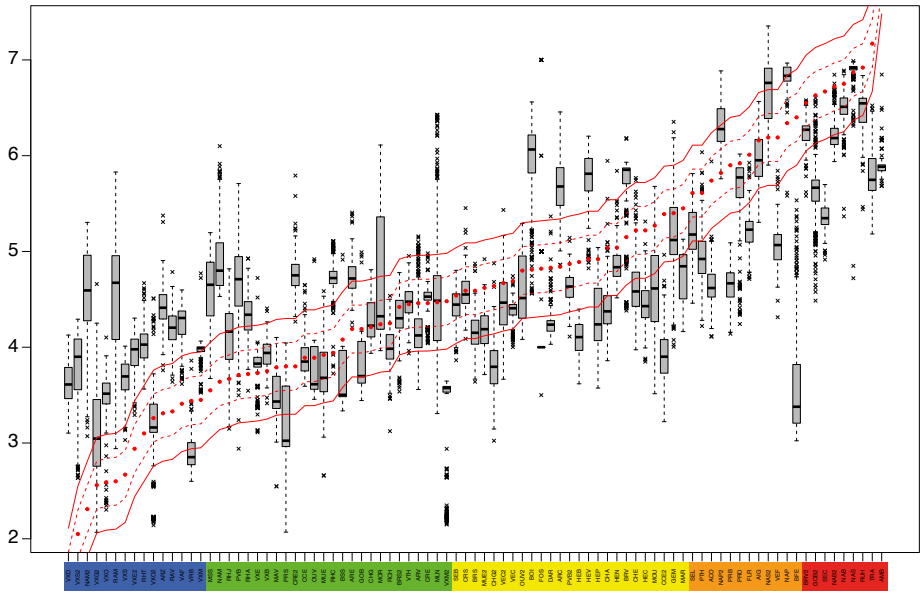
FigureS 8.6 Scatter plot of the relative frequency for the 4 most represented morphospecies in the HTS dataset. Counts (MORPH) and reads (MOL) are normalized for each sample. For each graph, the regression line for all samples is represented by a dashed line and the R2 and p-value are indicated.



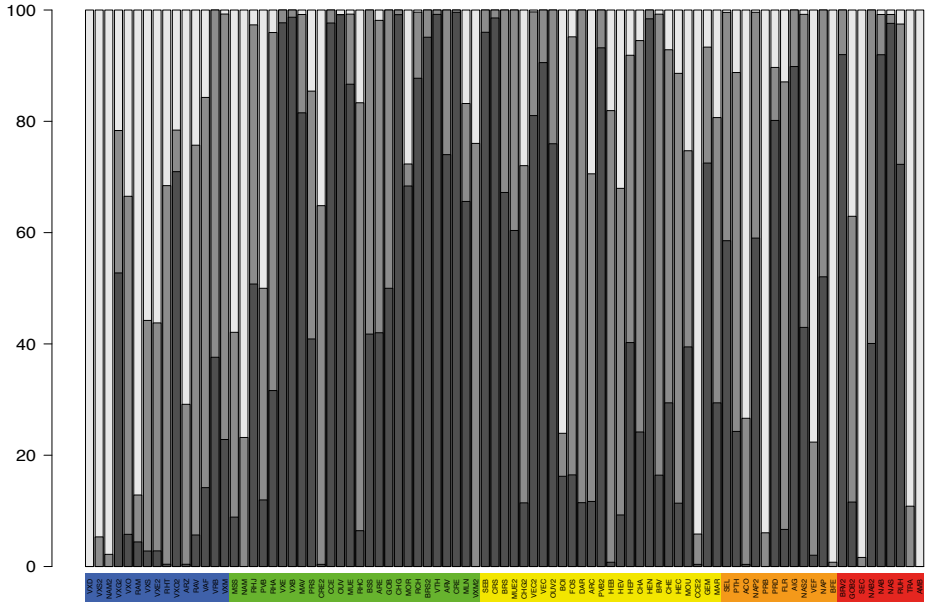
FigureS 8.7 Graphical table representing the cross-validation results. Samples are grouped into the five ecological classes (very good, good, average, bad and very bad) in the horizontal axis. Sample in each class are distributed in function of the absolute difference between the DI-CHmorpho and the cross-validation. The areas of the circles are proportional to the data for each class separately. The bars represent the number of sample in each class



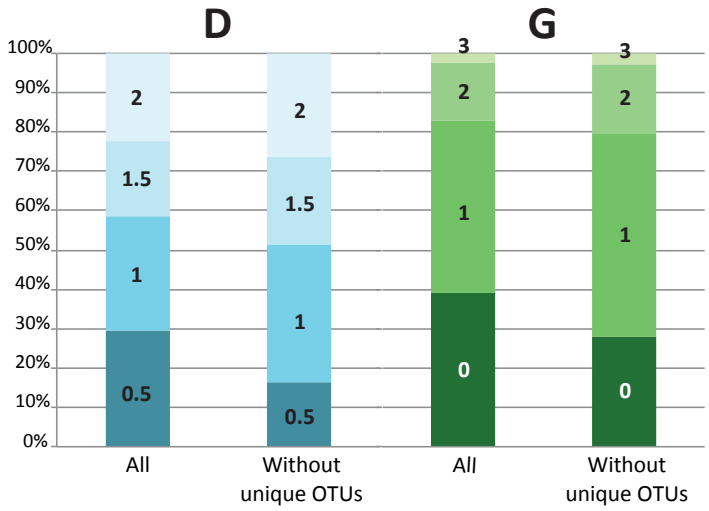
FigureS 8.8 Box plot of the DI-MOLTAXFREE Cross-Validation 25:75 test. The red dots correspond to the DI-CH value. A difference of 1 with the morphology is represented by a red continuous line and a difference of 0.5 is represented by a red dashed line. The sites are classed in function of their DI-CH, from best to worst water quality.



FigureS 8.9 Graphical representation of the DI-MOLTAXFREE Cross-Validation 25:75 test. For each sites, the proportion of replicates with a absolute difference to the DI-CH below 0.5 (dark grey), below 1 (grey) or above 1 (light grey) is represented. The sites are classed in function of their DI-CH, from best to worst water quality.



FigureS 8.10 Bar plots representing the proportion of D (blue) and G (green) values in function of their change during the cross-validation test. For the D value, the scale is from 1 to 8 with a 0.5 step and the splits correspond to the difference between the values. For the G value the scale is 0.5, 1, 2, 4, 8 and one change correspond to one change of category. Both values are presented with (All) and without (Without unique OTUs) the OTUs present in only one site in the dataset.



TableS 8.1 Illumina run code, station code, locations, sampling date and geographic references for each site used in this study.

Run	Code	Location	Date	N	E
DIATOM01	MAR	Marnot - embouchure	23.09.13	46.2917829744536	6.2525148226365
DIATOM01	CHE	Cherre - amont chemin Armand-Dufaux	10.09.13	46.2641266680935	6.2084973356855
DIATOM01	FOS	Fossaz - amont chemin du Milieu	10.09.13	46.2542189611206	6.1955359944258
DIATOM01	MOU	Moulin - aval route d'Hermance	09.09.13	46.2943644324306	6.2420058781010
DIATOM01	HEV	Hermance - les Verrières	10.09.13	46.2674503269817	6.2896224591666
DIATOM01	MLN	Moulanais - amont chemin de la Montagne	10.09.13	46.2023080494067	6.1957335256471
DIATOM01	CHA	Chamburaz - embouchure	10.09.13	46.3015570769916	6.2494452908081
DIATOM01	HEB	Hermance - embouchure	23.09.13	46.3037483305912	6.2439452136647
DIATOM01	HEP	Hermance - Pont de Bouringe	23.09.13	46.3006814993215	6.2474526366922
DIATOM01	TRA	Traînant - Traînant	23.09.13	46.2111335005438	6.1743805705553
DIATOM01	ACO	Aisy - Côte d'or	23.09.13	46.2672206526359	6.2251655579588
DIATOM01	HEN	Hermance - pont Neuf	23.09.13	46.2728733256966	6.2424192439071
DIATOM01	AMB	Aisy - embouchure	23.09.13	46.2709413692631	6.2174223159857
DIATOM01	HEC	Hermance - Pont de Crévy	23.09.13	46.2834739573656	6.2409531343499
DIATOM01	ARC	Aisy - route de Covéry	23.09.13	46.2560324193688	6.2450595739550
DIATOM01	PRB	Paradis - embouchure	24.09.13	46.2256169344669	6.2360049188309
DIATOM01	SEL	Seymaz - pont Ladame	24.09.13	46.2116235120562	6.2128208658723
DIATOM01	SEB	Seymaz - embouchure	24.09.13	46.1800909149041	6.1820310578403
DIATOM01	SEC	Seymaz - pont de Choulex/Montagnys	24.09.13	46.2240906477902	6.2320852189619
DIATOM01	PRD	Paradis - Les Doillets	24.09.13	46.2268271000039	6.2742197384710
DIATOM02	GEM	Grebattes - embouchure	13.03.14	46.2001104700812	6.0925547290088
DIATOM02	DAR	Maison-Carrée - Bois de Bay	13.03.14	46.1997615701834	6.0562836345793
DIATOM02	FLR	Montfleury - aval jardins familiaux	13.03.14	46.2092262912882	6.0661494899460
DIATOM02	NAS	Avril - Satigny	03.03.14	46.2110761677475	6.0399264175209
DIATOM02	NAB	Avril - Bourdigny	11.03.14	46.2166025516207	6.0466534300789
DIATOM02	NAP	Avril - Peney	11.03.14	46.2048790154452	6.0407982172074
DIATOM02	NAM	Maille - La Maille	11.03.14	46.2449085007345	6.0640161359387
DIATOM03	NAM2	Maille - La Maille	08.09.14	46.2449085007345	6.0640161359387
DIATOM03	NAB2	Avril - Bourdigny	08.09.14	46.2166025516207	6.0466534300789
DIATOM03	NAS2	Avril - Satigny	08.09.14	46.2110761677475	6.0399264175209
DIATOM03	NAP2	Avril - Peney	08.09.14	46.2048790154452	6.0407982172074
DIATOM03	BFE	Bois-des-frères - Embouchure	08.09.14	46.2114456185367	6.0962256626053
DIATOM03	ARE	Arve - Ecole de Medecine	15.09.14	46.1952666686722	6.1359482032658
DIATOM03	ARZ	Arve - Pont de Zone	15.09.14	46.1745776985265	6.2124621061332
DIATOM03	ARV	Arve - Vessy	15.09.14	46.1786191243934	6.1711212891023
DIATOM03	RHA	Rhône - amont Allondon	18.09.14	46.1777352623214	6.0142314448704
DIATOM03	RHT	Rhône - Touvière	18.09.14	46.1748119228603	5.9895711053417
DIATOM03	RHC	Rhône - Conflan	18.09.14	46.1359547550625	5.9639957127308
DIATOM03	RAV	Rhône - aval STEP	25.09.14	46.2023372927514	6.0906853294790
DIATOM03	RHJ	Rhône - amont Jonction	25.09.14	46.2017676187392	6.1224461399440
DIATOM03	RAM	Rhône - amont STEP/Aire	25.09.14	46.1924563371588	6.0919669409397
DIATOM03	RCH	Rhône - Chèvre	25.09.14	46.2006778445437	6.0702854653446
DIATOM03	BOI	NE - Boiron	02.10.14	47.0364662787530	7.0305721000608
DIATOM03	MOR	NE - Mortruz	02.10.14	47.0442896885140	7.0375526953099
DIATOM03	VTH	NE - Vieille Thielle - Pont	02.10.14	47.0475299503600	7.0433188055971
DIATOM03	RUH	NE - Ruhaut de Cressier	02.10.14	47.0484465259970	7.0430488500840
DIATOM03	PTH	NE - Petite Thielle	02.10.14	47.0504501150540	7.0502723851251
DIATOM03	AIG	NE - Aiguesdeurs	02.10.14	47.0539084615630	7.0706738395553
DIATOM03	VRB	NE - Vaux - Route de Bâle	02.10.14	47.0628512923830	7.0855399856431
DIATOM03	VAF	NE - Vaux - Aval fromagerie Lignières	02.10.14	47.0839690006470	7.0675440577301
DIATOM03	MSS	NE - Moulin - sous la Scie	02.10.14	47.0905172524950	7.0619671662762
DIATOM05	CHG	Chânat - amt Gobé	12.03.15	46.2534520918533	6.1410501431382
DIATOM05	GOB	Gobé - Amt Colovrex	12.03.15	46.2525912318470	6.1293858151625
DIATOM05	MAV	Marquet - amt Vireloup	12.03.15	46.2639383148193	6.1245396365019
DIATOM05	VEC	Vengeron - amt CFF	12.03.15	46.2470290116341	6.1464434264592
DIATOM05	BRV	Braille - aval bassin retention	16.03.15	46.2945206140113	6.1487539289172
DIATOM05	CRS	Creuson - amt rte Sauvergny	16.03.15	46.3008780570326	6.1387335190395
DIATOM05	BSS	Brassu - amt rte Suisse	16.03.15	46.3450848758620	6.2056197861548
DIATOM05	CRE	Creuson - emb	16.03.15	46.2849087011011	6.1310059299667

DIATOM05	BRS	Braille - amt rte Suisse	16.03.15	46.2941280195456	6.1662158289262
DIATOM05	PRS	Pry - amt rte Suisse	16.03.15	46.3537775471141	6.2107495244715
DIATOM05	VXB	Versoix - Bossy	17.03.15	46.2889371974045	6.1253935744691
DIATOM05	VXM	Versoix - Mâchefer	17.03.15	46.2786432136786	6.1534085741741
DIATOM05	MUE	Munet - emb	17.03.15	46.3261809698058	6.1325439069373
DIATOM05	PVB	Pissevache - rte Vieille Bâtie	17.03.15	46.2862720131969	6.1205920532548
DIATOM05	CCE	Crève-cœur - emb	17.03.15	46.2758000506382	6.1605473170487
DIATOM05	VXE	Versoix - emb	17.03.15	46.2750013312722	6.1695187672985
DIATOM05	VXS	Versoix - Sauvergny	19.03.15	46.3115934019758	6.1200411238364
DIATOM05	VXO	Versoix - aval Oudar	19.03.15	46.3074151163746	6.1205320782379
DIATOM05	VXD	Versoix - amt Divonne	19.03.15	46.3608438278437	6.1346982962355
DIATOM05	OUV	Oudar - aval STEP	19.03.15	46.3084398618494	6.1158335900318
DIATOM05	CRE2	Creuson - emb	21.09.15	46.2849087011011	6.1310059299667
DIATOM05	MUE2	Munet - emb	21.09.15	46.3261809698058	6.1325439069373
DIATOM05	PVB2	Pissevache - rte Vieille Bâtie	21.09.15	46.2862720131969	6.1205920532548
DIATOM05	CCE2	Crève-cœur - emb	21.09.15	46.2758000506382	6.1605473170487
DIATOM05	OUV2	Oudar - aval STEP	21.09.15	46.3084398618494	6.1158335900318
DIATOM05	VEF	Vengeron - Fortaille	25.09.15	46.2497517992879	6.1328261755823
DIATOM05	BRV2	Braille - aval bassin retention	25.09.15	46.2945206140113	6.1487539289172
DIATOM05	GOB2	Gobé - Amt Colovrex	25.09.15	46.2525912318470	6.1293858151625
DIATOM05	CHG2	Chânat - amt Gobé	25.09.15	46.2534520918533	6.1410501431382
DIATOM05	BRS2	Braille - amt rte Suisse	25.09.15	46.2941280195456	6.1662158289262
DIATOM05	VEC2	Vengeron - amt CFF	25.09.15	46.2470290116341	6.1464434264592
DIATOM05	VXO2	Versoix - aval Oudar	28.09.15	46.3074151163746	6.1205320782379
DIATOM05	VXS2	Versoix - Sauvergny	28.09.15	46.3115934019758	6.1200411238364
DIATOM05	VXG2	Versoix - Grilly	28.09.15	46.3248199023843	6.1315378736755
DIATOM05	VXM2	Versoix - Mâchefer	28.09.15	46.2786432136786	6.1534085741741
DIATOM05	VXE2	Versoix - emb	28.09.15	46.2750013312722	6.1695187672985

TableS 8.2 List of database entries description with their NCBI or Rsyst accession number. Environmental sequences (ENV) are marqued. The DB code corresponds to the name of the sequences in the Figure S2.

Available on http://onlinelibrary.wiley.com/doi/10.1111/1755-0998.12668/full

TableS 8.3 List of primers and tags used in this study.

		Forward 5'-3'		**Reverse 5'-3'**
Primer	*DIV4for*	GCGGTAATTCCAGCTCCAATAG	*DIV4rev3*	CTCTGACAATGGAATACGAATA
Tag	*B*	ACATGATG	*B*	CGCACGTG
	C	AGATGTAT	*C*	CGCAGACA
	D	AGTACTGA	*D*	CTAGCATG
	E	AGTGTCAG	*E*	CTCGCTCA
	G	ATCATCAT	*G*	CTGCATCG
	H	ATCTGAGA	*H*	GCACATGT
	I	GCTAGAGA	*I*	GCAGAGCG
	J	GCTGTCTA	*J*	GCGCGACG
	L	GTATCTAG	*L*	GTGACACA
	M	GTCACTGT	*M*	GTGCAGTA
	N	GTCTGCTG	*N*	TCGACTCG
	O	TCATGTGA	*O*	TCGAGAGA
	Q	TGAGTATG	*Q*	TGACGACT
	R	TGATCAGT	*R*	TGAGCTGA
	S	TGCAGATA	*S*	TGCACAGT
	T	TGTATCGT	*T*	TGCGATCG

TableS 8.4 Filtering process of the four Illumina runs used in this study

Statistics parameter	DIATOM01 - 2013	DIATOM02 - 2014	DIATOM03 - 2014	DIATOM05 - 2015	Total
Total number of reads	1176424	1055387	951878	632315	
Reject ambiguous forward	0	0	0	0	
Reject ambiguous reverse	0	0	0	0	
Low mean quality forward	52295	41746	58242	22663	
Low mean quality reverse	117546	255053	91663	59154	
Low mean quality contig	0	0	0	0	
Low base quality contig	61508	17095	31879	11398	
Not enough matching contig	2205	152394	1792	1384	
No primers forward	55701	52065	58047	22168	
Error in primers forward	4297	3075	3352	2079	
No primers reverse	45934	35218	44174	24144	
Error in primers reverse	4288	3659	2626	1593	
Mismatch found in primers	67105	153677	8252	16855	
Insufficient sequence length (dimers)	0	23222	0	0	
Total number of good reads	765545	318183	651851	470877	2206456
Number of ISU					3079
Number of OTU 99%					663
Number of OTU without chimera					440

TableS 8.5 List of OTUs with their number of reads per sample. The Taxonomy column shows the assignment given by the taxonomic assignment method with their respective DG values.

Available on http://onlinelibrary.wiley.com/doi/10.1111/1755-0998.12668/full

TableS 8.6 Number of OTUs from HTS analysis and species from morphological analysis for each site.

	MAR	CHE	FOS	MOU	HEV	MLN	CHA	HEB	HEP	TRA	ACO	HEN	AMB
OTU	34	24	1	22	30	17	26	26	17	27	32	52	7
Morphospecies	27	16	19	29	28	16	18	36	14	15	28	35	5

	HEC	ARC	PRB	SEL	SEB	SEC	PRD	GEM	DAR	FLR	NAS	NAB	NAP
OTU	49	23	39	33	27	32	37	32	26	23	15	14	22
Morphospecies	33	14	31	24	27	30	32	18	22	14	24	21	20

	NAM	NAM2	NAB2	NAS2	NAP2	BFE	ARE	ARZ	ARV	RHA	RHT	RHC	RAV
OTU	29	29	9	15	19	16	14	6	18	42	34	39	25
Morphospecies	15	14	20	25	24	13	24	24	27	26	27	28	27

	RHJ	RAM	RCH	BOI	MOR	VTH	RUH	PTH	AIG	VRB	VAF	MSS	CHG
OTU	36	10	40	21	26	44	18	20	37	21	35	36	38
Morphospecies	23	22	23	53	71	64	37	72	59	41	59	58	23

	GOB	MAV	VEC	BRV	CRS	BSS	CRE	BRS	PRS	VXB	VXM	MUE	PVB
OTU	5	17	31	7	37	6	20	42	19	56	18	24	25
Morphospecies	17	23	21	21	27	16	27	25	23	34	23	26	18

	CCE	VXE	VXO	VXD	OUV	VXS	CRE2	MUE2	PVB2	CCE2	OUV2	VEF	BRV2
OTU	29	8	20	14	8	23	35	19	16	14	11	5	25
Morphospecies	15	34	18	19	32	28	17	18	26	20	16	17	21

	GOB2	CHG2	BRS2	VEC2	VXO2	VXS2	VXG2	VXM2	VXE2
OTU	19	11	17	15	14	28	18	17	42
Morphospecies	34	14	34	23	31	28	19	15	24

TableS 8.7 List of species found during the morphological analysis with their relative abundance per site.

Available online

TableS 8.8 DI-CH values given by morphology (DI-CH), taxonomic assignment (Di-MOLTAXASSIGN) and leave-one-out cross-validation (DI-MOLTAXFREE) for each site. Classes are separated as follow: 1-3.5 very good; 3.5-4.5 good; 4.5-5.5 average; 5.5-6.5 bad; 6.5-8 very bad.

	MAR	CHE	FOS	MOU	HEV	MLN	CHA	HEB	HEP	TRA	ACO	HEN	AMB	HEC	ARC	PRB
DI-CH	5.45	5.22	4.82	5.27	4.92	4.47	5.04	4.89	4.92	7.17	5.74	5.04	7.98	5.22	4.84	5.90
DI-MOLTAXASSIGN	5.01	4.49	NA	4.65	4.56	3.71	4.78	5.60	5.00	7.10	4.92	5.39	7.50	4.39	6.25	4.81
DI-MOLTAXFREE	4.93	4.36	4.00	4.55	5.96	4.47	4.49	4.25	4.67	5.73	4.85	4.81	5.84	4.45	5.70	4.85

	SEL	SEB	SEC	PRD	GEM	DAR	FLR	NAS	NAB	NAP	NAM	NAM2	NAB2	NAS2	NAP2	BFE
DI-CH	5.61	4.54	6.67	5.92	5.40	4.83	6.01	6.87	6.75	6.34	3.64	2.31	6.72	6.19	5.82	6.40
DI-MOLTAXASSIGN	4.86	4.49	4.88	5.23	7.50	4.49	4.94	7.98	7.80	7.94	7.72	3.86	7.29	5.61	5.48	7.69
DI-MOLTAXFREE	5.21	4.58	5.47	5.89	5.02	4.26	5.08	6.93	6.49	6.85	4.64	4.45	6.26	6.92	6.30	3.24

	ARE	ARZ	ARV	RHA	RHT	RHC	RAV	RHJ	RAM	RCH	BOI	MOR	VTH	RUH	PTH	AIG
DI-CH	4.19	3.31	4.46	3.72	3.10	3.94	3.33	3.67	2.60	4.25	4.81	4.24	4.45	6.92	5.61	6.16
DI-MOLTAXASSIGN	4.61	4.91	4.85	4.85	3.01	4.95	4.52	4.79	4.42	4.29	5.24	5.58	5.33	7.93	5.15	6.77
DI-MOLTAXFREE	4.80	4.50	4.14	4.37	4.17	4.74	4.24	4.28	5.15	3.99	5.96	4.26	4.56	6.61	4.93	5.85

	VRB	VAF	MSS	CHG	GOB	MAV	VEC	BRV	CRS	BSS	CRE	BRS	PRS	VXB	VXM	MUE
DI-CH	3.44	3.41	3.55	4.21	4.19	3.79	4.69	5.15	4.59	4.08	4.46	4.59	3.80	3.75	3.45	3.92
DI-MOLTAXASSIGN	5.81	4.95	5.20	3.84	4.50	4.14	4.96	3.60	4.81	3.00	4.50	4.40	4.49	4.33	4.73	3.76
DI-MOLTAXFREE	2.70	4.39	4.88	4.04	3.64	3.37	4.41	5.87	4.58	3.47	4.57	4.38	2.99	4.06	4.00	3.54

	PVB	CCE	VXE	VXO	VXD	OUV	VXS	CRE2	MUE2	PVB2	CCE2	OUV2	VEF	BRV2	GOB2	CHG2
DI-CH	3.71	3.89	3.73	2.59	1.61	3.89	2.67	3.80	4.62	4.87	5.39	4.80	6.19	6.55	6.63	4.64
DI-MOLTAXASSIGN	4.02	4.92	4.50	4.61	3.04	3.98	4.32	5.30	4.77	4.67	3.98	5.00	NA	7.74	7.56	4.40
DI-MOLTAXFREE	4.93	3.82	3.80	3.57	3.80	3.62	3.94	4.79	4.37	4.56	3.46	4.96	5.03	6.34	5.67	3.72

	BRS2	VEC2	VXO2	VXS2	VXG2	VXM2	VXE2
DI-CH	4.42	4.67	3.26	2.05	2.56	4.48	2.94
DI-MOLTAXASSIGN	4.60	4.94	4.98	4.72	3.17	5.00	4.33
DI-MOLTAXFREE	4.49	4.69	3.16	4.18	2.89	3.58	4.14

TableS 8.9 Comparison of DG values given by morphology (D and G) and molecular (MOL-D and MOL-G) indices for each assigned OTUs. The ID correspond to the OTU number in the HTS dataset.

IDs	Taxonomy	D	G	MOL -D	MOL -G
491	Achnanthidium minutissimum	3	0.5	2.5	0.5
529	Achnanthidium minutissimum	3	0.5	2.5	0.5
387	Achnanthidium minutissimum	3	0.5	1.5	0.5
544	Achnanthidium minutissimum	3	0.5	4.5	0.5
462	Achnanthidium minutissimum	3	0.5	5	0.5
319	Achnanthidium minutissimum	3	0.5	5.5	0.5
310	Amphora ovalis	3.5	1	3.5	2
628	Cocconeis pediculus	5.5	2	4	0.5
86	Cocconeis placentula	5	1	5	4
326	Cocconeis placentula	5	1	5	4
403	Cocconeis placentula	5	1	5	0.5
413	Cocconeis placentula	5	1	5	2
508	Cocconeis placentula	5	1	5	2
509	Cocconeis placentula	5	1	5	4
525	Cocconeis placentula	5	1	5	4
575	Cocconeis placentula	5	1	5	2
586	Cocconeis placentula	5	1	5	0.5
618	Cocconeis placentula	5	1	5	1
159	Cocconeis placentula	5	1	4.5	4
395	Cocconeis placentula	5	1	4.5	2
398	Cocconeis placentula	5	1	4.5	1
412	Cocconeis placentula	5	1	4.5	4
454	Cocconeis placentula	5	1	4.5	2
524	Cocconeis placentula	5	1	4.5	2
230	Cocconeis placentula	5	1	5.5	4
492	Cocconeis placentula	5	1	5.5	2
265	Cocconeis placentula	5	1	4	4
280	Cocconeis placentula	5	1	4	2
262	Cocconeis placentula	5	1	3.5	8
263	Cocconeis placentula	5	1	3.5	2
453	Cocconeis placentula	5	1	3.5	8
474	Cocconeis placentula	5	1	3.5	2
584	Cocconeis placentula	5	1	3.5	4
526	Cocconeis placentula	5	1	2.5	8
527	Cocconeis placentula	5	1	2	8
146	Cyclotella meneghiniana	6	1	6.5	8
548	Cyclotella meneghiniana	6	1	6.5	2
550	Cymbella helvetica	2	1	3	2
44	Eolimna minima	7	1	6.5	2
191	Eolimna minima	7	1	6.5	8
202	Eolimna minima	7	1	6.5	2
461	Eolimna minima	7	1	6.5	0.5
605	Eolimna minima	7	1	6.5	2
439	Eolimna minima	7	1	6	4
531	Eolimna minima	7	1	6	0.5
10	Eolimna minima	7	1	4	4
249	Eolimna subminuscula	7	4	6.5	1
331	Fistulifera saprophila	7	2	4	2
223	Fragilaria crotonensis	4	1	5.5	4

411	Gomphonema micropus	3	1	3.5	4
440	Gomphonema micropus	3	1	3.5	4
574	Gomphonema micropus	3	1	3.5	2
612	Gomphonema micropus	3	1	3.5	4
585	Gomphonema micropus	3	1	4	4
214	Gyrosigma acuminatum	4	1	5	4
240	Gyrosigma acuminatum	4	1	5	2
451	Gyrosigma acuminatum	4	1	5	0.5
362	Gyrosigma acuminatum	4	1	5.5	2
150	Mayamaea spp	6	1	5.5	4
318	Mayamaea spp	6	1	7	0.5
556	Mayamaea spp	6	1	7	0.5
358	Melosira varians	4.5	2	5	2
381	Melosira varians	4.5	2	5	2
542	Melosira varians	4.5	2	5	0.5
566	Navicula cryptocephala	4	1	6	2
633	Navicula cryptocephala	4	1	8	0.5
596	Nitzschia linearis	4.5	1	7	0.5
243	Pinnularia subgibba	7.5	2	8	8
455	Pinnularia subgibba	7.5	2	8	8
516	Planothidium lanceolatum	4	1	3.5	2
580	Planothidium lanceolatum	4	1	3.5	4
391	Planothidium lanceolatum	4	1	4.5	2
229	Sellaphora seminulum	8	4	8	8
397	Sellaphora seminulum	8	4	8	2
558	Surirella angusta	4.5	1	5	0.5
394	Surirella brebissoni	4.5	2	4	2
447	Surirella minuta	4	1	4	4
61	Synedra ulna	4	1	3.5	8

CHAPTER 9
ENVIRONMENTAL DNA SURVEY OF BIOFILM EUKARYOTES: IMPLICATIONS FOR RIVERS BIOMONITORING

Project in progress

9.1. Project description

This last chapter started recently as a pilot project with the aim to explore the bioindicator potential of other than diatoms groups of protists. We wanted to take the opportunity of having DNA extracted from all the biofilm samples to check whether such groups as foraminifera and ciliates can be used as indicators of water quality in rivers. The results presented here are preliminary and additional experiments and analysis will be conducted to complete this project.

9.2. Abstract

Environmental DNA metabarcoding has proved to be a powerful tool to describe the diversity of microbial and meiofaunal communities in aquatic ecosystems. The eDNA survey of some eukaryotic groups, e.g. diatoms, has been used as complementary approach for biomonitoring of watercourses (chapters 7 and 8). However, the bioindicative value of other eukaryotes present in eDNA samples remains largely unknown. Here, we compare the potential for indication of ecological status of watercourses by different groups of protists (diatoms, ciliates, foraminifera) and metazoans. We analysed 78 epilitic biofilm samples from the Geneva basin, using as reference the ecological status established by Swiss Diatom Index DI-CH. In addition to the diatoms V4 data, the eDNA datasets were obtained with three different 18S rRNA gene markers: ciliates specific V4, universal eukaryotes V9, and foraminiferal specific 37f/41f region. Our results show that the best correlation was obtained either with diatoms or with diatoms and other algae (Chrysophyceae and Florideophyceae). Ciliates were much less informative about ecological status and foraminifera were not informative at all. Interestingly, some metazoans were found to be good bioindicators, in particular flatworms present only in watercourses of very good and good water quality.

9.3. Introduction

Rivers play an important function in our environment. Therefore it is crucial to evaluate with precision their ecological status. Several tools are available to achieve this purpose and studying the composition of protist communities is one of them. Using protists as bioindicators clearly presents advantages as stated by Payne (2013) who mentioned several key points that make protists easy to use and define them as efficient indicators. Four groups of protists are widely used as bioindicators: diatoms, ciliates, foraminifera and testate amoebae (reviewed in Pawlowski *et al.* 2016b).

In our work we have concentrated on three of these four bioindicator groups, excluding the testate amoebae, which are uncommon in biofilm samples. Traditional approaches focus on the morphological identification of specimens but the use of molecular data is becoming more widespread. Within the last decade, metabarcoding

surveys based on high-throughput amplicon sequencing technologies have unveiled huge protist diversity previously unknown and also opened the possibility to screen a large number of habitats within a short time.

Diatoms are the most important protists in the field of water quality bioindication. Several indices based on their morphology have been established, among them the Biological Diatom Index in France (Lenoir & Coste 1996; Coste *et al.* 2009), the Trophic Diatom Index in UK (Kelly *et al.* 2001), or the Swiss Diatom Index (DI-CH) in Switzerland (Hürlimann & Niederhauser 2007). Several metabarcoding studies highlight the potential of diatom molecular data to assess water quality (Kermarrec *et al.* 2013, 2014; Zimmermann *et al.* 2014; Visco *et al.* 2015 - Chapter 7). Yet despite the great efforts made by the authors of these studies, a lot of diatom species used as bioindicators are not yet represented in the DNA barcode database, impeding the use of total dataset of eDNA sequences. Recently, Apothéloz-Perret-Gentil *et al.* (2017 - Chapter 8) proposed a taxonomy-free approach to calculate the index based on molecular data that bypass the step of species identification.

Ciliates are common in most freshwater environments and can be very abundant in organically enriched waterbodies. They are commonly used as bioindicators in waste water treatment plants (Nicolau *et al.* 2001). Their importance as water quality indicators in rivers and lakes has also been highlighted (Foissner & Berger 1996; Sola *et al.* 1996; Berger & Foissner 2003). In Germany, an index based on fungi, ciliate and other protist species (DIN 38 410, Berger *et al.* 1997) has been established to monitor the water quality.

Foraminifera are widely used as bioindicators in marine environment (Schönfeld *et al.* 2012). Their tests preserved in the sediments have been used to assess the impact of pollution due to oil drilling and spills (e.g. Jorissen *et al.* 2009; Denoyelle *et al.* 2010; Schwing *et al.* 2015), heavy metals (e.g. Cadre *et al.* 2003; Bergin *et al.* 2006; Frontalini *et al.* 2009), and industrial aquaculture (Vidovic *et al.* 2009, 2014). A partial region of the 18S rDNA fragment has been established as DNA barcode for this group (Pawlowski 2000; Pawlowski & Holzmann 2014) and an up-to-date public database is available at http://forambarcoding.unige.ch. In the field of bioindication, the metabarcoding of foraminifera had been used to measure the impact of salmon farming in marine sediment (Pawlowski *et al.* 2014a, 2016a; Pochon *et al.* 2015). Those studies show that the foraminiferal communities respond strongly to fish farms

impact, confirming their potential as bioindicators in marine habitats. Until now, foraminifera have not yet been used as bioindicators in freshwater environments although they have been found in almost all aquatic and terrestrial settings (Lejzerowicz *et al.* 2010; Chapter 4).

Here, we analysed the community of diatoms, ciliates and foraminifera, as well as other protists and metazoans present in 78 biofilm samples from different rivers of the Geneva basin. A fragment of 18S rDNA of the three bioindicator groups was amplified by using specific primers for each group. We also amplified the V9 region of 18S rDNA by using universal eukaryotic primers. We investigated the diversity obtained by high-throughput sequencing (HTS) and the potential of the different groups for bioindication, using the taxonomy-free approach developed in Apothéloz-Perret-Gentil *et al*. (2017 - Chapter 8).

9.4. Materials and methods

9.4.1. Sampling

The study was based on 78 samples collected and analysed in our previous study described in chapter 8 (Apothéloz-Perret-Gentil *et al*. 2017 – Chapter 8). Locations, sampling dates, geographic references and a map are available in the supplementary data (TableS 9.1, FigureS 9.1). The ecological status of each sample was determined using Swiss Diatom Index (DI-CH) (Swiss Federal Council 1998; Visco *et al.* 2015 - Chapter 7; Apothéloz-Perret-Gentil *et al.* 2017 - Chapter 8).

9.4.2. PCR and high-throughput sequencing

In addition to the diatom V4 region, already amplified and sequenced in a previous study (Apothéloz-Perret-Gentil *et al.* 2017 - Chapter 8), the samples were amplified for 3 other markers: the V9 region of 18S rRNA gene was amplified using universal eukaryotic primers, the V4 region was amplified using ciliates specific primers and the two hypervariable regions in 5' part of the 18S rRNA gene were amplified using foraminifera specific primers. In each case, a unique combination of tags was used for each sample in order to multiplex them for an Illumina library. Individual tags are composed of 8 nucleotides attached at each primer 5'- extremities. The PCR conditions for each marker are summarized in the TableS 9.2.

Duplicates of PCR were performed on three different extractions per sample. In total 6 PCR replicates were pooled for one sample. PCR products were purified using Sephadex G-50 superfine resin (GE Healthcare) and quantified using QuBit HS dsDNA (Invitrogen). The same amount of each sample was pooled and a final purification step was performed with High Pure PCR Product Purification kit (Roche Applied Science). Libraries were prepared with Illumina TruSeq® PCR free Preparation Kit and quantified with qPCR using KAPA Library Quantification Kit. The libraries were sequenced on a MiSeq instrument using paired-end sequencing for 500 cycles with a standard kit v2.

9.4.3. HTS data analysis

Operational Taxonomic Units (OTUs) were obtained following the method described in Pawlowski et al. 2014. Just after quality filtering and assembly steps, the librairies from the same marker were combined. Then, de-replication was performed in order to obtain Individual Sequence Units (ISUs). An abundance threshold of 10 was used for the minimum number of reads required for each ISU (Bokulich *et al.* 2013). They were then grouped at 98% using complete-linkage clustering method. Finally, chimeric sequences were removed after manual inspection of Uchime (Edgar *et al.* 2011) candidates. Diatoms sequences obtained in the Chapter 8 were clustered at 98%. OTUs were assigned to the first BLAST hit using nBLAST (Altschul *et al.* 1990) with a similarity threshold of 95%. The BLAST was performed against NCBI database and a local database of each taxonomic group. All computer analyses were performed using R (R Core Team 2013). The non-metric multidimensional scaling (NMDS) plot was performed with vegan package (Oksanen *et al.* 2013).

9.4.4. Calculation of the molecular index

The molecular index was inferred from HTS data following the same workflow as in Chapter 8. Two ecological values based on the distribution of the relative frequency of each OTU across the samples were calculated; one for the optimal condition of living and the other for the tolerance rate. The two values correspond to the autoecological value D and weighting parameter G used in DI-CH (Hürlimann & Niederhauser 2007). To calculate the index, the weighted average equation of Zelinka and Marvan (1961) was used. A 25/75 cross-validation test and a single permutation cross-validation test were applied to each dataset. In the first case, the ecological values were calculated for 75% of the samples and the evaluation was

performed on the 25% remaining samples. The sites were randomly chosen and calculations were repeated 100 times. For each calculation, the difference with the reference was calculated. In the single-permutation cross-validation test, we calculated the ecological values on the entire dataset except one site that was evaluated.

9.5. Results

9.5.1. HTS data

We analysed four datasets corresponding to diatoms V4 region, ciliates V4 region, foraminiferal 37f/41f region and eukaryotes V9 region of 18S rRNA gene. For diatoms we used data obtained in Chapter 8. For other regions, the new datasets were obtained. The V4 region of ciliates succeeded to amplify in all samples, except one, while foraminifera marker only amplified in 54% of the samples (FigureS 9.2). For the V9 marker, 16 samples were excluded because negative PCR controls turned out to be positive. In total, five Illumina libraries were prepared and sequenced. A summary of the numbers of reads and the filtering processes are shown in TableS 9.3. Finally, data were clustered into 265, 408, 789, and 1987 OTUs for the diatoms, ciliates, foraminifera, and eukaryotes, respectively.

For diatoms (Figure 9.1A), about a quarter of the reads are represented by Achnanthales (e.g. *Achnanthidium*, *Planothidium*, *Cocconeis* genera) and another quarter by Naviculales (mostly *Navicula* and *Eolimna* species). Bacillariales (almost only *Nitzschia* species), Cymbellales (*Gomphonema* and *Encyonema* species) and Thalassiophysales (*Amphora* species) represent about 10% of the diatom diversity each. Eunotiales (*Eunotia* species), Fragilariales (*Diatoma*, *Fragilaria* and *Asterionella* species), Melosirales (*Melosira varians*), Surirellales (*Surirella* and *Cymatopleura* species) and Thalassiosirales (*Cyclotella* and *Stephanodiscus* species) represent together about 10% of the total dataset. Finally, about 13% of the reads were assigned to environmental sequences of diatoms, which are not obiviously assigned to any order of diatoms. The percentages of reads and OTUs for each group are similar in the diatom dataset.

For ciliates (Figure 9.1B), half of the reads is represented by Oligohymenophorea (almost 40% of *Vorticella* species) following by Spirotrichea (almost only *Stichotrichia* species) with 15% of the dataset. Litostomatea (more than 80% of *Paraspathidium* species), Phyllopharyngea (*Heliophrya*, *Chilodonella*, *Trithigmostoma*, *Chilodonella*, *Trithigmostoma*, *Chlamydodontida* genera) and Prostomatea (*Placus*, *Urotricha*, *Plagiocampa* and unknown genera) represent about 5% of the ciliate diversity each. Armophorea (only one OTU assigned to *Metopus* species), Colpodea (90% of *Platyophrya* species), Heterotrichea (*Stentor* species) and Nassophorea (almost only *Nassula* species) represent together about 5% of the reads and 10% of the OTUs In this case also the percentages of reads and OTU were similar for all the highly represented groups, except for the unclassified ciliate that represent 15% of the reads but only 5% of the OTUs.

For foraminifera (Figure 9.1C), the most abundant clade is the FW 3 comprising almost 60% of the reads following by the FW 4 clade with 30% of the dataset. The clade FW 1 represents 8% of the reads. Interestingly, those three clades were equally represented in term of diversity (30% of the OTUs each). The clades FW 2 and FW5 are less represented with about 2% and 0.5% of the total number of reads respectively. The clade FW2 represents about 8% of the OTUs against only 1.5% for the FW 5 clade. The marine clades E and M are represented with very few reads and OTUs (0.02% and 0.01% of the reads and 1.3% and 0.25% of the OTUs respectively).

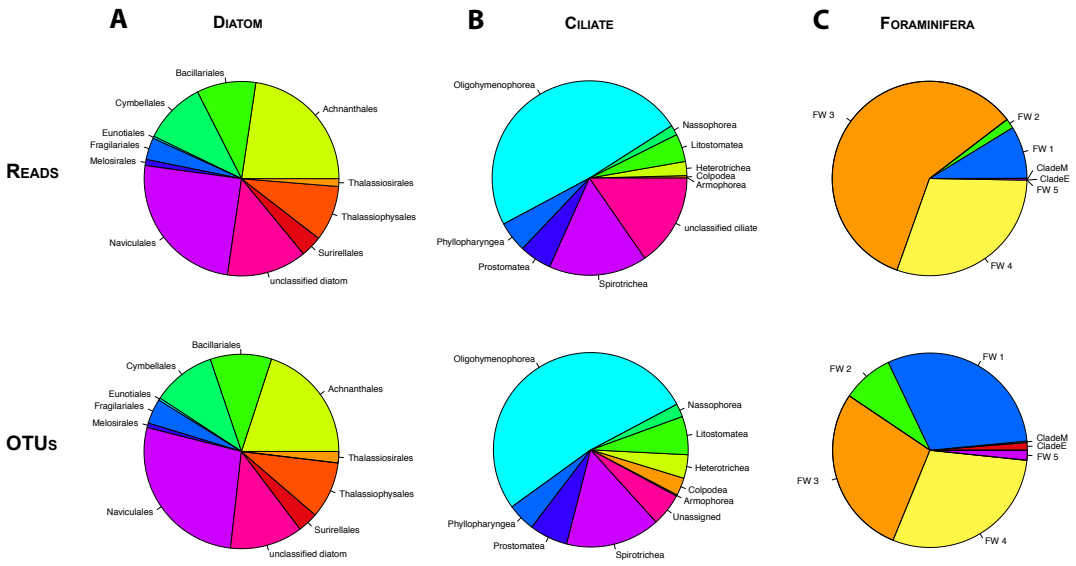


Figure 9.1 Repartition of assigned reads and OTUs of diatoms (A), ciliate (B) and foraminifera (C) (taxonomic rank: Order, Class and Clade for the three groups respectively) for the entire dataset.

Compared to the three taxon-oriented datasets, the analysis of the V9 marker presents an overall view of eukaryotic diversity in biofilm samples. The assemblage is dominated by diatoms (38% of the reads and 15% of the OTUs) and other algae (Chlorophyta, Chrysophyceae and Florideophyceae). Together the algae (including diatoms) represent more than 62% of the reads and 27% of the OTUs. The second most abundant group are metazoans to which belong almost 25% of the reads. They are the most diverse group with fungi (17% and 16% of the OTUs respectively). Their assemblage is dominated by Gastropods (47%), followed by Platyhelminthes (18%), Arthropods (14%), Cnidarians (9%) and Rotifera. Ciliates and Foraminifera are represented by relatively few reads. Ciliates make about 1.62% of the total number, while foraminifera comprise only 56 reads (about 0.001%). (Figure 9.2, TableS 9.4).

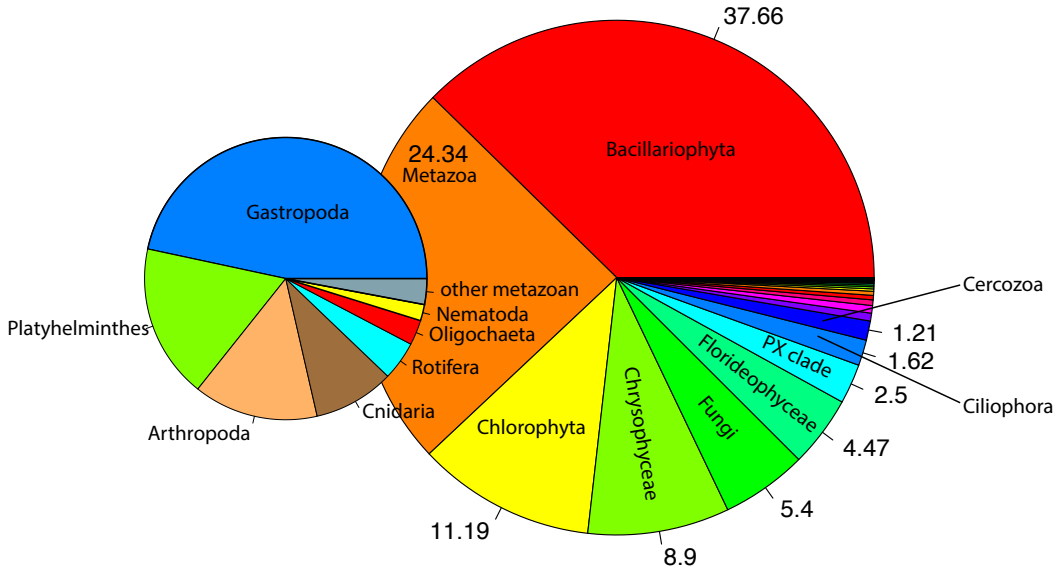


Figure 9.2 Proportion of the reads of major taxonomic groups in the V9 dataset.

9.5.2. Bioindication

We perform community analysis of each group with NMDS to explore their potential as bioindicators. 95% confidence ellipses based on the standard error of the mean community were drawn for each water quality class (Figure 9.3). As expected, diatoms community inferred from V4 data shows a gradient for different water qualities and appear to be well separated (Figure 9.3A). For ciliates, one site of very good water quality (VXO2) was very different in terms of community structure and therefore not considered for the analysis. The analysis of ciliates community allows differentiating very good quality sites, while the other sites were lumped together (Figure 9.3B). In the case of foraminifera, no community pattern was visible at all (Figure 9.3C).

Analysis of eukaryotic community using the V9 marker gives very interesting results. The analysis of entire dataset shows emerging two community groups, one corresponding to water quality defined as very good to average (blue, green and yellow) and the other corresponding to poor or very poor water quality (orange and red) (Figure 9.3D). As algae represent the majority of the communities in the V9 marker, we analysed them separately from other groups, including or not the

diatoms. When diatoms are included, we observe a quite well separated gradient for the water quality classes (Figure 9.3E). The gradient is less evident when diatoms are excluded from analyses (Figure 9.3F).

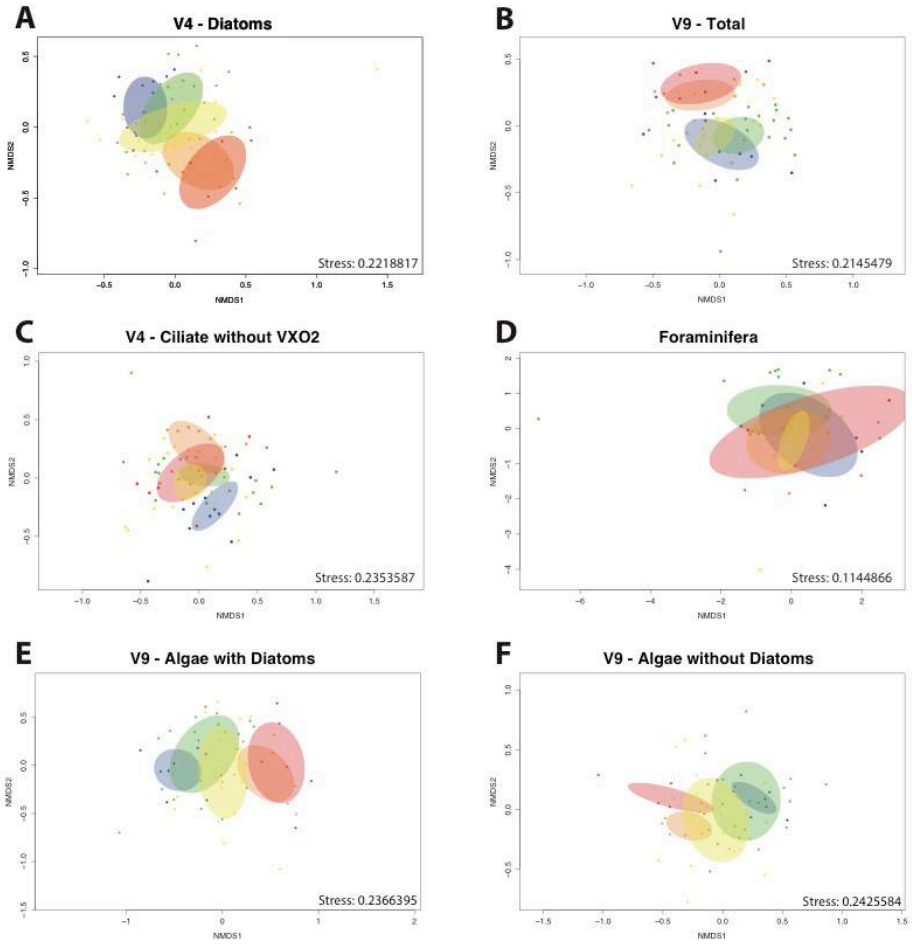


Figure 9.3 Non-metric multidimensional scaling of the different communities. The stress value is indicated for each marker. Sites are coloured in function of their DI-CH value (blue-very good, green-good, yellow-average, orange-poor, red-very poor). For each quality group, a 95% confidence ellipse of the standard error of the mean is drawn.

We used “taxonomy-free” approach to evaluate the correlation between the values of DI-CH and molecular indices inferred for each group or marker. This has been done by providing each OTU with autoecological value and weighting parameter as described in Chapter 8. Molecular indices were calculated for each sample with the single-permutation cross-validation test and the correlation with the DI-CH value is indicated in Figure 9.4. The V4 marker of diatoms and the V9 marker of algae without

diatoms gave the best correlation (R2 = 0.63 for both), followed by the V9 maker of algae with diatoms with a R2 = 0.61. Much weaker correlation was obtained with V4 of ciliates (R2 = 0.35 and p-value = 0.0023), V9 of total eukaryotes (R2 = 0.36, p-value = 0.0041) and the 37f/41f marker of foraminifera (R2 = 0.41, p-value = 0.0077).

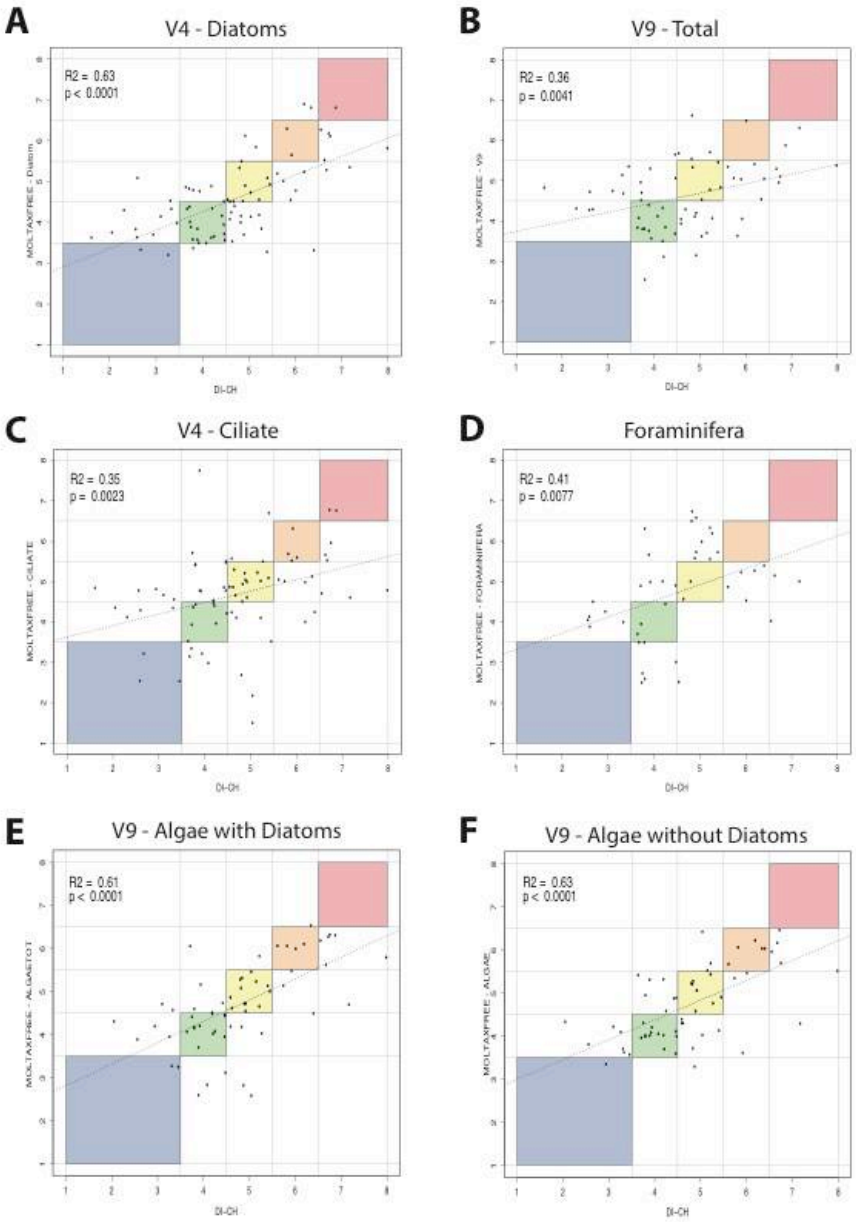


Figure 9.4 Relationship between the calculated molecular index and the DI-CH value for each marker. Coloured boxes represent the ecological status given by the DI-CH (blue: very good, green: good, yellow: average, orange: poor, red: very poor). The regression line for all samples is represented by dashed line and the R2 and p-value are indicated for each graph.

To complement this correlation, a 25/75 cross-validation test was performed and for each site the difference with the reference DI-CH value was calculated. The frequencies of all differences were plotted into a single graph (Figure 9.5). An optimal plot consists of a normal distribution with a small standard deviation, e.g. a bell as tight as possible. The analyses of V4 diatoms, V4 ciliates and V9 algae with and without diatoms gave comparable results, with a slightly better distribution for the two markers containing diatoms. The total V9 marker showed a bigger standard deviation than the diatom and ciliate marker. The foraminiferal marker did not show any bell-shaped signal.

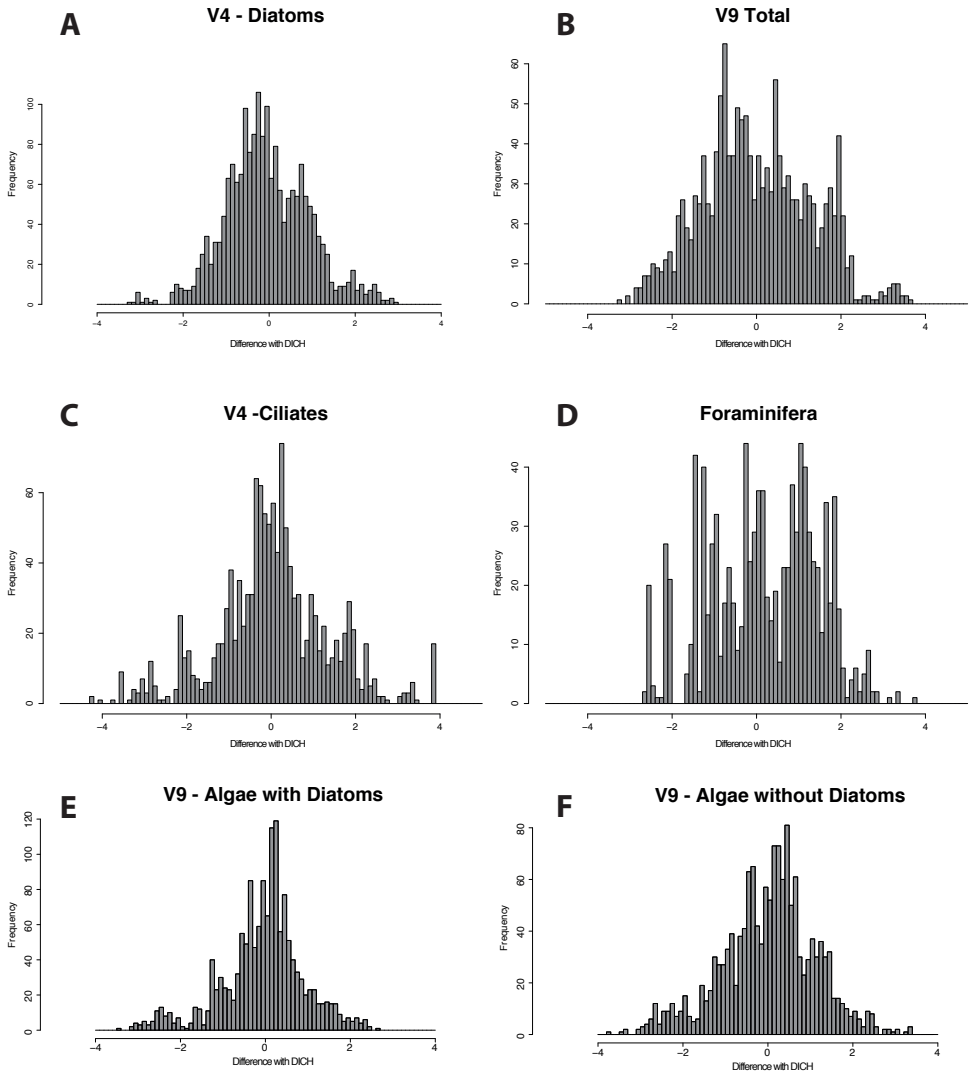


Figure 9.5 25/75 cross-validation test on the four selected markers with 100 replicates. The frequency of difference with the DI-CH value is plotted.

In addition to evaluating the bioindicator potential of protist taxa, we also examine the metazoan data obtained through analysis of V9 marker in search for indicator species. The Figure 9.6 shows the relative frequency of each metazoan OTUs across the entire dataset. The sites are sorted in function of their ecological status, therefore sequences found only in one or two classes may be good indicator species. We indicate some potential candidates for bioindication with an arrow at the left side of the figure.

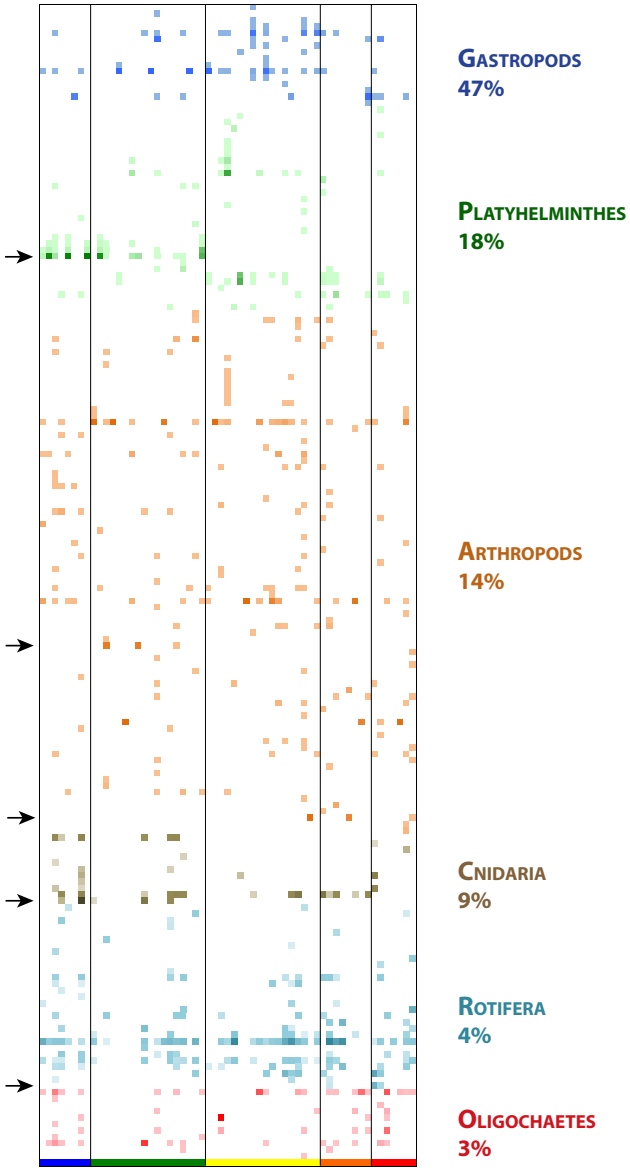


Figure 9.6 Heat map of the relative abundance of 6 most represented groups of metazoan present in the V9 dataset. Each row corresponds to an OTU and each column to a sample. Samples are sorted in function of their DI-CH value. The arrows indicate potential indicator species.

9.6. Discussion

The aim of this project was to explore the diversity of different groups of eukaryotes present in the epilithon of rivers and streams of Geneva basin. Our preliminary results provide interesting insights into the potential of some of these groups to be used as bioindicators of water quality.

The analysis of ciliates V4 data shows that their epilithic community are consistent with other studies on freshwater streams (Dopheide *et al.* 2008; Boscaro *et al.* 2016). All the ciliates classes present in Boscaro *et al*. (2016) were found in our dataset, except two classes present at very low abundance in the latter study. Ciliates have already been used as indicators of organic enrichment in freshwater streams (Madoni & Bassanini 1999; Madoni & Braghiroli 2007; Madoni *et al.* 2008), but all these studies analysed their diversity in sediment samples. Much less is known about ciliates diversity in biofilm samples, although they are considered as one of the most important consumers of biofilm products (Dopheide *et al.* 2008) and their communities are sensitive to the impact of human activity in streams (Dopheide *et al.* 2009).

In view of our results, the response of biofilm ciliates to the environmental pollution is much less significant than in the case of diatoms. The correlation between molecular indices inferred from ciliates data and the DI-CH used as reference is not very good even if the 25/75 cross-validation show a promising bell-shaped graph. This could be explained by the fact that the community of ciliates in the biofilm may not be the same as in sediment and that their bioindicative signal is affected by trophic factors rather than chemical parameters. Madoni & Braghiroli (2007) observed that algivorous ciliates were dominant in low polluted sites while bacterivorous ciliates dominated in polluted ones. Another explanation may be directly based on the calibration of the index. Indeed in our case the ciliates index has been calibrated on the morphological diatom index. Perhaps the ciliates sensitivity to chemical parameters is not exactly the same as diatoms and it might be necessary to do chemical analysis to assess the efficiency of ciliates as bioindicators in streams.

The second analysed group, the foraminifera perform even less well than ciliates. The main reason is that they occurred relatively seldom in biofilm samples. One of the conditions to be a suitable group for bioindication is that the species has to be

well represented in every sample across the year and easy to analyse (Arndt *et al.* 1987). The fact that foraminifera are not found and sequenced in all samples hugely limits their use in routine biomonitoring. Moreover, the diversity of foraminifera in freshwater habitats compared to marine ecosystem is relatively low. It is therefore not surprising that our results concerning this group are not very conclusive. Nevertheless, one would need to examine a larger sampling dataset, including also sediment samples before declaring that freshwater foraminifera are not as good bioindicators as their marine relatives.

Compared to ciliates and foraminifera, the protists that perform as well as diatoms are the other groups of algae, in particular the green algae (Chlorophyta), the golden algae (Chrysophyceae) and the red algae of the superfamily Florideophyceae. The potential of these different algal groups as bioindicators is not surprising and has already been suggested before by some authors (Tolotti *et al.* 2003; Bellinger & Sigee 2015). Several attempts to create an index based on periphyton have already been conducted (Hill *et al.* 2000; Schaumburg *et al.* 2004; Schneider & Lindstrøm 2011). However the routine use of algae as bioindicators was mainly limited because of the difficulties in their morphological identification (Schneider & Lindstrøm 2011; Bellinger & Sigee 2015). The metabarcoding approach, which allows distinguishing the species based on their DNA sequences, widely opens the door to the use of multi-taxon-based phytobenthos as bioindicators. To set up such multi-taxon approach, it would be interesting to compare different markers, in particular the V4 region of 18S rRNA gene and the chloroplastic gene *rbc*L, commonly used for diatoms, and possibly easily adapted to work with other algal groups.

An interesting finding of this study was obtained by analysis of metazoan V9 sequences. These sequences counted for about a quarter of the total number of reads and comprised all main groups of freshwater metazoans, including gastropods, platyhelminths and arthropods. Such high diversity of so many metazoan phyla in small volume of biofilm samples can be explained only by the presence of extracellular DNA (reviewed in Barnes & Turner 2016; Goldberg *et al.* 2016). The extracellular DNA is considered as an important factor in the formation and stabilisation of bacterial biofilm structure (Whitchurch *et al.* 2002; Okshevsky & Meyer 2015). Knowing that the free DNA molecules can be preserved in water and transported for kilometres (Deiner & Altermatt 2014; Deiner *et al.* 2016), we can deduce that the biofilm acts as a filter that traps DNA from the water.

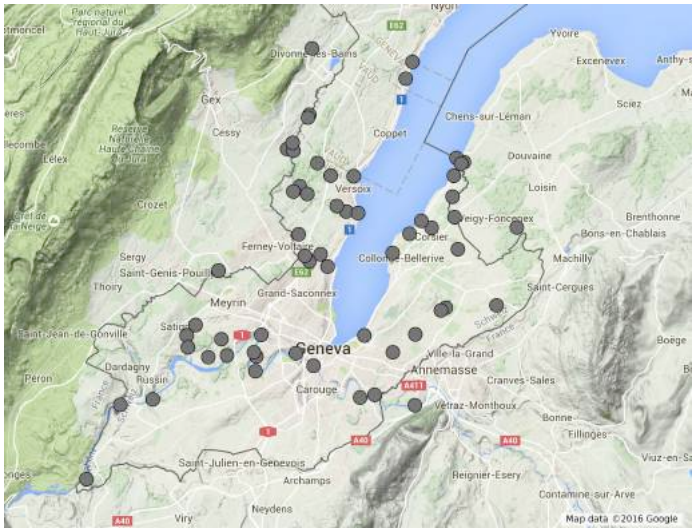
Mapping the relative frequency of metazoan OTUs in function of the ecological status of sample reveals some potential candidates for bioindicator species. The case of one platyhelminthes OTU, indicated with an arrow in the Figure 9.3, is particularly interesting. This OTU was identified as *Polycelis felina* and was detected only in sites of very good or good water quality. This species has already been described as sensible to different pollutants and considered as a good bioindicator of streams (Stubbington *et al.* 2011; Manenti & Bianchi 2014).

This example highlights the power of extracellular DNA approach and opens perspectives for its further use for bioindication. Indeed, even a small biofilm sample contains enough DNA from metazoan species to obtain information about the quality of the river. The ecology of metazoan has been studied more closely than that of most of unicellular groups. Thus exploring eDNA could be an interesting source of information about the diversity of metazoans living in a given river, complementary to traditional surveys. It might help finding new potential indicative species or developing an index directly based on metazoan extracellular DNA data.

To conclude, it is important to notice that the multitaxon approach could provide an interesting alternative to the biomonitoring based on single-taxon bioindicators. For example, the biotic index based on the whole phytobenthos community, rather than diatoms only, could be much more sensitive and useful for environmental impact assessment. At the same time, a detailed analysis of DNA sequences present in biofilm samples could be an invaluable source of information about the presence/absence of some indicator species.

9.7. Supplementary data

FigureS 9.1 Map of the sampling site



FigureS 9.2 For each sample, DICH values and presence (+) or absence (-) of HTS data per markers (Diatom, Ciliate, Foram and V9). All samples amplified with the V9 marker, however those for which the PCR negative control showed amplification were discarded. DI-CH values are coloured as follow: Blue (1-3.5) very good; Green (3.5-4.5) good; Yellow (4.5-5.5) average; Orange (5.5-6.5) poor; Red (6.5-8) very poor

Station	DICH	Diatom	Ciliate	Foram	V9
1 MAR	5.45	+	+	-	+
2 CHE	5.22	+	+	+	+
3 FOS	4.82	+	+	+	+
4 MOU	5.27	+	+	+	+
5 HEV	4.92	+	+	+	+
6 MLN	4.47	+	+	+	+
7 CHA	5.04	+	+	+	+
8 HEB	4.89	+	+	+	+
9 HEP	4.92	+	+	+	+
10 TRA	7.17	+	+	+	+
11 ACO	5.74	+	+	-	+
12 HEN	5.04	+	-	-	+
13 AMB	7.98	+	+	-	+
14 HEC	5.22	+	+	+	+
15 ARC	4.84	+	+	-	+
16 PRB	5.90	+	+	+	-
17 SEL	5.61	+	+	+	+
18 SEB	4.54	+	+	+	-
19 SEC	6.67	+	+	+	+
20 PRD	5.92	+	+	-	+
21 GEM	5.40	+	+	-	+
22 DAR	4.83	+	+	+	+
23 FLR	6.01	+	+	+	+
24 NAS	6.87	+	+	-	+
25 NAB	6.75	+	+	-	+
26 NAP	6.34	+	+	-	+
27 NAM	3.64	+	+	+	+
28 NAM2	2.31	+	+	-	-
29 NAB2	6.72	+	+	-	+
30 NAS2	6.19	+	+	-	+
31 NAP2	5.82	+	+	-	+
32 BFE	6.40	+	+	+	+
33 ARE	4.19	+	+	+	+
34 ARZ	3.31	+	+	-	+
35 ARV	4.46	+	+	-	+
36 RHA	3.72	+	+	+	-
37 RHT	3.10	+	+	-	+
38 RHC	3.94	+	+	-	+
39 RAV	3.33	+	+	+	+

Station	DICH	Diatom	Ciliate	Foram	V9
40 RHJ	3.67	+	+	+	-
41 RAM	2.60	+	+	+	-
42 RCH	4.25	+	+	+	+
43 CHG	4.21	+	+	-	+
44 GOB	4.19	+	+	-	+
45 MAV	3.79	+	+	+	+
46 VEC	4.69	+	+	-	-
47 BRV	5.15	+	+	-	+
48 CRS	4.59	+	+	-	+
49 BSS	4.08	+	+	-	+
50 CRE	4.46	+	+	-	+
51 BRS	4.59	+	+	-	+
52 PRS	3.80	+	+	+	+
53 VXB	3.75	+	+	+	+
54 VXM	3.45	+	+	-	+
55 MUE	3.92	+	+	+	+
56 PVB	3.71	+	+	+	+
57 CCE	3.89	+	+	+	+
58 VXE	3.73	+	+	+	-
59 VXS	2.59	+	+	+	-
60 VXO	1.61	+	+	-	-
61 VXD	3.89	+	+	-	+
62 OUV	2.67	+	+	+	-
63 CRE2	3.80	+	+	+	+
64 MUE2	4.62	+	+	-	+
65 PVB2	4.87	+	+	-	+
66 CCE2	5.39	+	+	+	-
67 OUV2	4.80	+	+	+	+
68 VEF	6.19	+	+	+	-
69 BRV2	6.55	+	+	+	+
70 GOB2	6.63	+	+	+	-
71 CHG2	4.64	+	+	+	-
72 BRS2	4.42	+	+	-	+
73 VEC2	4.67	+	+	-	-
74 VXO2	3.26	+	+	-	+
75 VXS2	2.05	+	+	-	+
76 VXG2	2.56	+	+	+	+
77 VXM2	4.48	+	+	+	+
78 VXE2	2.94	+	+	+	+

TableS 9.1 Code, location, sampling date and geographic references for all the sites used in this study

Code	Location	Date	N	E
MAR	Marnot - embouchure	23.09.13	46.2917829744536	6.2525148226365
CHE	Cherre - amont chemin Armand-Dufaux	10.09.13	46.2641266680935	6.2084973356855
FOS	Fossaz - amont chemin du Milieu	10.09.13	46.2542189611206	6.1955359944258
MOU	Moulin - aval route d'Hermance	09.09.13	46.2943644324306	6.2420058781010
HEV	Hermance - les Verrières	10.09.13	46.2674503269817	6.2896224591666
MLN	Moulanais - amont chemin de la Montagne	10.09.13	46.2023080494067	6.1957335256471
CHA	Chamburaz - embouchure	10.09.13	46.3015570769916	6.2494452908081
HEB	Hermance - embouchure	23.09.13	46.3037483305912	6.2439452136647
HEP	Hermance - Pont de Bouringe	23.09.13	46.3006814993215	6.2474526366922
TRA	Traînant - Traînant	23.09.13	46.2111335005438	6.1743805705553
ACO	Aisy - Côte d'or	23.09.13	46.2672206526359	6.2251655579588
HEN	Hermance - pont Neuf	23.09.13	46.2728733256966	6.2424192439071
AMB	Aisy - embouchure	23.09.13	46.2709413692631	6.2174223159857
HEC	Hermance - Pont de Crévy	23.09.13	46.2834739573656	6.2409531343499
ARC	Aisy - route de Covéry	23.09.13	46.2560324193688	6.2450595739550
PRB	Paradis - embouchure	24.09.13	46.2256169344669	6.2360049188309
SEL	Seymaz - pont Ladame	24.09.13	46.2116235120562	6.2128208658723
SEB	Seymaz - embouchure	24.09.13	46.1800909149041	6.1820310578403
SEC	Seymaz - pont de Choulex/Montagnys	24.09.13	46.2240906477902	6.2320852189619
PRD	Paradis - Les Doillets	24.09.13	46.2268271000039	6.2742197384710
GEM	Grebattes - embouchure	13.03.14	46.2001104700812	6.0925547290088
DAR	Maison-Carrée - Bois de Bay	13.03.14	46.1997615701834	6.0562836345793
FLR	Montfleury - aval jardins familiaux	13.03.14	46.2092262912882	6.0661494899460
NAS	Avril - Satigny	03.03.14	46.2110761677475	6.0399264175209
NAB	Avril - Bourdigny	11.03.14	46.2166025516207	6.0466534300789
NAP	Avril - Peney	11.03.14	46.2048790154452	6.0407982172074
NAM	Maille - La Maille	11.03.14	46.2449085007345	6.0640161359387
NAM2	Maille - La Maille	08.09.14	46.2449085007345	6.0640161359387
NAB2	Avril - Bourdigny	08.09.14	46.2166025516207	6.0466534300789
NAS2	Avril - Satigny	08.09.14	46.2110761677475	6.0399264175209
NAP2	Avril - Peney	08.09.14	46.2048790154452	6.0407982172074

BFE	Bois-des-frères - Embouchure	08.09.14	46.2114456185367	6.0962256626053
ARE	Arve - Ecole de Medecine	15.09.14	46.1952666686722	6.1359482032658
ARZ	Arve - Pont de Zone	15.09.14	46.1745776985265	6.2124621061332
ARV	Arve - Vessy	15.09.14	46.1786191243934	6.1711212891023
RHA	Rhône - amont Allondon	18.09.14	46.1777352623214	6.0142314448704
RHT	Rhône - Touvière	18.09.14	46.1748119228603	5.9895711053417
RHC	Rhône - Conflan	18.09.14	46.1359547550625	5.9639957127308
RAV	Rhône - aval STEP	25.09.14	46.2023372927514	6.0906853294790
RHJ	Rhône - amont Jonction	25.09.14	46.2017676187392	6.1224461399440
RAM	Rhône – amont STEP/Aire	25.09.14	46.1924563371588	6.0919669409397
RCH	Rhône - Chèvre	25.09.14	46.2006778445437	6.0702854653446
CHG	Chânat - amt Gobé	12.03.15	46.2534520918533	6.1410501431382
GOB	Gobé - Amt Colovrex	12.03.15	46.2525912318470	6.1293858151625
MAV	Marquet - amt Vireloup	12.03.15	46.2639383148193	6.1245396365019
VEC	Vengeron - amt CFF	12.03.15	46.2470290116341	6.1464434264592
BRV	Braille - aval bassin retention	16.03.15	46.2945206140113	6.1487539289172
CRS	Creuson - amt rte Sauvergny	16.03.15	46.3008780570326	6.1387335190395
BSS	Brassu - amt rte Suisse	16.03.15	46.3450848758620	6.2056197861548
CRE	Creuson - emb	16.03.15	46.2849087011011	6.1310059299667
BRS	Braille - amt rte Suisse	16.03.15	46.2941280195456	6.1662158289262
PRS	Pry - amt rte Suisse	16.03.15	46.3537775471141	6.2107495244715
VXB	Versoix - Bossy	17.03.15	46.2889371974045	6.1253935744691
VXM	Versoix - Mâchefer	17.03.15	46.2786432136786	6.1534085741741
MUE	Munet - emb	17.03.15	46.3261809698058	6.1325439069373
PVB	Pissevache - rte Vieille Bâtie	17.03.15	46.2862720131969	6.1205920532548
CCE	Crève-cœur - emb	17.03.15	46.2758000506382	6.1605473170487
VXE	Versoix - emb	17.03.15	46.2750013312722	6.1695187672985
VXS	Versoix - Sauvergny	19.03.15	46.3115934019758	6.1200411238364
VXO	Versoix - aval Oudar	19.03.15	46.3074151163746	6.1205320782379
VXD	Versoix – amt Divonne	19.03.15	46.3608438278437	6.1346982962355
OUV	Oudar - aval STEP	19.03.15	46.3084398618494	6.1158335900318
CRE2	Creuson - emb	21.09.15	46.2849087011011	6.1310059299667
MUE2	Munet - emb	21.09.15	46.3261809698058	6.1325439069373
PVB2	Pissevache - rte Vieille Bâtie	21.09.15	46.2862720131969	6.1205920532548
CCE2	Crève-cœur - emb	21.09.15	46.2758000506382	6.1605473170487

OUV2	Oudar - aval STEP	21.09.15	46.3084398618494	6.1158335900318
VEF	Vengeron - Fortaille	25.09.15	46.2497517992879	6.1328261755823
BRV2	Braille - aval bassin retention	25.09.15	46.2945206140113	6.1487539289172
GOB2	Gobé - Amt Colovrex	25.09.15	46.2525912318470	6.1293858151625
CHG2	Chânat - amt Gobé	25.09.15	46.2534520918533	6.1410501431382
BRS2	Braille - amt rte Suisse	25.09.15	46.2941280195456	6.1662158289262
VEC2	Vengeron - amt CFF	25.09.15	46.2470290116341	6.1464434264592
VXO2	Versoix - aval Oudar	28.09.15	46.3074151163746	6.1205320782379
VXS2	Versoix - Sauvergny	28.09.15	46.3115934019758	6.1200411238364
VXG2	Versoix - Grilly	28.09.15	46.3248199023843	6.1315378736755
VXM2	Versoix - Mâchefer	28.09.15	46.2786432136786	6.1534085741741
VXE2	Versoix - emb	28.09.15	46.2750013312722	6.1695187672985

TableS 9.2 PCR conditions and primers sequences for each marker. The grey lines of primers and PCR conditions correspond to the nested PCR

Marker	**Primer Forward**	**Primer Reverse**	**PCR conditions (Annealing / Cycles)**	**References**
V9	1380F	1510R	47° / 35x	Amaral-Zettler et al., 2009
	1389F	1510R	47° / 15x	
Ciliate	CilF	CilR	49° / 30x	Stoeck et al., 2014
	TAReuk454 FWD1	TAReuk REV3	57° / 10x 47° / 35x	
Foraminifera	14F3	17	50° / 35x	Pawlowski 2000
	14F1	17	50° / 25x	

TableS 9.3 Filtering process for the Illumina runs

Statistics parameter	Foram 1	Foram 2	Ciliate	V9 1	V9 2
Total number of reads	4329912	3935450	5858230	4151404	4193332
Reject ambiguous forward	0	0	0	0	0
Reject ambiguous reverse	0	0	0	0	0
Low mean quality forward	170385	264605	225978	125770	119183
Low mean quality reverse	241854	351466	528071	115837	96930
Low mean quality contig	0	0	0	0	0
Low base quality contig	322940	323360	189979	272678	355758
Not enough matching contig	22173	34800	72145	39138	44716
No primers forward	139069	119482	272279	703574	263493
No primers reverse	103908	85469	265128	489411	224365
Mismatch found in primers	12378	11127	77818	18162	28531
Insufficient sequence length (dimers)	7	7	0	0	0
Total number of good reads	3317198	2745134	4226242	2386833	3060356
Number of ISU	14226		4059	8295	
Number of OTU 98%	1283		608	2092	
Number of OTU without chimera	789		408	1987	

TableS 9.4 Repartition of high taxonomic groups in the entire dataset based on V9 reads

Number reads	%	Group	
1450963	37.66	Stramenopiles	Bacillariophyta
937709	24.34	Opisthokonta	Metazoa
431235	11.19	Viridiplantae	Chlorophyta
342950	8.90	Stramenopiles	Chrysophyceae
208209	5.40	Opisthokonta	Fungi
172086	4.47	Rhodophyta	Florideophyceae
96269	2.50	Stramenopiles	PX clade
62225	1.62	Alveolata	Ciliophora
46547	1.21	Rhizaria	Cercozoa
18937	0.49	Amoebozoa	Discosea
18388	0.48	Alveolata	Apicomplexa
13537	0.35	Euglenozoa	Kinetoplastida
10845	0.28	Heterolobosea	Schizopyrenida
9762	0.25	Amoebozoa	Tubulinea

5712	0.15	Opisthokonta	Choanoflagellida
5070	0.13	Stramenopiles	Oomycetes
4092	0.11	Viridiplantae	Streptophyta
3813	0.10	Stramenopiles	Eustigmatophyceae
2119	0.06	Alveolata	Dinophyceae
2070	0.05	Alveolata	environmental samples
1819	0.05	Amoebozoa	Mycetozoa
1147	0.03	Opisthokonta	Nucleariidae and Fonticula group
921	0.02	Stramenopiles	environmental samples
738	0.02	Stramenopiles	Labyrinthulomycetes
643	0.02	Heterolobosea	unclassified Heterolobosea
549	0.01	Apusozoa	Rigifilida
540	0.01	Opisthokonta	Opisthokonta incertae sedis
503	0.01	Stramenopiles	Placididea
440	0.01	Stramenopiles	Synurophyceae
298	0.01	Amoebozoa	Darbyshirella
264	0.01	Haptophyceae	Isochrysidales
253	0.01	Cryptophyta	Cryptomonadales
227	0.01	Amoebozoa	Angulamoeba
210	0.01	Apusozoa	Apusomonadidae
185	0.00	Rhizaria	environmental samples
158	0.00	Amoebozoa	environmental samples
126	0.00	Euglenozoa	Euglenida
100	0.00	Amoebozoa	Archamoebae
95	0.00	Stramenopiles	Blastocystis
89	0.00	Centroheliozoa	Acanthocystidae
82	0.00	Haptophyceae	environmental samples
57	0.00	Euglenozoa	Diplonemida
56	0.00	Rhizaria	Foraminifera
36	0.00	Cryptophyta	environmental samples
34	0.00	Amoebozoa	Gracilipodida
34	0.00	Cryptophyta	Pyrenomonadales
34	0.00	Amoebozoa	Telaepolella
26	0.00	Heterolobosea	Tulamoebidae
26	0.00	Stramenopiles	unclassified stramenopiles
21	0.00	unclassified eukaryotes	Paratrimastix
15	0.00	Amoebozoa	Ischnamoeba
14	0.00	Alveolata	Colpodellidae
14	0.00	Parabasalia	Cristamonadida
11	0.00	Rhodophyta	Bangiophyceae

CHAPTER 10

GENERAL DISCUSSION AND PERSPECTIVES

The main themes of my thesis are the diversity of foraminifera and other groups of protists and their potential application to environmental impact assessment. We investigate these issues using the tools of DNA barcoding and metabarcoding that become recently very popular thanks to the tremendous advances in high-throughput sequencing (HTS) technologies. We approach the subject from two angles. In the first part, we contribute to the development of reference database of foraminifera DNA barcodes, by describing new species and characterizing them genetically (Chapter 2-3-4). We also investigate the environmental diversity of foraminifera (Chapter 5) and analyse the potential impact of intragenomic polymorphisms on interpretation of metabarcoding data (Chapter 6). In the second part, we applied the HTS metabarcoding to biomonitoring and bioassessment of aquatic ecosystems using foraminifera and other protists (Chapter 7, 8, 9). All these studies raised many questions, both at academic and applied levels. Here, we will discuss some of these questions, at first those related to the diversity and evolutionary origin of freshwater foraminifera, then to the key technical challenges related to the application of HTS metabarcoding to biomonitoring.

10.1. Metabarcoding applied to freshwater foraminifera

Metabarcoding surveys of the Geneva basin revealed a high genetic diversity of freshwater foraminifera (Chapter 5). The 18S phylogeny showed that the freshwater phylotypes cluster into 5 clades belonging to the assemblage of monothalamous foraminifera. However, the phylogenetic position of the freshwater clades in relation to those of marine species remained uncertain. As long as these clades were composed exclusively of short environmental sequences, it was not possible to resolve their phylogenetic relationships. Thanks to the description and molecular characterisation of cultured freshwater species, presented in this thesis, it will be possible now to obtain sequences of other genes (actin, tubulin) or metatrascriptomic data that will allow inferring stronger phylogenies based on multi-genes analysis. Such phylogenies should help addressing the important questions about the origins of freshwater foraminifera and their evolutionary history.

Remarkably, until now the foraminifera are usually considered as exclusively marine group, although the foraminifera-like freshwater species have been described more than a century ago. Our study definitely shows that foraminifera are common in freshwater settings. Yet, the origin of these freshwater lineages remains enigmatic. In view of our results, it appears that foraminifera colonized freshwater habitats several time during their evolution. This is not surprising, given that many other groups of protists are represented in both marine and freshwater habitats. Usually, both environments are well separated in phylogenies, suggesting that marine to freshwater transitions were rare and ancient events (see Logares *et al.* 2009; Heger *et al.* 2010). Vermeij & Dudley (2000) explained the prevalence of ancient transitions by the fact that these ancient freshwater environments were less diverse and therefore less competitors and predators were present in the habitats, leaving empty ecological niches for colonization by marine species. Indeed, only few examples of recent marine-freshwater colonization are known: in dinoflagellates (Logares *et al.* 2007) and in diatoms (Alverson *et al.* 2007). The observations of freshwater-marine transitions are even more rare (Alverson *et al.* 2007; Shalchian-Tabrizi *et al.* 2008).

Apparently, foraminifera are not an exception to this rule. Although in our metabarcoding study we examined only a very small area of Geneva basin, sequences of foraminifera were recovered also from soil samples collected all over the world (Lejzerowicz *et al.* 2010) as well as the samples collected in Asia and Europe (Chapter 4). All these sequences branch in the same clades suggesting that freshwater foraminifera have global distribution and that the marine-freshwater transitions occur only few times in evolution of foraminifera. However, the exact number of these transition events is difficult to determine. There are four large clades that group the majority of freshwater phylotypes, but our extensive metabarcoding survey reveals the presence of some smaller independent freshwater clades, such as FW5, or the presence of freshwater phylotypes that branch within the clades of marine species. The later case is exemplified by the clade M, which comprises two environmental phylotypes found in our study as well as the soil species from Australia *Edaphoallogromia australica*. We expect that metabarcoding surveys of other geographical areas may unveil novel lineages of freshwater foraminifera, suggesting that the diversity of this poorly known group is high and that the transitions between marine and freshwater environments are more frequent than generally accepted.

10.2. Metabarcoding applied to biomonitoring

As discussed in the previous section, the investigation of foraminiferal diversity using HTS metabarcoding could be very useful to answer fundamental evolutionary and ecological issues. However, as shown in the second part of my thesis, to be successfully used for routine uses, such as water quality assessment, various challenges related to this method needs to be addressed. We will focus here on few of them, including the type of sampled material, the abundance and other issues.

10.2.1. Type of sampled material

In our studies, we sampled two different types of substrates: sediment and biofilm. The eukaryotic community inferred from biofilm samples (Chapter 9) was consistent with morphological studies (Cutler *et al.* 2015). Phototrophic organisms dominate the assemblage but heterotrophic organisms such as fungi, ciliates or cercozoans are also important players in the functioning of the biofilm ecosystem (see Battin *et al.* 2016). Bacteria, particularly cyanobacteria, are also known to be a key component of biofilm sample, however we did not investigate them in our studies. Since some of them are known to be good bioindicators (Mateo *et al.* 2015; Monteagudo & Moreno 2016; Teta *et al.* 2017), we think that further multitaxon metabarcoding studies should include this microbial component.

The other type of substrate examined in this study was the sediment sample. The sediment is a very complex environment dependent on abiotic factors such as a grain size (mud, sand), temperature or light availability (Delgado *et al.* 1991; Jesus *et al.* 2009). Here, we analysed sediment samples studying the diversity of freshwater foraminifera (Chapter 5). We did not found a significant difference between the distribution of four freshwater foraminiferal clades in the two types of samples, although the highly diverse clade FW4 was present mainly in biofilm samples.

Despite the importance of both biofilm and sediments for the functioning of aquatic ecosystem (Ancion *et al.* 2013; Gerbersdorf & Wieprecht 2015; Reid *et al.* 2016), to our knowledge, only few studies compare those two substrates. Two studies investigate the effect of both sample types in the assessment of water quality. Potapova & Charles (2005) showed that the diversity and abundance of algal assemblage were found to be different between sediment and biofilm sample at the same sampling site, although this difference did not significantly affect the

assessment of water quality. Reid *et al.* (2016) showed the potential of biofilm samples for the indication of heavy metal pollution compared to sediment by studying the community of heterotrophic organisms (bacteria and ciliates).

Although both types of substrates seem suitable for bioassessment, some technical issues should be considered for routine assessment. For example, the protocols of DNA extraction from sediments may be more complex and expensive than from biofilm samples. Moreover, PCR inhibitors are more abundant in sediment sample that can lead to severe amplification problems (Tsai & Olson 1992; Miller 2001). Those two aspects may prompt using biofilm samples when the biological signal is comparable between both substrata.

The presence of extracellular DNA could also be a challenging issue in DNA-based biomonitoring. Indeed, it is well known that free DNA molecules can be adsorbed and preserved in sediment (Mao *et al.* 2014; Turner *et al.* 2015; Torti *et al.* 2015). However, it is also known that free DNA is an integral part of the biofilm (Whitchurch *et al.* 2002; Steinberger & Holden 2005; Vilain *et al.* 2009). Therefore, the same precautions have to be applied in the case of both types of samples. On the other hand, the extracellular DNA may also be a powerful tool and not only a constraint in biomonitoring. In Chapter 9, we report the presence of some potential indicator metazoan species, which were probably only present in the biofilm sample as free DNA. It is possible that the biofilm acts as a filter that retains the DNA molecules present in the water. In this case, the analysis of biofilm DNA samples could be used to obtain a global overview of river biodiversity. However, this needs to be tested by further metabarcoding studies that compare the composition of environmental DNA isolated from water, biofilm and sediment samples.

10.2.2. Quantitative issue

One of the most important issues in biomonitoring is the estimation of species abundance. In conventional morphology-based surveys, the specimens belonging to each species are counted and the values of species absolute or relative abundance are used for index calculation. Indeed, the relative abundance of the indicator species is often a key parameter in the inference of biotic indices. However, in the HTS datasets, the number of reads does not directly correspond to number of specimens (Stoeck *et al.* 2014; Elbrecht & Leese 2015). Differences in abundance

estimation between DNA-based and morphological studies are caused by biological and technical factors.

Among biological factors, the most important is biomass variations per species that can lead to different amount of DNA in environmental samples. This factor seems particularly important in macro-invertebrates studies (Elbrecht & Leese 2015) and concern both mitochondrial and nuclear genes. In small-sized taxa, such as protists and meiofauna, the abundance issue is mainly related to the variations of gene copies number that can drastically change from one species to another (see Chapter 1.1.4). In this case, the studies using multicopy rRNA genes are particularly sensitive.

Technical biases originate essentially during the DNA extraction and the PCR amplification steps (Brooks *et al.* 2015). Different extraction protocols may lead to different amount of extracted DNA. Several studies highlight significant difference in the abundance of some bacterial (Feinstein *et al.* 2009; Henderson *et al.* 2013) or diatom (Vasselon *et al.* 2017) taxa in function of the DNA extraction methods. Storage condition prior to the extraction also seems to affect the abundance ratio between taxa (Bahl *et al.* 2012). However, only some taxa seemed to be concerned (Henderson *et al.* 2013) and, in the case of water quality assessment with diatoms, the storage conditions did not change inferred index value (Vasselon *et al.* 2017).

The final amount of sequences per species is also highly dependant on primer efficiency, which may differ from species to species (Elbrecht & Leese 2015, 2017; Piñol *et al.* 2015). However, the importance of the abundance biases depends on taxonomic group of bioindicators. Several studies indicate that the relative abundance of sequences match relatively well the relative abundance of individuals in unicellular organisms (bacteria, protists), even if they do not reflect directly the real number of living specimens (Pawlowski *et al.* 2014; Giner *et al.* 2016). This assumption was confirmed by our studies (Chapter 7, 8). There are also studies showing that the relative abundance of some macro-invertebrates, e.g. marine polychaetes (Lejzerowicz *et al.* 2015) or Chironomidae in freshwater environments (Carew *et al.* 2013) follows similar patterns in molecular and morphological data.

10.2.3. The uncertainties of taxonomic assignment

Another key issue in HTS metabarcoding is taxonomic assignation of the sequences or cluster of sequences (OTUs) to morphospecies. This issue is particularly important when metabarcoding is used to infer the biotic indices based on ecological values or categories assigned to each morphospecies (for example, AMBI, ITI for marine invertebrates, or DI-CH and other indices used for diatoms). The inference of such indices from molecular data requires a direct link between OTUs and morphospecies. However, establishing of such link might not be straightforward for several reasons.

At first, the reference DNA database of bioindicator species is far from being complete (see Chapter 1.1.4). Available databases are not exhaustive and this can be an issue, particularly when assignation to the species level is needed. As shown by many studies, including ours (Chapter 7-8), the majority of sequences cannot be assigned to species even in well-studied groups such as diatoms (Kermarrec *et al.* 2014; Zimmermann *et al.* 2015).

Secondly, most of morphospecies are genetically variable and comprise often many cryptic species. It is rare that there is only one OTU that is matching perfectly to a given morphospecies. Usually, a morphospecies is represented by a group of OTUs that are phylogenetically closely related and can be easily assigned to the same ecological category. However, frequently the clade of OTUs spans more than one morphospecies. The taxonomic assignment can be additionally impeded by short length of gene fragments used in HTS metabarcoding studies, which often is a cause of limited taxonomic resolution of analysed marker, for example in the case of 18S rRNA gene of *Navicula spp* (Chapter 7). In this situation, it can be very difficult to decide, which OTU belongs to which morphospecies. This makes the interpretation of molecular data more complicated because OTUs that cannot be identified may belong to different ecological categories.

10.2.4. Accurate assessment of diversity

The correct interpretation of metabarcoding data can also be impeded by the presence of pseudogenes or intragenomic polymorphic sequences (Brown *et al.* 2015), particularly in the case of rRNA genes (see Chapter 1.2.3).This issue may be especially critical in the ecological quality assessment based on diversity metrics (Yu *et al.* 2012; Ji *et al.* 2013; Leray & Knowlton 2015; Evans *et al.* 2016). The presence

of pseudogenes and intragenomic polymorphism may artificially inflate the species richness. Indeed, different haplotype sequences may cluster into different OTUs, particularly if the variation between them is high (Chapter 6).

Another challenging aspect of HTS metabarcoding is the presence of extracellular or free DNA in environmental samples. As previously said, those DNA can come from different sources and can be preserved for a long time in the environment, particularly in sediment (Mao *et al.* 2014; Turner *et al.* 2015; Torti *et al.* 2015). Moreover, it can be transported over large distance (Deiner & Altermatt 2014). The rate of degradation of DNA depends on numerous parameters and is therefore difficult to predict (Barnes *et al.* 2014; Pilliod *et al.* 2014; Eichmiller *et al.* 2016). This makes uncertain the interpretation of metabarcoding data, particularly in the case of recent environmental changes, such as renaturation of a watercourse or eradication of invasive species. The solution to overcome this issue could be the use of RNA, which is a more labile molecule and will therefore give a better temporal representation of the diversity. Indeed, RNA proved to infer better quality assessment than DNA compared to the traditional morphology (Pawlowski *et al.* 2014; Chapter 7), however, the significant increase in cost and time requirements is a major limitation to the use of RNA in routine assessment.

10.2.5. The need of standardization

Finally, the methods to generate and analyse HTS metabarcoding data can drastically change the interpretation of the results and therefore a major concern for the use of HTS biomonitoring is the requirement to standardize the different protocols for routine assessment. Several attempts have been made to evaluate the variations observed at each step in the process, from the sampling (Pochon *et al.* 2015; Aylagas *et al.* 2016) to extraction protocols (Vasselon *et al.* 2017) and data analysis (Mysara *et al.* 2017). A standardized protocol for using molecular data to infer marine macro-invertebrates benthic index have already been published (Aylagas & Rodríguez-Ezpeleta 2016). However, the rapid development and frequent changes of HTS technologies makes such standardisation difficult, which is one of the main concerns raised by the opponents of the HTS biomonitoring.

10.3. Perspectives

To conclude, despite these various challenges, HTS metabarcoding proved to be a powerful tool for the assessments of diversity and ecological status of environment. Molecular technics are cost and time-effective compared to the traditional ones and are less subjected to human errors such as taxonomic misidentification. Since the experienced taxonomists become more and more rare (Cotterill & Foissner 2010), the HTS metabarcoding could become a key player in the large scale monitoring of aquatic networks.

However, in my opinion, our knowledge of the organismal biology could never be replaced totally by a metabarcoding approach and it would be a pity if the molecular tools would take completely over the traditional morphological methods. The efficiency and specificity of DNA metabarcoding makes it particularly appropriate when a lot of samples need to be assessed with accuracy. For the moment, the pilot metabarcoding studies focused on the same groups of bioindicators as the traditional morphology-based surveys. The aim of these studies was to test the effectiveness of molecular methods in well-established conditions. However, it is maybe time to be more audacious and use the novel tools and technologies to explore the potential of new groups of bioindicators. Molecular tools can help to investigate the bioindicator potential of microbial eukaryote groups, for which no suitable literature on the morphological taxonomy and ecology is available (Mitchell & Meisterfeld 2005). It is important to take maximum advantage of both approaches. The challenge lies in keeping the direct connexion between applied and academic science using the advances of current and future biotechnologies for the best of environmental protection.

LITERATURE CITED

Aguiar FC, Feio MJ, Ferreira MT (2011) Choosing the best method for stream bioassessment using macrophyte communities: Indices and predictive models. *Ecological Indicators*, **11**, 379–388.

Aird D, Ross MG, Chen W-S *et al.* (2011) Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biology*, **12**, R18.

Allen AP, Gillooly JF (2006) Assessing latitudinal gradients in speciation rates and biodiversity at the global scale. *Ecology Letters*, **9**, 947–954.

Altin DZ, Habura A, Goldstein ST (2009) A New Allogromiid Foraminifer Niveus Flexilis Nov. Gen., Nov. Sp., from Coastal Georgia, Usa: Fine Structure and Gametogenesis. *The Journal of Foraminiferal Research*, **39**, 73–86.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *Journal of Molecular Biology*, **215**, 403–410.

Alve E, Korsun S, Schönfeld J *et al.* (2016) Foram-AMBI: A sensitivity index based on benthic foraminiferal faunas from North-East Atlantic and Arctic fjords, continental shelves and slopes. *Marine Micropaleontology*, **122**, 1–12.

Alve E, Lepland A, Magnusson J, Backer-Owe K (2009) Monitoring strategies for re-establishment of ecological reference conditions: possibilities and limitations. *Marine Pollution Bulletin*, **59**, 297–310.

Alverson AJ, Jansen RK, Theriot EC (2007) Bridging the Rubicon: Phylogenetic analysis reveals repeated colonizations of marine and fresh waters by thalassiosiroid diatoms. *Molecular Phylogenetics and Evolution*, **45**, 193–210.

Alverson AJ, Kolnick L (2005) Intragenomic Nucleotide Polymorphism Among Small Subunit (18s) Rdna Paralogs in the Diatom Genus Skeletonema (bacillariophyta)1. *Journal of Phycology*, **41**, 1248–1257.

Amaral-Zettler LA (2012) Eukaryotic diversity at pH extremes. *Frontiers in Microbiology*, **3**, 441.

Amato A, Kooistra WHCF, Ghiron JHL *et al.* (2007) Reproductive isolation among sympatric cryptic species in marine diatoms. *Protist*, **158**, 193–207.

Amend AS, Seifert KA, Bruns TD (2010) Quantifying microbial communities with 454 pyrosequencing: does read abundance count? *Molecular Ecology*, **19**, 5555–5565.

Ancion P-Y, Lear G, Dopheide A, Lewis GD (2013) Metal concentrations in stream biofilm and sediments and their potential to explain biofilm microbial community structure. *Environmental Pollution (Barking, Essex: 1987)*, **173**, 117–124.

Apothéloz-Perret-Gentil L, Cordonier A, Straub F *et al.* (2017) Taxonomy-free molecular diatom index for high-throughput eDNA biomonitoring. *Molecular Ecology Resources*.

Apothéloz-Perret-Gentil L, Holzmann M, Pawlowski J (2013) Arnoldiellina fluorescens gen. et sp. nov. – A new green autofluorescent foraminifer from the Gulf of Eilat (Israel). *European Journal of Protistology*, **49**, 210–216.

Arjen de Groot G, Laros I, Geisen S (2016) Molecular identification of soil eukaryotes and focused approaches targeting protist and faunal groups using high-throughput metabarcoding. *Microbial Environmental Genomics (MEG)*, 125–140.

Armitage PD, Moss D, Wright JF, Furse MT (1983) The performance of a new biological water quality score system based on macroinvertebrates over a wide range of unpolluted running-water sites. *Water Research*, **17**, 333–347.

Arndt U, Nobel W, Schweizer B (1987) *Bioindikatoren. Möglichkeiten, Grenzen und neue Erkenntnisse*. Ulmer, Stuttgart.

Auer CA (2003) Tracking genes from seed to supermarket: techniques and trends. *Trends in Plant Science*, **8**, 591–597.

Authman MM, Abbas HH (2007) Accumulation and Distribution of Copper and Zinc in Both Water and Some Vital Tissues of Two Fish Species(Tilapia zillii and Mugil cephalus) of Lake Qarun, Fayoum Province, Egypt. *Pakistan Journal of Biological Sciences*, **10**, 2106–2122.

Aylagas E, Borja Á, Irigoien X, Rodríguez-Ezpeleta N (2016) Benchmarking DNA Metabarcoding for Biodiversity-Based Monitoring and Assessment. *Frontiers in Marine Science*, **3**.

Aylagas E, Borja Á, Rodríguez-Ezpeleta N (2014) Environmental Status Assessment Using DNA Metabarcoding: Towards a Genetics Based Marine Biotic Index (gAMBI). *PLOS ONE*, **9**, e90529.

Aylagas E, Borja Á, Tangherlini M *et al.* (2017) A bacterial community-based index to assess the ecological status of estuarine and coastal environments. *Marine Pollution Bulletin*, **114**, 679–688.

Aylagas E, Rodríguez-Ezpeleta N (2016) Analysis of Illumina MiSeq Metabarcoding Data: Application to Benthic Indices for Environmental Monitoring. *Methods in Molecular Biology (Clifton, N.J.)*, **1452**, 237–249.

Bahl MI, Bergström A, Licht TR (2012) Freezing fecal samples prior to DNA extraction affects the Firmicutes to Bacteroidetes ratio determined by downstream quantitative PCR analysis. *FEMS Microbiology Letters*, **329**, 193–197.

Baird DJ, Hajibabaei M (2012) Biomonitoring 2.0: a new paradigm in ecosystem assessment made possible by next-generation DNA sequencing. *Molecular Ecology*, **21**, 2039–2044.

Barnes MA, Turner CR (2016) The ecology of environmental DNA and implications for conservation genetics. *Conservation Genetics*, **17**, 1–17.

Barnes MA, Turner CR, Jerde CL *et al.* (2014) Environmental conditions influence eDNA persistence in aquatic systems. *Environmental Science & Technology*, **48**, 1819–1827.

Bass D, Stentiford GD, Littlewood DTJ, Hartikainen H (2015) Diverse Applications of Environmental DNA Methods in Parasitology. *Trends in Parasitology*, **31**, 499–513.

Battin TJ, Besemer K, Bengtsson MM, Romani AM, Packmann AI (2016) The ecology and biogeochemistry of stream biofilms. *Nature Reviews Microbiology*, **14**, 251–263.

Bazzanti M, Mastrantuono L, Pilotto F (2017) Depth-related response of macroinvertebrates to the reversal of eutrophication in a Mediterranean lake: Implications for ecological assessment. *The Science of the Total Environment*, **579**, 456–465.

Beckers B, Op De Beeck M, Thijs S *et al.* (2016) Performance of 16s rDNA Primer Pairs in the Study of Rhizosphere and Endosphere Bacterial Microbiomes in Metabarcoding Studies. *Frontiers in Microbiology*, **7**.

Bellinger EG, Sigee DC (2015) *Freshwater Algae: Identification and Use as Bioindicators*. John Wiley & Sons.

Belore ML, Winter JG, Duthie HC (2002) Use of Diatoms and Macroinvertebrates as Bioindicators of Water Quality in Southern Ontario Rivers. *Canadian Water Resources Journal / Revue canadienne des ressources hydriques*, **27**, 457–484.

Bere T (2016) Challenges of diatom-based biological monitoring and assessment of streams in developing countries. *Environmental Science and Pollution Research*, **23**, 5477–5486.

Berger H, Foissner W (2003) Illustrated guide and ecological notes to ciliate indicator species (Protozoa, Ciliophora) in running waters, lakes, and sewage plants. In: *Handbuch Angewandte Limnologie: Grundlagen - Gewässerbelastung - Restaurierung - Aquatische Ökotoxikologie - Bewertung - Gewässerschutz*, p. . Wiley-VCH Verlag GmbH & Co. KGaA.

Berger H, Foissner W, Kohmann F (1997) Bestimmung und Ökologie der Mikrosaprobien nach DIN 38 410.

Bergin F, Kucuksezgin F, Uluturhan E *et al.* (2006) The response of benthic foraminifera and ostracoda to heavy metal pollution in Gulf of Izmir (Eastern Aegean Sea). *Estuarine, Coastal and Shelf Science*, **66**, 368–386.

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

Bernhard JM, Edgcomb VP, Visscher PT *et al.* (2013) Insights into foraminiferal influences on microfabrics of microbialites at Highborne Cay, Bahamas. *Proceedings of the National Academy of Sciences*, **110**, 9830–9834.

Beszteri B, John U, Medlin LK (2007) An assessment of cryptic genetic diversity within the Cyclotella meneghiniana species complex (Bacillariophyta) based on nuclear and plastid genes, and amplified fragment length polymorphisms. *European Journal of Phycology*, **42**, 47–60.

Bik HM, Fournier D, Sung W, Bergeron RD, Thomas WK (2013) Intra-genomic variation in the ribosomal repeats of nematodes. *PloS One*, **8**, e78230.

Bik HM, Halanych KM, Sharma J, Thomas WK (2012) Dramatic shifts in benthic microbial eukaryote communities following the Deepwater Horizon oil spill. *PloS One*, **7**, e38550.

Binh CTT, Tong T, Gaillard J-F, Gray KA, Kelly JJ (2014) Acute Effects of TiO2 Nanomaterials on the Viability and Taxonomic Composition of Aquatic Bacterial Communities Assessed via High-Throughput Screening and Next Generation Sequencing. *PLOS ONE*, **9**, e106280.

Birk S, Willby NJ, Kelly MG *et al.* (2013) Intercalibrating classifications of ecological status: Europe's quest for common management objectives for aquatic ecosystems. *Science of The Total Environment*, **454**–**455**, 490–499.

Birungi Z, Masola B, Zaranyika MF, Naigaga I, Marshall B (2007) Active biomonitoring of trace heavy metals using fish (Oreochromis niloticus) as bioindicator species. The case of Nakivubo wetland along Lake Victoria. *Physics and Chemistry of the Earth, Parts A/B/C*, **32**, 1350–1358.

Blanc H (1886) Un nouveau foraminifère de la faune profonde du Lac. *Bibliothèque Universelle*, **3**, 362–366.

Blanc H (1888) *Gromia brunnerii un nouveau foraminifère. Recueil zoologique Suisse*. H. Geog.

Blanco-Bercial L, Cornils A, Copley N, Bucklin A (2014) DNA Barcoding of Marine Copepods: Assessment of Analytical Approaches to Species Identification. *PLOS Currents Tree of Life*.

Blaxter ML (2004) The promise of a DNA taxonomy. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **359**, 669–679.

Boehler S, Strecker R, Heinrich P *et al.* (2017) Assessment of urban stream sediment pollutants entering estuaries using chemical analysis and multiple bioassays to characterise biological activities. *Science of The Total Environment*, **593**–**594**, 498–507.

Boehme P, Amendt J, Zehner R (2012) The use of COI barcodes for molecular identification of forensically important fly species in Germany. *Parasitology Research*, **110**, 2325–2332.

Bohan DA, Vacher C, Tamaddoni-Nezhad A *et al.* (2017) Next-Generation Global Biomonitoring: Large-scale, Automated Reconstruction of Ecological Networks. *Trends in Ecology & Evolution*.

Bohmann K, Evans A, Gilbert MTP *et al.* (2014) Environmental DNA for wildlife biology and biodiversity monitoring. *Trends in Ecology & Evolution*, **29**, 358–367.

Böhmer J, Arbačiauskas K, Benstead R *et al.* (2014) Water Framework Directive Intercalibration Technical Report: Central Baltic Lake Benthic invertebrate ecological assessment methods - EU Science Hub - European Commission. *EU Science Hub*.

Bokulich NA, Subramanian S, Faith JJ *et al.* (2013) Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nature Methods*, **10**, 57–59.

Borja A, Dauer DM (2008) Assessing the environmental quality status in estuarine and coastal systems: Comparing methodologies and indices. *Ecological Indicators*, **8**, 331–337.

Borja A, Franco J, Pérez V (2000) A Marine Biotic Index to Establish the Ecological Quality of Soft-Bottom Benthos Within European Estuarine and Coastal Environments. *Marine Pollution Bulletin*, **40**, 1100–1114.

Boscaro V, Rossi A, Vannini C *et al.* (2016) Strengths and Biases of High-Throughput Sequencing Data in the Characterization of Freshwater Ciliate Microbiomes. *Microbial Ecology*, 1–11.

Boulton AJ (1999) An overview of river health assessment: philosophies, practice, problems and prognosis. *Freshwater Biology*, **41**, 469–479.

Bowser SS, Habura A, Pawlowski J (2006) Molecular evolution of Foraminifera. *Genomics and evolution of microbial eukaryotes*, 78–93.

Braak CJF ter, Dame H van (1989) Inferring pH from diatoms: a comparison of old and new calibration methods. *Hydrobiologia*, **178**, 209–223.

Brandes J, Kuhajek JM, Goodwin E, Wood SA (2016) Molecular Characterisation and Co-cultivation of Bacterial Biofilm Communities Associated with the Mat-Forming Diatom Didymosphenia geminata. *Microbial Ecology*, **72**, 514–525.

Brannock PM, Halanych KM (2015) Meiofaunal community analysis by high-throughput sequencing: Comparison of extraction, quality filtering, and clustering methods. *Marine Genomics*, **23**, 67–75.

Brannock PM, Ortmann AC, Moss AG, Halanych KM (2016) Metabarcoding reveals environmental factors influencing spatio-temporal variation in pelagic micro-eukaryotes. *Molecular Ecology*, **25**, 3593–3604.

Brooks JP, Edwards DJ, Harwich MD *et al.* (2015) The truth about metagenomics: quantifying and counteracting bias in 16S rRNA studies. *BMC microbiology*, **15**, 66.

Brown EA, Chain FJJ, Crease TJ, MacIsaac HJ, Cristescu ME (2015) Divergence thresholds and divergent biodiversity estimates: can metabarcoding reliably describe zooplankton communities? *Ecology and Evolution*, **5**, 2234–2251.

Cadre VL, Debenay J-P, Lesourd M (2003) Low pH effects on Ammonia beccarii test deformation: Implications for using test deformations as a pollution indicator. *The Journal of Foraminiferal Research*, **33**, 1–9.

Capo E, Debroas D, Arnaud F, Domaizon I (2015) Is Planktonic Diversity Well Recorded in Sedimentary DNA? Toward the Reconstruction of Past Protistan Diversity. *Microbial Ecology*, **70**, 865–875.

Carew ME, Pettigrove VJ, Metzeling L, Hoffmann AA (2013) Environmental monitoring using next generation sequencing: rapid identification of macroinvertebrate bioindicator species. *Frontiers in Zoology*, **10**, 45.

Carlisle DM, Hawkins CP, Meador MR, Potapova M, Falcone J (2008) Biological assessments of Appalachian streams based on predictive models for fish, macroinvertebrate, and diatom assemblages. *Journal of the North American Benthological Society*, **27**, 22.

Cedhagen T, Gooday AJ, Pawlowski J (2009) A new genus and two new species of saccamminid foraminiferans (Protista, Rhizaria) from the deep Southern Ocean. *Zootaxa*, **2096**, 9–22.

Cedhagen T, Pawlowski J (2002) Toxisarcon Synsuicidica N. Gen., N. Sp., a Large Monothalamous Foraminiferan from the West Coast of Sweden. *Journal of Foraminiferal Research*, **32**, 351–357.

CEMAGREF (1982) Etude des méthodes biologiques d'appréciation quantitavive de la qualité des eaux. Rapport QE Lyon Bassin Rhône-Méditerranée-Corse. *AFNOR norm NF T 90-354*.

Cerveny D, Turek J, Grabic R *et al.* (2016) Young-of-the-year fish as a prospective bioindicator for aquatic environmental contamination monitoring. *Water Research*, **103**, 334–342.

Chambouvet A, Alves-de-Souza C, Cueff V *et al.* (2011) Interplay Between the Parasite Amoebophrya sp. (Alveolata) and the Cyst Formation of the Red Tide Dinoflagellate Scrippsiella trochoidea. *Protist*, **162**, 637–649.

Chand Dakal T, Giudici P, Solieri L (2016) Contrasting Patterns of rDNA Homogenization within the Zygosaccharomyces rouxii Species Complex. *PloS One*, **11**, e0160744.

Chang J, Carpenter EJ (1991) Species-specific phytoplankton growth rates via diel DNA synthesis cycles. V. Application to natural populations in Long Island Sound. *Mar. Ecol. Prog. Ser*, **78**, 115–122.

Chanudet V, Guédant P, Rode W *et al.* (2016) Evolution of the physico-chemical water quality in the Nam Theun 2 Reservoir and downstream rivers for the first 5 years after impoundment. *Hydroécologie Appliquée*, **19**, 27–61.

Chapman PM (1990) The sediment quality triad approach to determining pollution-induced degradation. *Science of The Total Environment*, **97**, 815–825.

Charif D, Lobry JR (2007) Seqin{R} 1.0-2: a contributed package to the {R} project for statistical computing devoted to biological sequences retrieval and analysis.

Chariton AA, Court LN, Hartley DM, Colloff MJ, Hardy CM (2010) Ecological assessment of estuarine sediments by pyrosequencing eukaryotic ribosomal DNA. *Frontiers in Ecology and the Environment*, **8**, 233–238.

Chariton AA, Stephenson S, Morgan MJ *et al.* (2015) Metabarcoding of benthic eukaryote communities predicts the ecological condition of estuaries. *Environmental Pollution (Barking, Essex: 1987)*, **203**, 165–174.

Chen Y, Dai Y, Wang Y *et al.* (2016) Distribution of bacterial communities across plateau freshwater lake and upslope soils. *Journal of Environmental Sciences (China)*, **43**, 61–69.

Chen Q-H, Xu R-L, Tam NFY, Cheung SG, Shin PKS (2008) Use of ciliates (Protozoa: Ciliophora) as bioindicator to assess sediment quality of two constructed mangrove sewage treatment belts in Southern China. *Marine Pollution Bulletin*, **57**, 689–694.

Chessman BC (1995) Rapid assessment of rivers using macroinvertebrates: A procedure based on habitat-specific sampling, family level identification and a biotic index. *Australian Journal of Ecology*, **20**, 122–129.

Claparède R-É, Lachmann CFJ (1859) *Études sur les infusoires et les rhizopodes*. Kessmann.

Cline J, Braman JC, Hogrefe HH (1996) PCR fidelity of pfu DNA polymerase and other thermostable DNA polymerases. *Nucleic Acids Research*, **24**, 3546–3551.

Coats DW, Bockstahler KR (1995) Occurrence of the parasitic dinoflagellate Amoebophrya ceratii in Chesapeake Bay populations of Gymnodinium sanguineum. *Oceanographic Literature Review*, **9**, 763.

Coleman AW (1988) The Autofluorescent Flagellum: A New Phylogenetic Enigma1. *Journal of Phycology*, **24**, 118–120.

Coolen MJL, Orsi WD, Balkema C *et al.* (2013) Evolution of the plankton paleome in the Black Sea from the Deglacial to Anthropocene. *Proceedings of the*

National Academy of Sciences of the United States of America, **110**, 8609–8614.

Cooper MK, Phalen DN, Donahoe SL, Rose K, Šlapeta J (2016) The utility of diversity profiling using Illumina 18S rRNA gene amplicon deep sequencing to detect and discriminate Toxoplasma gondii among the cyst-forming coccidia. *Veterinary Parasitology*, **216**, 38–45.

Cordonier A, Gallina N, Nirel PM (2010) Essay on the characterization of environmental factors structuring communities of epilithic diatoms in the major rivers of the canton of Geneva, Switzerland. *Vie Milieu/Life Environ.*, **60**, 223–231.

Coste M, Boutry S, Tison-Rosebery J, Delmas F (2009) Improvements of the Biological Diatom Index (BDI): Description and efficiency of the new version (BDI-2006). *Ecological Indicators*, **9**, 621–650.

Cotterill FPD, Foissner W (2010) A pervasive denigration of natural history misconstrues how biodiversity inventories and taxonomy underpin scientific knowledge. *Biodiversity and conservation*, **19**, 291–303.

Cowart DA, Pinheiro M, Mouchel O *et al.* (2015) Metabarcoding is powerful yet still blind: a comparative analysis of morphological and molecular surveys of seagrass communities. *PloS One*, **10**, e0117562.

Cunha DGF, Sabogal-Paz LP, Dodds WK (2016) Land use influence on raw surface water quality and treatment costs for drinking supply in São Paulo State (Brazil). *Ecological Engineering*, **94**, 516–524.

Cushman JA (Joseph A (1928) *Foraminifera; their classification and economic use*. Sharon, Mass.

Cutler NA, Chaput DL, Oliver AE, Viles HA (2015) The spatial organization and microbial community structure of an epilithic biofilm. *FEMS microbiology ecology*, **91**.

Dam HV, Mertens A, Sinkeldam J (1994) A coded checklist and ecological indicator values of freshwater diatoms from The Netherlands. *Netherland Journal of Aquatic Ecology*, **28**, 117–133.

Davies TJ, Barraclough TG, Savolainen V, Chase MW (2004) Environmental causes for plant biodiversity gradients. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **359**, 1645–1656.

Debenay J-P, Della Patrona L, Herbland A, Goguenheim H (2009) The impact of easily oxidized material (EOM) on the meiobenthos: Foraminifera abnormalities in shrimp ponds of New Caledonia; implications for environment and paleoenvironment survey. *Marine Pollution Bulletin*, **59**, 323–335.

Decelle J, Romac S, Sasaki E, Not F, Mahé F (2014) Intracellular Diversity of the V4 and V9 Regions of the 18S rRNA in Marine Protists (Radiolarians) Assessed by High-Throughput Sequencing. *PLOS ONE*, **9**, e104297.

Deegan LA, Golden HE, Harvey CJ, Peterson BJ (1999) Influence of Environmental Variability on the Growth of Age-0 and Adult Arctic Grayling. *Transactions of the American Fisheries Society*, **128**, 1163–1175.

Deflandre G (1953) Traité de Zoologie. *In P.*, **1**, 97–148.

Deiner K, Altermatt F (2014) Transport distance of invertebrate environmental DNA in a natural river. *PloS One*, **9**, e88786.

Deiner K, Fronhofer EA, Mächler E, Walser J-C, Altermatt F (2016) Environmental DNA reveals that rivers are conveyer belts of biodiversity information. *Nature Communications*, **7**, 12544.

Delgado M, De Jonge VN, Peletier H (1991) Effect of sand movement on the growth of benthic diatoms. *Journal of Experimental Marine Biology and Ecology*, **145**, 221–231.

Dell'Anno A, Carugati L, Corinaldesi C, Riccioni G, Danovaro R (2015) Unveiling the Biodiversity of Deep-Sea Nematodes through Metabarcoding: Are We Ready to Bypass the Classical Taxonomy? *PLoS ONE*, **10**.

Dellinger M, Labat A, Perrouault L, Grellier P (2014) Haplomyxa saranae gen. nov. et sp. nov., a new naked freshwater foraminifer. *Protist*, **165**, 317–329.

DeNicola DM (2000) A review of diatoms found in highly acidic environments. *Hydrobiologia*, **433**, 111–122.

Denoyelle M, Jorissen FJ, Martin D, Galgani F, Miné J (2010) Comparison of benthic foraminifera and macrofaunal indicators of the impact of oil-based drill mud disposal. *Marine Pollution Bulletin*, **60**, 2007–2021.

Directive 2000/60/EC (2000) establishing a framework for Community action in the field of water policy. *Official Journal L 327*, 1–73.

Dopheide A, Lear G, Stott R, Lewis G (2008) Molecular Characterization of Ciliate Diversity in Stream Biofilms. *Applied and Environmental Microbiology*, **74**, 1740–1747.

Dopheide A, Lear G, Stott R, Lewis G (2009) Relative Diversity and Community Structure of Ciliates in Stream Biofilms According to Molecular and Microscopy Methods. *Applied and Environmental Microbiology*, **75**, 5261–5272.

Dowle EJ, Morgan-Richards M, Trewick SA (2013) Molecular evolution and the latitudinal biodiversity gradient. *Heredity*, **110**, 501–510.

Dowle EJ, Pochon X, C. Banks J, Shearer K, Wood SA (2016) Targeted gene enrichment and high-throughput sequencing for environmental biomonitoring: a case study using freshwater macroinvertebrates. *Molecular Ecology Resources*, **16**, 1240–1254.

Dowle E, Pochon X, Keeley N, Wood SA (2015) Assessing the effects of salmon farming seabed enrichment using bacterial community diversity and high-throughput sequencing. *FEMS microbiology ecology*, **91**, fiv089.

Duivenvoorden LJ, Roberts DT, Tucker GM (2017) Serpentine geology links to water quality and heavy metals in sediments of a stream system in central Queensland, Australia. *Environmental Earth Sciences*, **76**, 320.

Eckert KA, Kunkel TA (1991) DNA polymerase fidelity and the polymerase chain reaction. *PCR methods and applications*, **1**, 17–24.

Edgar RC (2013) UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nature Methods*, **10**, 996–998.

Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R (2011) UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*, **27**, 2194–2200.

Egan SP, Barnes MA, Hwang C-T *et al.* (2013) Rapid Invasive Species Detection by Combining Environmental DNA with Light Transmission Spectroscopy. *Conservation Letters*, **6**, 402–409.

Egge ES, Eikrem W, Edvardsen B (2015a) Deep-branching novel lineages and high diversity of haptophytes in the Skagerrak (Norway) uncovered by 454 pyrosequencing. *The Journal of Eukaryotic Microbiology*, **62**, 121–140.

Egge ES, Johannessen TV, Andersen T *et al.* (2015b) Seasonal diversity and dynamics of haptophytes in the Skagerrak, Norway, explored by high-throughput sequencing. *Molecular Ecology*, **24**, 3026–3042.

Eichmiller JJ, Best SE, Sorensen PW (2016) Effects of Temperature and Trophic State on Degradation of Environmental DNA in Lake Water. *Environmental Science & Technology*, **50**, 1859–1867.

Eiler A, Drakare S, Bertilsson S *et al.* (2013) Unveiling distribution patterns of freshwater phytoplankton by a next generation sequencing based approach. *PloS One*, **8**, e53516.

Elbrächter M (1994) Green autofluorescence — a new taxonomic feature for living dinoflagellate cysts and vegetative cells. *Review of Palaeobotany and Palynology*, **84**, 101–105.

Elbrecht V, Leese F (2015) Can DNA-Based Ecosystem Assessments Quantify Species Abundance? Testing Primer Bias and Biomass—Sequence Relationships with an Innovative Metabarcoding Protocol. *PLOS ONE*, **10**, e0130324.

Elbrecht V, Leese F (2017) *Validation and development of COI metabarcoding primers for freshwater macroinvertebrate bioassessment*. PeerJ Preprints.

Elbrecht V, Vamos E, Meissner K, Aroviita J, Leese F (2017) *Assessing strengths and weaknesses of DNA metabarcoding based macroinvertebrate identification for routine stream monitoring*. PeerJ Preprints.

Escobar D, Zea S, Sánchez JA (2012) Phylogenetic relationships among the Caribbean members of the Cliona viridis complex (Porifera, Demospongiae, Hadromerida) using nuclear and mitochondrial DNA sequences. *Molecular Phylogenetics and Evolution*, **64**, 271–284.

Esling P, Lejzerowicz F, Pawlowski J (2015) Accurate multiplexing and filtering for high-throughput amplicon-sequencing. *Nucleic Acids Research*, **43**, 2513–2524.

Evans KM, Mann DG (2009) A Proposed Protocol for Nomenclaturally Effective DNA Barcoding of Microalgae. *Phycologia*, **48**, 70–74.

Evans NT, Olds BP, Renshaw MA *et al.* (2016) Quantification of mesocosm fish and amphibian species diversity via environmental DNA metabarcoding. *Molecular Ecology Resources*, **16**, 29–41.

Evans KM, Wortley AH, Mann DG (2007) An assessment of potential diatom “barcode” genes (cox1, rbcL, 18S and ITS rDNA) and their effectiveness in determining relationships in Sellaphora (Bacillariophyta). *Protist*, **158**, 349–364.

Ewing B, Green P (1998) Base-Calling of Automated Sequencer Traces Using Phred. II. Error Probabilities. *Genome Research*, **8**, 186–194.

Falasco E, Bona F, Badino G, Hoffmann L, Ector L (2009) Diatom teratological forms and environmental alterations: a review. *Hydrobiologia*, **623**, 1–35.

Feinstein LM, Sul WJ, Blackwood CB (2009) Assessment of Bias Associated with Incomplete Extraction of Microbial DNA from Soil. *Applied and Environmental Microbiology*, **75**, 5428–5433.

Feio MJ, Aguiar FC, Almeida SFP, Ferreira MT (2012) AQUAFLORA: A predictive model based on diatoms and macrophytes for streams water quality assessment. *Ecological Indicators*, **18**, 586–598.

Feio MJ, Almeida SFP, Craveiro SC, Calado AJ (2009) A comparison between biotic indices and predictive models in stream water quality assessment based on benthic diatom communities. *Ecological Indicators*, **9**, 497–507.

Feio MJ, Poquet JM (2011) Predictive Models for Freshwater Biological Assessment: Statistical Approaches, Biological Elements and the Iberian Peninsula Experience: A Review. *International Review of Hydrobiology*, **96**, 321–346.

Ferrera I, Giner CR, Reñé A *et al.* (2016) Evaluation of Alternative High-Throughput Sequencing Methodologies for the Monitoring of Marine Picoplanktonic Biodiversity Based on rRNA Gene Amplicons. *Frontiers in Marine Science*, **3**.

Ficetola GF, Miaud C, Pompanon F, Taberlet P (2008) Species detection using environmental DNA from water samples. *Biology Letters*, **4**, 423–425.

Flakowski J, Bolivar I, Fahrni J, Pawlowski J (2005) Actin Phylogeny of Foraminifera. *The Journal of Foraminiferal Research*, **35**, 93–102.

Foissner W, Berger H (1996) A user-friendly guide to the ciliates (Protozoa, Ciliophora) commonly used by hydrobiologists as bioindicators in rivers, lakes, and waste waters, with notes on their ecology. *Freshwater Biology*, **35**, 375–482.

Fonseca VG, Nichols B, Lallias D *et al.* (2012) Sample richness and genetic diversity as drivers of chimera formation in nSSU metagenetic analyses. *Nucleic Acids Research*, **40**, e66.

Forster D, Dunthorn M, Mahé F *et al.* (2016) Benthic protists: the under-charted majority. *FEMS microbiology ecology*, **92**.

Frontalini F, Buosi C, Da Pelo S *et al.* (2009) Benthic foraminifera as bio-indicators of trace element pollution in the heavily contaminated Santa Gilla lagoon (Cagliari, Italy). *Marine Pollution Bulletin*, **58**, 858–877.

Fujita S, Iseki M, Yoshikawa S *et al.* (2005) Identification and characterization of a fluorescent flagellar protein from the brown alga Scytosiphon lomentaria (Scytosiphonales, Phaeophyceae): A flavoprotein homologous to Old Yellow Enzyme. *European Journal of Phycology*, **40**, 159–167.

Gabriels W, Lock K, De Pauw N, Goethals PLM (2010) Multimetric Macroinvertebrate Index Flanders (MMIF) for biological assessment of rivers and lakes in Flanders (Belgium). *Limnologica - Ecology and Management of Inland Waters*, **40**, 199–207.

Galal TM, Farahat EA (2015) The invasive macrophyte Pistia stratiotes L. as a bioindicator for water pollution in Lake Mariut, Egypt. *Environmental Monitoring and Assessment*, **187**, 701.

Galan M, Razzauti M, Bard E *et al.* (2016) 16S rRNA Amplicon Sequencing for Epidemiological Surveys of Bacteria in Wildlife. *mSystems*, **1**.

Gardner MJ, Hall N, Fung E *et al.* (2002) Genome sequence of the human malaria parasite Plasmodium falciparum. *Nature*, **419**, 498–511.

Garros C, Ngugi N, Githeko AE, Tuno N, Yan G (2008) Gut content identification of larvae of the Anopheles gambiae complex in western Kenya using a barcoding approach. *Molecular Ecology Resources*, **8**, 512–518.

Geisen S, Tveit AT, Clark IM *et al.* (2015) Metatranscriptomic census of active protists in soils. *The ISME Journal*, **9**, 2178–2190.

Gerbersdorf SU, Wieprecht S (2015) Biostabilization of cohesive sediments: revisiting the role of abiotic conditions, physiology and diversity of microbes, polymeric secretion, and biofilm architecture. *Geobiology*, **13**, 68–97.

Giampaoli S, Berti A, Di Maggio RM *et al.* (2014) The environmental biological signature: NGS profiling for forensic comparison of soils. *Forensic Science International*, **240**, 41–47.

Gibson JF, Shokralla S, Curry C *et al.* (2015) Large-Scale Biomonitoring of Remote and Threatened Ecosystems via High-Throughput Sequencing. *PLOS ONE*, **10**, e0138432.

Gillman LN, Keeling DJ, Gardner RC, Wright SD (2010) Faster evolution of highly conserved DNA in tropical plants. *Journal of Evolutionary Biology*, **23**, 1327–1330.

Giner CR, Forn I, Romac S *et al.* (2016) Environmental Sequencing Provides Reasonable Estimates of the Relative Abundance of Specific Picoeukaryotes. *Applied and Environmental Microbiology*, **82**, 4757–4766.

Godhe A, Asplund ME, Härnström K *et al.* (2008) Quantification of Diatom and Dinoflagellate Biomasses in Coastal Marine Seawater Samples by Real-Time PCR. *Applied and Environmental Microbiology*, **74**, 7174–7182.

Goldberg CS, Sepulveda A, Ray A, Baumgardt J, Waits LP (2013) Environmental DNA as a new method for early detection of New Zealand mudsnails (Potamopyrgus antipodarum). *Freshwater Science*, **32**, 792–800.

Goldberg CS, Turner CR, Deiner K *et al.* (2016) Critical considerations for the application of environmental DNA methods to detect aquatic species. *Methods in Ecology and Evolution*, **7**, 1299–1307.

Goldstein ST, Barker WW (1990) Gametogenesis in the Monothalamous Agglutinated Foraminifer Cribrothalammina alba. *The Journal of Protozoology*, **37**, 20–27.

Gong J, Dong J, Liu X, Massana R (2013) Extremely high copy numbers and polymorphisms of the rDNA operon estimated from single cell analysis of oligotrich and peritrich ciliates. *Protist*, **164**, 369–379.

Gooday AJ (2002) Organic-walled allogromiids: aspects of their occurrence, diversity and ecology in marine habitats. *Journal of Foraminiferal Research*, **32**, 384–399.

Gooday (2009) Historical records of coastal eutrophication-induced hypoxia. *Biogeosciences*, **6**, 1707–1745.

Gooday AJ, Anikeeva OV, Pawlowski J (2011) New genera and species of monothalamous Foraminifera from Balaclava and Kazach'ya Bays (Crimean Peninsula, Black Sea). *Marine Biodiversity*, **41**, 481–494.

Gooday AJ, Anikeeva OV, Sergeeva NG (2006) Tinogullmia lukyanovae sp. nov.—a monothalamous, organic-walled foraminiferan from the coastal Black Sea. *Journal of the Marine Biological Association of the United Kingdom*, **86**, 43–49.

Gooday AJ, Bowser SS, Cedhagen T *et al.* (2005) Monothalamous foraminiferans and gromiids (Protista) from western Svalbard: A preliminary survey *Published in collaboration with the University of Bergen and the Institute of Marine Research, Norway, and the Marine Biological Laboratory, University of Copenhagen, Denmark*. *Marine Biology Research*, **1**, 290–312.

Gooday AJ, Holzmann M, Caulle C *et al.* (2017) Giant protists (xenophyophores, Foraminifera) are exceptionally diverse in parts of the abyssal eastern Pacific licensed for polymetallic nodule exploration. *Biological Conservation*, **207**, 106–116.

Gooday AJ, Holzmann M, Guiard J, Cornelius N, Pawlowski J (2004) A new monothalamous foraminiferan from 1000 to 6300 m water depth in the Weddell Sea: morphological and molecular characterisation. *Deep Sea Research Part II: Topical Studies in Oceanography*, **51**, 1603–1616.

Gooday AJ, Jorissen FJ (2012) Benthic foraminiferal biogeography: controls on global distribution patterns in deep-water settings. *Annual Review of Marine Science*, **4**, 237–262.

Gooday AJ, Pawlowski J (2004) Conqueria laevis gen. and sp. nov., a new soft-walled, monothalamous foraminiferan from the deep Weddell Sea. *Journal of the Marine Biological Association of the United Kingdom*, **84**, 919–924.

Gooday AJ, da Silva AA, Koho KA, Lecroq B, Pearce RB (2010) The “mica sandwich”; a remarkable new genus of Foraminifera (Protista, Rhizaria) from the Nazare Canyon (Portuguese margin, NE Atlantic). *Micropaleontology*, **56**, 345–357.

Gould SJ, Subramani S (1988) Firefly luciferase as a tool in molecular and cell biology. *Analytical Biochemistry*, **175**, 5–13.

Gouy M, Guindon S, Gascuel O (2010) SeaView Version 4: A Multiplatform Graphical User Interface for Sequence Alignment and Phylogenetic Tree Building. *Molecular Biology and Evolution*, **27**, 221–224.

Gray HJ (2015) Aquatic macrophytes and periphyton communities as bioindicators of lake trophic status in Riding Mountain National Park, Manitoba. University of British Columbia.

Gribble KE, Anderson DM (2007) High intraindividual, intraspecific, and interspecific variability in large-subunit ribosomal DNA in the heterotrophic dinoflagellates Protoperidinium, Diplopsalis, and Preperidinium (Dinophyceae). *Phycologia*, **46**, 315–324.

Groendahl S, Kahlert M, Fink P (2017) The best of both worlds: A combined approach for analyzing microalgal diversity via metabarcoding and morphology-based methods. *PloS One*, **12**, e0172808.

Grospietsch T (1958) Wechseltierchen (Rhizopoden). *Sammlung: Einführung in die Kleinlebewelt. Kosmos*, **III**.

Groussin M, Pawlowski J, Yang Z (2011) Bayesian relaxed clock estimation of divergence times in foraminifera. *Molecular Phylogenetics and Evolution*, **61**, 157–166.

Gruber AR, Lorenz R, Bernhart SH, Neuböck R, Hofacker IL (2008) The Vienna RNA Websuite. *Nucleic Acids Research*, **36**, W70–W74.

Guardiola M, Uriz MJ, Taberlet P *et al.* (2015) Deep-Sea, Deep-Sequencing: Metabarcoding Extracellular DNA from Sediments of Marine Canyons. *PloS One*, **10**, e0139633.

Guillou L, Viprey M, Chambouvet A *et al.* (2008) Widespread occurrence and genetic diversity of marine parasitoids belonging to Syndiniales (Alveolata). *Environmental Microbiology*, **10**, 3349–3365.

Habura A, Goldstein ST, Broderick S, Bowser SS (2008) A bush, not a tree: The extraordinary diversity of cold-water basal foraminiferans extends to warm-water environments. *Limnology and Oceanography*, **53**, 1339–1351.

Habura A, Goldstein ST, Parfrey LW, Bowser SS (2006) Phylogeny and Ultrastructure of Miliammina fusca: Evidence for Secondary Loss of Calcification in a Miliolid Foraminifer. *Journal of Eukaryotic Microbiology*, **53**, 204–210.

Habura A, Pawlowski J, Hanes SD, Bowser SS (2004) Unexpected Foraminiferal Diversity Revealed by Small-subunit rDNA Analysis of Antarctic Sediment. *Journal of Eukaryotic Microbiology*, **51**, 173–179.

Habura A, Wegener L, Travis JL, Bowser SS (2005) Structural and Functional Implications of an Unusual Foraminiferal β-Tubulin. *Molecular Biology and Evolution*, **22**, 2000–2009.

Hajibabaei M, Janzen DH, Burns JM, Hallwachs W, Hebert PDN (2006) DNA barcodes distinguish species of tropical Lepidoptera. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 968–971.

Hajibabaei M, Shokralla S, Zhou X, Singer GAC, Baird DJ (2011) Environmental barcoding: a next-generation sequencing approach for biomonitoring applications using river benthos. *PloS One*, **6**, e17497.

Hajibabaei M, Spall JL, Shokralla S, van Konynenburg S (2012) Assessing biodiversity of a freshwater benthic macroinvertebrate community through non-destructive environmental barcoding of DNA from preservative ethanol. *BMC ecology*, **12**, 28.

Hamsher SE, Evans KM, Mann DG, Poulíčková A, Saunders GW (2011) Barcoding diatoms: exploring alternatives to COI-5P. *Protist*, **162**, 405–422.

Hanner R, Becker S, Ivanova NV, Steinke D (2011) FISH-BOL and seafood identification: geographically dispersed case studies reveal systemic market substitution across Canada. *Mitochondrial DNA*, **22 Suppl 1**, 106–122.

Hartikainen H, Ashford OS, Berney C *et al.* (2014) Lineage-specific molecular probing reveals novel diversity and ecological partitioning of haplosporidians. *The ISME journal*, **8**, 177–186.

Haynes JR (1981) *Foraminifera*. Wiley Online Library.

Hebert PD, Cywinska A, Ball SL, others (2003) Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London B: Biological Sciences*, **270**, 313–321.

Hebert PDN, Penton EH, Burns JM, Janzen DH, Hallwachs W (2004) Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper

butterfly Astraptes fulgerator. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 14812–14817.

Hedley RH (1960) The Iron-Containing Shell of Gromia Oviformis (Rhizopoda). *Journal of Cell Science*, **s3**-**101**, 279–293.

Heger TJ, Mitchell EAD, Todorov M *et al.* (2010) Molecular phylogeny of euglyphid testate amoebae (Cercozoa: Euglyphida) suggests transitions between marine supralittoral and freshwater/terrestrial environments are infrequent. *Molecular Phylogenetics and Evolution*, **55**, 113–122.

Henderson G, Cox F, Kittelmann S *et al.* (2013) Effect of DNA Extraction Methods and Sampling Techniques on the Apparent Structure of Cow and Sheep Rumen Microbial Communities. *PLOS ONE*, **8**, e74787.

Hewlett R (2000) Implications of taxonomic resolution and sample habitat for stream classification at a broad geographic scale. *Journal of the North American Benthological Society*, **19**, 352–361.

Heyse G, Jönsson F, Chang W-J, Lipps HJ (2010) RNA-dependent control of gene amplification. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 22134–22139.

Hill BH, Herlihy AT, Kaufmann PR *et al.* (2000) Use of periphyton assemblage data as an index of biotic integrity. *Journal of the North American Benthological Society*, **19**, 50–67.

Hill BH, Stevenson RJ, Pan Y *et al.* (2001) Comparison of correlations between environmental characteristics and stream diatom assemblages characterized at genus and species levels. *Journal of the North American Benthological Society*, **20**, 299–310.

Hilsenhoff WL (1988) Rapid Field Assessment of Organic Pollution with a Family-Level Biotic Index. *Journal of the North American Benthological Society*, **7**, 65–68.

Hoffmann C, Schubert G, Calvignac-Spencer S (2016) Aquatic biodiversity assessment for the lazy. *Molecular Ecology*, **25**, 846–848.

Hofmann G, Werum M, Lange-Bertalot H (2011) *Diatomeen im Süßwasser-Benthos von Mitteleuropa: Bestimmungsflora Kieselalgen für die ökologische Praxis ; über 700 der häufigsten Arten und ihre Ökologie*. Gantner.

Hollingsworth PM, Forrest LL, Spouge JL *et al.* (2009) A DNA barcode for land plants. *Proceedings of the National Academy of Sciences*, **106**, 12794–12797.

Holzmann M, Habura A, Giles H, Bowser SS, Pawlowski J (2003) Freshwater foraminiferans revealed by analysis of environmental DNA samples. *The Journal of Eukaryotic Microbiology*, **50**, 135–139.

Holzmann M, Pawlowski J (2002) Freshwater Foraminiferans from Lake Geneva: Past and Present. *Journal of Foraminiferal Research*, **32**, 344–350.

Hoogenraad HR, de Groot AA (1940) Zoetwaterrhizopoden en heliozoën. In: *Sijthoff* Fauna von Nederland., pp. 1–303.

Hottinger L, Halicz E, Reiss Z (1993) *Recent foraminiferida from the Gulf of Aqaba, Red Sea*. Opera Sazu, Ljubljana.

Hou Y, Sierra R, Bassen D *et al.* (2013) Molecular Evidence for β-tubulin Neofunctionalization in Retaria (Foraminifera and Radiolarians). *Molecular Biology and Evolution*, **30**, 2487–2493.

Huang Z, Han L, Zeng L, Xiao W, Tian Y (2016) Effects of land use patterns on stream water quality: a case study of a small-scale watershed in the Three Gorges Reservoir Area, China. *Environmental Science and Pollution Research International*, **23**, 3943–3955.

Huemer P, Karsholt O, Mutanen M (2014) DNA barcoding as a screening tool for cryptic diversity: an example from Caryocolum, with description of a new species (Lepidoptera, Gelechiidae). *ZooKeys*, **404**, **404**, 91–111.

Hürlimann J, Niederhauser P (2007) Méthodes d'analyse et d'appréciation des cours d'eau. Diatomées Niveau R (region). *Etat de l'environnement n° 0740. Office fédéral de l'environnement, Berne*.

Huxley-Jones E, Shaw JLA, Fletcher C, Parnell J, Watts PC (2012) Use of DNA barcoding to reveal species composition of convenience seafood. *Conservation Biology: The Journal of the Society for Conservation Biology*, **26**, 367–371.

Jesus B, Brotas V, Ribeiro L *et al.* (2009) Adaptations of microphytobenthos assemblages to sediment type and tidal position. *Continental Shelf Research*, **29**, 1624–1634.

Ji Y, Ashton L, Pedley SM *et al.* (2013) Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding. *Ecology Letters*, **16**, 1245–1257.

Jiang Y, Xu H, Hu X *et al.* (2011) An approach to analyzing spatial patterns of planktonic ciliate communities for monitoring water quality in Jiaozhou Bay, northern China. *Marine Pollution Bulletin*, **62**, 227–235.

Jiang Y, Xu H, Hu X, Warren A, Song W (2013) Functional groups of marine ciliated protozoa and their relationships to water quality. *Environmental Science and Pollution Research International*, **20**, 5272–5280.

Johnson RK (2007) *Bedömningsgrunder för bottenfauna i sjöar och vattendrag: användarmanual och bakgrundsdokument*. Sveriges lantbruksuniversitet.

Jordaan K, Bezuidenhout CC (2016) Bacterial community composition of an urban river in the North West Province, South Africa, in relation to physico-chemical water quality. *Environmental Science and Pollution Research*, **23**, 5868–5880.

Jorissen FJ, Bicchi E, Duchemin G *et al.* (2009) Impact of oil-based drill mud disposal on benthic foraminiferal assemblages on the continental margin off Angola. *Deep Sea Research Part II: Topical Studies in Oceanography*, **56**, 2270–2291.

Jousson O, Bartoli P, Zaninetti L, Pawlowski J (1998) Use of the ITS rDNA for elucidation of some life-cycles of Mesometridae (Trematoda, Digenea). *International Journal for Parasitology*, **28**, 1403–1411.

Jousson O, Pawlowski J, Zaninetti L *et al.* (2000) Invasive alga reaches California. *Nature*, **408**, 157–158.

Joy MK, Death RG (2002) Predictive modelling of freshwater fish as a biomonitoring tool in New Zealand. *Freshwater Biology*, **47**, 2261–2275.

Jyväsjärvi J, Aroviita J, Hämäläinen H (2014) An extended Benthic Quality Index for assessment of lake profundal macroinvertebrates: addition of indicator taxa by multivariate ordination and weighted averaging. *Freshwater Science*, **33**, 995–1007.

Kahlert M, Albert R-L, Anttila E-L *et al.* (2009) Harmonization is more important than experience—results of the first Nordic–Baltic diatom intercalibration exercise 2007 (stream monitoring). *Journal of Applied Phycology*, **21**, 471–482.

Karp DS, Judson S, Daily GC, Hadly EA (2014) Molecular diagnosis of bird-mediated pest consumption in tropical farmland. *SpringerPlus*, **3**, 630.

Keck F, Bouchez A, Franc A, Rimet F (2016) Linking phylogenetic similarity and pollution sensitivity to develop ecological assessment methods: a test with river diatoms. *Journal of Applied Ecology*, n/a-n/a.

Kelly MG, Adams C, Graves a C (2001) *The Trophic Diatom Index: a User's Manual; Revised Edition*. Environment Agency.

Kelly M, Bennett C, Coste M *et al.* (2008) A comparison of national approaches to setting ecological status boundaries in phytobenthos assessment for the European Water Framework Directive: results of an intercalibration exercise. *Hydrobiologia*, **621**, 169–182.

Kelly M, Bennett C, Coste M *et al.* (2009) A comparison of national approaches to setting ecological status boundaries in phytobenthos assessment for the European Water Framework Directive: results of an intercalibration exercise. *Hydrobiologia*, **621**, 169–182.

Kelly MG, Penny CJ, Whitton BA (1995) Comparative performance of benthic diatom indices used to assess river water quality. *Hydrobiologia*, **302**, 179–188.

Kelly M, Urbanic G, Acs E *et al.* (2014) Comparing aspirations: intercalibration of ecological status concepts across European lakes for littoral diatoms. *Hydrobiologia*, **734**, 125–141.

Kennedy MP, Lang P, Tapia Grimaldo J *et al.* (2016) The Zambian Macrophyte Trophic Ranking scheme, ZMTR: A new biomonitoring protocol to assess the trophic status of tropical southern African rivers. *Aquatic Botany*, **131**, 15–27.

Kermarrec L, Franc A, Rimet F *et al.* (2013) Next-generation sequencing to inventory taxonomic diversity in eukaryotic communities: a test for freshwater diatoms. *Molecular Ecology Resources*, **13**, 607–619.

Kermarrec L, Franc A, Rimet F *et al.* (2014) A next-generation sequencing approach to river biomonitoring using benthic diatoms. *Freshwater Science*, **33**, 349–363.

Kim E, Harrison JW, Sudek S *et al.* (2011) Newly identified and diverse plastid-bearing branch on the eukaryotic tree of life. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 1496–1500.

Králová-Hromadová I, Bazsalovicsová E, Oros M, Scholz T (2012) Sequence structure and intragenomic variability of ribosomal ITS2 in monozoic tapeworms of the genus Khawia (Cestoda: Caryophyllidea), parasites of cyprinid fish. *Parasitology Research*, **111**, 1621–1627.

Krammer K, Lange-Bertalot H (1986) *Bacillariophyceae*. Stuttgart, Germany.

Kuhar U, Germ M, Gaberščik A, Urbanič G (2011) Development of a River Macrophyte Index (RMI) for assessing river ecological status. *Limnologica - Ecology and Management of Inland Waters*, **41**, 235–243.

Lallias D, Hiddink JG, Fonseca VG *et al.* (2015) Environmental metabarcoding reveals heterogeneous drivers of microbial eukaryote diversity in contrasting estuarine ecosystems. *The ISME journal*, **9**, 1208–1221.

Lambert AS, Dabrin A, Morin S *et al.* (2016) Temperature modulates phototrophic periphyton response to chronic copper exposure. *Environmental Pollution*, **208**, **Part B**, 821–829.

Lang I, Kaczmarska I (2011) A protocol for a single-cell PCR of diatoms from fixed samples: method validation using Ditylum brightwellii (T. West) Grunow. *Diatom Research*, **26**, 43–49.

Lange-Bertalot H (2001) *Diatoms of the European Inland Waters and Comparable Habitats*. Ruggell, Lichtenstein.

Lange-Bertalot H, Metzeltin D (1996) *Indicators of oligotrophy: 800 taxa representative of three ecologically distinct lake types : carbonate buffered, oligodystrophic, weakly buffered soft water*. Koeltz Scientific Books.

Laroche O, Wood SA, Tremblay LA *et al.* (2016) First evaluation of foraminiferal metabarcoding for monitoring environmental impact from an offshore oil drilling site. *Marine Environmental Research*, **120**, 225–235.

Laval-Peuto M, Rassoulzadegan F (1988) Autofluorescence of marine planktonic Oligotrichina and other ciliates. *Hydrobiologia*, **159**, 99–110.

Lavoie I, Somers KM, Paterson AM, Dillon PJ (2005) Assessing scales of variability in benthic diatom community structure. *Journal of Applied Phycology*, **17**, 509–513.

Le Bescot N, Mahé F, Audic S *et al.* (2016) Global patterns of pelagic dinoflagellate diversity across protist size classes unveiled by metabarcoding. *Environmental Microbiology*, **18**, 609–626.

Le Coadou L, Le Ménach K, Labadie P *et al.* (2017) Quality survey of natural mineral water and spring water sold in France: Monitoring of hormones, pharmaceuticals, pesticides, perfluoroalkyl substances, phthalates, and alkylphenols at the ultra-trace level. *Science of The Total Environment*.

Lecroq B, Gooday AJ, Cedhagen T, Sabbatini A, Pawlowski J (2009a) Molecular analyses reveal high levels of eukaryotic richness associated with enigmatic deep-sea protists (Komokiacea). *Marine Biodiversity*, **39**, 45–55.

Lecroq B, Gooday AJ, Tsuchiya M, Pawlowski J (2009b) A new genus of xenophyophores (Foraminifera) from Japan Trench: morphological description, molecular phylogeny and elemental analysis. *Zoological Journal of the Linnean Society*, **156**, 455–464.

Lecroq B, Lejzerowicz F, Bachar D *et al.* (2011) Ultra-deep sequencing of foraminiferal microbarcodes unveils hidden richness of early monothalamous lineages in deep-sea sediments. *Proceedings of the National Academy of Sciences*, **108**, 13177–13182.

Lee JJ, Anderson OR (1991) Symbiosis in foraminifera. In: *Biology of foraminifera* Academic Press., pp. 157–220. London.

Lee S, Basu S, Tyler CW, Wei IW (2004) Ciliate populations as bio-indicators at Deer Island Treatment Plant. *Advances in Environmental Research*, **8**, 371–378.

Lee DF, Lu J, Chang S, Loparo JJ, Xie XS (2016) Mapping DNA polymerase errors by single-molecule sequencing. *Nucleic Acids Research*, **44**, e118.

Leidy J (1879) Freshwater Rhizopods of North America. U.S. Geological Survey, **12**, 277-280.

Lejzerowicz F, Esling P, Pillet L *et al.* (2015) High-throughput sequencing and morphology perform equally well for benthic monitoring of marine ecosystems. *Scientific Reports*, **5**, 13932.

Lejzerowicz F, Pawlowski J, Fraissinet-Tachet L, Marmeisse R (2010) Molecular evidence for widespread occurrence of Foraminifera in soils. *Environmental Microbiology*, **12**, 2518–2526.

Lejzerowicz F, Voltsky I, Pawlowski J (2013) Identifying active foraminifera in the Sea of Japan using metatranscriptomic approach. *Deep Sea Research Part II: Topical Studies in Oceanography*, **86**–**87**, 214–220.

Lenoir A, Coste M (1996) Development of a practical diatom index of overall water quality applicable to the French National Water Board Network. In: *Use of algae for monitoring rivers II*, pp. 29–43. Innsbruck, Austria.

Lentendu G, Wubet T, Chatzinotas A *et al.* (2014) Effects of long-term differential fertilization on eukaryotic microbial communities in an arable soil: a multiple barcoding approach. *Molecular ecology*, **23**, 3341–3355.

Leray M, Knowlton N (2015) DNA barcoding and metabarcoding of standardized samples reveal patterns of marine benthic diversity. *Proceedings of the National Academy of Sciences of the United States of America*, **112**, 2076–2081.

Llamas B, Valverde G, Fehren-Schmitz L *et al.* (2017) From the field to the laboratory: Controlling DNA contamination in human ancient DNA research in the high-throughput sequencing era. *STAR: Science & Technology of Archaeological Research*, **3**, 1–14.

Lobo EA, Callegaro VLM (2003) Use of epilithic diatoms as bioindicators from lotic systems in southern Brazil, with special emphasis on eutrophication. *Acta Limnologica Brasiliensia*, **16**, 25–40.

Lobo EA, Heinrich CG, Schuch M, Wetzel CE, Ector L (2016) Diatoms as Bioindicators in Rivers. In: *River Algae* (ed JR ON), pp. 245–271. Springer International Publishing.

Loeblich AR, Tappan H (1960) Saedeleeria, new genus of the family Allogromiidae (Foraminifera). *Proceedings of The Biological Society of Washington*, **73**, 195–196.

Loeblich AR, Tappan HN (1988) *Foraminiferal genera and their classification*. Van Nostrand Reinhold Co.

Logares R, Audic S, Bass D *et al.* (2014) Patterns of rare and abundant marine microbial eukaryotes. *Current biology: CB*, **24**, 813–821.

Logares R, Audic S, Santini S *et al.* (2012) Diversity patterns and activity of uncultured marine heterotrophic flagellates unveiled with pyrosequencing. *The ISME Journal*, **6**, 1823–1833.

Logares R, Bråte J, Bertilsson S *et al.* (2009) Infrequent marine–freshwater transitions in the microbial world. *Trends in Microbiology*, **17**, 414–422.

Logares R, Mangot J-F, Massana R (2015) Rarity in aquatic microbes: placing protists on the map. *Research in Microbiology*, **166**, 831–841.

Logares R, Shalchian-Tabrizi K, Boltovskoy A, Rengefors K (2007) Extensive dinoflagellate phylogenies indicate infrequent marine–freshwater transitions. *Molecular Phylogenetics and Evolution*, **45**, 887–903.

Lohan KMP, Fleischer RC, Carney KJ, Holzer KK, Ruiz GM (2016) Amplicon-Based Pyrosequencing Reveals High Diversity of Protistan Parasites in Ships' Ballast Water: Implications for Biogeography and Infectious Diseases. *Microbial Ecology*, **71**, 530–542.

Longet D, Pawlowski J (2007) Higher-level phylogeny of Foraminifera inferred from the RNA polymerase II (RPB1) gene. *European Journal of Protistology*, **43**, 171–177.

Lopes ML, Rodrigues AM, Quintino V (2014) Ecological effects of contaminated sediments following a decade of no industrial effluents emissions: the Sediment Quality Triad approach. *Marine Pollution Bulletin*, **87**, 117–130.

Lowe WH (1974) Environmental Requirements and Pollution Tolerance of Freshwater Diatoms. *National Environmental Research Center | US EPA*.

Luddington IA, Kaczmarska I, Lovejoy C (2012) Distance and character-based evaluation of the V4 region of the 18S rRNA gene for the identification of diatoms (Bacillariophyceae). *PloS One*, **7**, e45664.

MacDougall MJ, Paterson AM, Winter JG *et al.* (2016) Response of periphytic diatom communities to multiple stressors influencing lakes in the Muskoka River Watershed, Ontario, Canada. *Freshwater Science*, **36**, 77–89.

MacGillivary ML, Kaczmarska I (2011) Survey of the efficacy of a short fragment of the rbcL gene as a supplemental DNA barcode for diatoms. *The Journal of Eukaryotic Microbiology*, **58**, 529–536.

Madoni P, Bassanini N (1999) Longitudinal changes in the ciliated protozoa communities along a fluvial system polluted by organic matter. *European Journal of Protistology*, **35**, 391–402.

Madoni P, Braghiroli S (2007) Changes in the ciliate assemblage along a fluvial system related to physical, chemical and geomorphological characteristics. *European Journal of Protistology*, **43**, 67–75.

Madoni P, Braghiroli S, Fioravanti M, Galassi L (2008) Assessment of the running water quality by comparing ciliate and macroinvertebrate community structure. *Italian Journal of Zoology*, **75**, 243–252.

Mahé F, Rognes T, Quince C, de Vargas C, Dunthorn M (2015) Swarm v2: highly-scalable and high-resolution amplicon clustering. *PeerJ PrePrints*, **e1503**.

Mahé F, de Vargas C, Bass D *et al.* (2017) Parasites dominate hyperdiverse soil protist communities in Neotropical rainforests. *Nature Ecology & Evolution*, **1**, 91.

Majewski W (2005) Benthic foraminiferal communities: distribution and ecology in Admiralty Bay, King George Island, West Antarctica. *Polish Polar Research*, **26**, 159–214.

Majewski W, Lecroq B, Sinniger F, Pawlowski J (2007) Monothalamous foraminifera from Admiralty Bay, King George Island, West Antarctica. *Polish Polar Research*, **28**, 187–210.

Manenti R, Bianchi B (2014) Distribution of the triclad Polycelis felina (Planariidae) in Aezkoa Mountains: effect of stream biotic features. *Acta Zool. Bulg*, **66**, 271–275.

Mann DG, Droop SJM (1996) Biodiversity, biogeography and conservation of diatoms. In: *Biogeography of Freshwater Algae* Developments in Hydrobiology. (ed Kristiansen J), pp. 19–32. Springer Netherlands.

Mann DG, Sato S, Trobajo R, Vanormelingen P, Souffreau C (2010) DNA barcoding for species identification and discovery in diatoms. *Cryptogamie Algologie*, **31**, 557–577.

Manoylov KM (2014) Taxonomic identification of algae (morphological and molecular): species concepts, methodologies, and their implications for ecological bioassessment. *Journal of Phycology*, **50**, 409–424.

Mao D, Luo Y, Mathieu J *et al.* (2014) Persistence of extracellular DNA in river sediment facilitates antibiotic resistance gene propagation. *Environmental Science & Technology*, **48**, 71–78.

Martin G, Reyes Fernandez M de los (2012) Diatoms as Indicators of Water Quality and Ecological Status: Sampling, Analysis and Some Ecological Remarks. In: *Ecological Water Quality - Water Treatment and Reuse* (ed Voudouris K), p. . InTech.

Martín-Cereceda M, Serrano S, Guinea A (1996) A comparative study of ciliated protozoa communities in activated-sludge plants. *FEMS Microbiology Ecology*, **21**, 267–276.

Massana R, del Campo J, Sieracki ME, Audic S, Logares R (2014) Exploring the uncultured microeukaryote majority in the oceans: reevaluation of ribogroups within stramenopiles. *The ISME journal*, **8**, 854–866.

Massana R, Gobet A, Audic S *et al.* (2015) Marine protist diversity in European coastal waters and sediments as revealed by high-throughput sequencing. *Environmental Microbiology*, **17**, 4035–4049.

Mateo P, Leganés F, Perona E, Loza V, Fernández-Piñas F (2015) Cyanobacteria as bioindicators and bioreporters of environmental analysis in aquatic ecosystems. *Biodiversity and Conservation*, **24**, 909–948.

Mattia FD, Imazio S, Grassi F *et al.* (2008) Study of genetic relationships between wild and domesticated grapevine distributed from Middle East Regions to European countries. *RENDICONTI LINCEI*, **19**, 223.

Mazor RD, Reynoldson TB, Rosenberg DM, Resh VH (2006) Effects of biotic assemblage, classification, and assessment method on bioassessment performance. *Canadian Journal of Fisheries and Aquatic Sciences*, **63**, 394–411.

McCormick PV, Cairns J (1994) Algae as indicators of environmental change. *Journal of Applied Phycology*, **6**, 509–526.

McEnery M, Lee JJ (1976) Allogromia laticollaris: a foraminiferan with an unusual apogamic metagenic fife cycles. *Journal of Protozoology*, **23**, 94–108.

McFarland B, Carse F, Sandin L (2010) Littoral macroinvertebrates as indicators of lake acidification within the UK. *Aquatic Conservation: Marine and Freshwater Ecosystems*, **20**, S105–S116.

McGee BL, Pinkney AE, Velinsky DJ *et al.* (2009) Using the Sediment Quality Triad to characterize baseline conditions in the Anacostia River, Washington, DC, USA. *Environmental Monitoring and Assessment*, **156**, 51–67.

McInerney P, Adams P, Hadi MZ (2014) Error Rate Comparison during Polymerase Chain Reaction by DNA Polymerase. *Molecular Biology International*, **2014**, e287430.

Medinger R, Nolte V, Pandey RV *et al.* (2010) Diversity in a hidden world: potential and limitation of next-generation sequencing for surveys of molecular diversity of eukaryotic microorganisms. *Molecular Ecology*, **19 Suppl 1**, 32–40.

Meiklejohn KA, Wallman JF, Dowton M (2013) DNA Barcoding Identifies all Immature Life Stages of a Forensically Important Flesh Fly (Diptera: Sarcophagidae). *Journal of Forensic Sciences*, **58**, 184–187.

Meisterfeld R, Holzmann M, Pawlowski J (2001) Morphological and molecular characterization of a new terrestrial allogromiid species: Edaphoallogromia australica gen. et spec. nov. (Foraminifera) from Northern Queensland (Australia). *Protist*, **152**, 185–192.

Mercereau-Puijalon O, Barale J-C, Bischoff E (2002) Three multigene families in Plasmodium parasites: facts and questions. *International Journal for Parasitology*, **32**, 1323–1344.

Miler O, Porst G, McGoff E *et al.* (2013) Morphological alterations of lake shores in Europe: A multimetric ecological assessment approach using benthic macroinvertebrates. *Ecological Indicators*, **34**, 398–410.

Miller DN (2001) Evaluation of gel filtration resins for the removal of PCR-inhibitory substances from soils and sediments. *Journal of Microbiological Methods*, **44**, 49–58.

Miranda LN, Zhuang Y, Zhang H, Lin S (2012) Phylogenetic analysis guided by intragenomic SSU rDNA polymorphism refines classification of “Alexandrium tamarense” species complex. *Harmful Algae*, **16**, 35–48.

Misra AK, Tiwari PK, Venturino E (2016) Modeling the impact of awareness on the mitigation of algal bloom in a lake. *Journal of Biological Physics*, **42**, 147–165.

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

Mojtahid M, Jorissen F, Pearson TH (2008) Comparison of benthic foraminiferal and macrofaunal responses to organic pollution in the Firth of Clyde (Scotland). *Marine Pollution Bulletin*, **56**, 42–76.

Moniz MBJ, Kaczmarska I (2009) Barcoding diatoms: Is there a good marker? *Molecular Ecology Resources*, **9 Suppl s1**, 65–74.

Moniz MBJ, Kaczmarska I (2010) Barcoding of diatoms: nuclear encoded ITS revisited. *Protist*, **161**, 7–34.

Monteagudo L, Moreno JL (2016) Benthic freshwater cyanobacteria as indicators of anthropogenic pressures. *Ecological Indicators*, **67**, 693–702.

Moreira LB, Castro ÍB, Hortellani MA *et al.* (2017) Effects of harbor activities on sediment quality in a semi-arid region in Brazil. *Ecotoxicology and Environmental Safety*, **135**, 137–151.

Moreira D, López-García P (2014) The rise and fall of Picobiliphytes: how assumed autotrophs turned out to be heterotrophs. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*, **36**, 468–474.

Morrisey DJ, Howitt L, Underwood AJ, Stark JS (1992) Spatial variation in soft-sediment benthos. *Marine ecology progress series. Oldendorf*, **81**, 197–204.

Moyer GR, Díaz-Ferguson E, Hill JE, Shea C (2014) Assessing Environmental DNA Detection in Controlled Lentic Systems. *PLOS ONE*, **9**, e103767.

Mrva M (2009) Re-discovery of Lieberkuehnia wageneri Claparède et Lachmann, 1859 (Rhizaria, Foraminifera): Taxonomical and Morphological Studies Based on a Slovak Population. *Acta Protozoologica*, **2009**, 111117.

Munz NA, Burdon FJ, de Zwart D *et al.* (2017) Pesticides drive risk of micropollutants in wastewater-impacted streams during low flow conditions. *Water Research*, **110**, 366–377.

Murray JW (2006) *Ecology and Applications of Benthic Foraminifera*. Cambridge University Press.

Mysara M, Njima M, Leys N, Raes J, Monsieurs P (2017) From reads to operational taxonomic units: an ensemble processing pipeline for MiSeq amplicon sequencing data. *GigaScience*, **6**, 1–10.

Nauss RN (1949) Reticulomyxa filosa Gen. Et Sp. Nov., A New Primitive Plasmodium. *Bulletin of the Torrey Botanical Club*, **76**, 161–173.

Nguyen-Viet H, Bernard N, Mitchell E a. D *et al.* (2007) Relationship between testate amoeba (protist) communities and atmospheric heavy metals accumulated in Barbula indica (bryophyta) in Vietnam. *Microbial Ecology*, **53**, 53–65.

Nicolau A, Dias N, Mota M, Lima N (2001) Trends in the use of protozoa in the assessment of wastewater treatment. *Research in Microbiology*, **7**, 621–630.

van Nieukerken EJ, Wagner DL, Baldessari M *et al.* (2012) Antispila oinophylla new species (Lepidoptera, Heliozelidae), a new North American grapevine leafminer invading Italian vineyards: taxonomy, DNA barcodes and life cycle. *ZooKeys*, 29–77.

Nitsche F, Arndt H (2015) Comparison of similar Arctic and Antarctic morphotypes of heterotrophic protists regarding their genotypes and ecotypes. *Protist*, **166**, 42–57.

Nolte V, Pandey RV, Jost S *et al.* (2010) Contrasting seasonal niche separation between rare and abundant taxa conceals the extent of protist diversity. *Molecular Ecology*, **19**, 2908–2915.

O'Brien A, Townsend K, Hale R, Sharley D, Pettigrove V (2016) How is ecosystem health defined and measured? A critical review of freshwater and estuarine studies. *Ecological Indicators*, **69**, 722–729.

Oksanen J, Blanchet FG, Kindt R *et al.* (2013) Package "vegan." *Community ecology package, version*, **2**.

Okshevsky M, Meyer RL (2015) The role of extracellular DNA in the establishment, maintenance and perpetuation of bacterial biofilms. *Critical Reviews in Microbiology*, **41**, 341–352.

Orfanidis S, Dencheva K, Nakou K *et al.* (2014) Benthic macrophyte metrics as bioindicators of water quality: towards overcoming typological boundaries and methodological tradition in Mediterranean and Black Seas. *Hydrobiologia*, **740**, 61–78.

Orsini L, Procaccini G, Sarno D, Montresor M (2004) Multiple rDNA ITS-types within the diatom Pseudo-nitzschia delicatissima (Bacillariophyceae) and their relative abundances across a spring bloom in the Gulf of Naples. *Marine ecology-progress series*, **271**, 87–98.

Pall K, Moser V (2009) Austrian Index Macrophytes (AIM-Module 1) for lakes: a Water Framework Directive compliant assessment system for lakes using aquatic macrophytes. *Hydrobiologia*, **633**, 83.

Parfrey LW, Katz LA (2010) Genome Dynamics Are Influenced by Food Source in Allogromia laticollaris Strain CSH (Foraminifera). *Genome Biology and Evolution*, **2**, 678–685.

Park MG, Yih W, Coats DW (2004) Parasites and Phytoplankton, with Special Emphasis on Dinoflagellate Infections1. *Journal of Eukaryotic Microbiology*, **51**, 145–155.

Patterson RT, Lamoureux EDR, Neville LA, Macumber AL (2013) Arcellacea (testate lobose amoebae) as pH indicators in a pyrite mine-acidified lake, Northeastern Ontario, Canada. *Microbial Ecology*, **65**, 541–554.

Pauls SU, Blahnik RJ, Zhou X, Wardwell CT, Holzenthal RW (2010) DNA barcode data confirm new species and reveal cryptic diversity in Chilean Smicridea (Smicridea) (Trichoptera:Hydropsychidae). *Journal of the North American Benthological Society*, **29**, 1058–1074.

Pawlowski J (2000) Introduction to the molecular systematics of foraminifera. *Micropaleontology*, **46**, 1–12.

Pawlowski J, Audic S, Adl S *et al.* (2012) CBOL protist working group: barcoding eukaryotic richness beyond the animal, plant, and fungal kingdoms. *PLoS biology*, **10**, e1001419.

Pawlowski J, Bolivar I, Fahrni JF, Vargas CD, Bowser SS (1999) Molecular Evidence That Reticulomyxa Filosa Is A Freshwater Naked Foraminifer. *Journal of Eukaryotic Microbiology*, **46**, 612–617.

Pawlowski J, Christen R, Lecroq B *et al.* (2011a) Eukaryotic Richness in the Abyss: Insights from Pyrotag Sequencing. *PLoS ONE*, **6**, e18169.

Pawlowski J, Esling P, Lejzerowicz F *et al.* (2016a) Benthic monitoring of salmon farms in Norway using foraminiferal metabarcoding. *Aquaculture Environment Interactions*, **8**, 371–386.

Pawlowski J, Esling P, Lejzerowicz F, Cedhagen T, Wilding TA (2014a) Environmental monitoring through protist next-generation sequencing metabarcoding: assessing the impact of fish farming on benthic foraminifera communities. *Molecular Ecology Resources*.

Pawlowski J, Fahrni JF, Brykczynska U, Habura A, Bowser SS (2002a) Molecular data reveal high taxonomic diversity of allogromiid Foraminifera in Explorers Cove (McMurdo Sound, Antarctica). *Polar Biology*, **25**, 96–105.

Pawlowski J, Fahrni J, Lecroq B *et al.* (2007) Bipolar gene flow in deep-sea benthic foraminifera. *Molecular Ecology*, **16**, 4089–4096.

Pawlowski J, Fontaine D, da Silva AA, Guiard J (2011b) Novel lineages of Southern Ocean deep-sea foraminifera revealed by environmental DNA sequencing. *Deep Sea Research Part II: Topical Studies in Oceanography*, **58**, 1996–2003.

Pawlowski J, Holzmann M (2002) Molecular phylogeny of Foraminifera a review. *European Journal of Protistology*, **38**, 1–10.

Pawlowski J, Holzmann M (2014) A Plea for Dna Barcoding of Foraminifera. *The Journal of Foraminiferal Research*, **44**, 62–67.

Pawlowski J, Holzmann M, Berney C *et al.* (2002b) PHYLOGENY OF ALLOGROMIID FORAMINIFERA INFERRED FROM SSU rRNA GENE SEQUENCES. *The Journal of Foraminiferal Research*, **32**, 334–343.

Pawlowski J, Holzmann M, Berney C *et al.* (2003) The evolution of early Foraminifera. *Proceedings of the National Academy of Sciences*, **100**, 11494–11498.

Pawlowski J, Holzmann M, Tyszka J (2013) New supraordinal classification of Foraminifera: Molecules meet morphology. *Marine Micropaleontology*, **100**, 1–10.

Pawlowski J, Lecroq B (2010) Short rDNA barcodes for species identification in foraminifera. *The Journal of Eukaryotic Microbiology*, **57**, 197–205.

Pawlowski J, Lee JJ (1991) Taxonomic Notes on Some Tiny, Shallow Water Foraminifera from the Northern Gulf of Elat (Red Sea). *Micropaleontology*, **37**, 149.

Pawlowski J, Lee JJ (1992) The Life Cycle of Rotaliella elatiana N. Sp.: A Tiny Macroalgavorous Foraminifer from the Gulf of Elat1. *The Journal of Protozoology*, **39**, 131–143.

Pawlowski J, Lejzerowicz F, Apotheloz-Perret-Gentil L, Visco J, Esling P (2016b) Protist metabarcoding and environmental biomonitoring: Time for change. *European Journal of Protistology*.

Pawlowski J, Lejzerowicz F, Esling P (2014b) Next-generation environmental diversity surveys of foraminifera: preparing the future. *The Biological Bulletin*, **227**, 93–106.

Pawlowski J, Majewski W (2011) Magnetite-Bearing Foraminifera from Admiralty Bay, West Antarctica, with Description of Psammophaga Magnetica, Sp. Nov. *The Journal of Foraminiferal Research*, **41**, 3–13.

Payne RJ (2013) Seven reasons why protists make useful bioindicators. *Acta Protozoologica*, **3**.

Penard E (1899) *Les rhizopodes de faune profonde dans le lac Léman*. W. Kündig.

Penard E (1902) *Faune rhizopodique du bassin du Léman ...* H. Kündig, Genève.

Penard E (1905) *Les sarcodinés des Grands Lacs*. Genève : Henry Kündig.

Penard E (1907) *Recherches biologiques sur deux Lieberkühnia.* Fischer,.

Pereira TJ, Baldwin JG (2016) Contrasting evolutionary patterns of 28S and ITS rRNA genes reveal high intragenomic variation in Cephalenchus (Nematoda): Implications for species delimitation. *Molecular Phylogenetics and Evolution*, **98**, 244–260.

Pernice MC, Forn I, Gomes A *et al.* (2015) Global abundance of planktonic heterotrophic protists in the deep ocean. *The ISME Journal*, **9**, 782–792.

Phinn SR, Dekker AG, Brando VE, Roelfsema CM (2005) Mapping water quality and substrate cover in optically complex coastal and reef waters: an integrated approach. *Marine Pollution Bulletin*, **51**, 459–469.

Pillet L, Fontaine D, Pawlowski J (2012) Intra-genomic ribosomal RNA polymorphism and morphological variation in Elphidium macellum suggests inter-specific hybridization in foraminifera. *PloS One*, **7**, e32373.

Pilliod DS, Goldberg CS, Arkle RS, Waits LP (2014) Factors influencing detection of eDNA from a stream-dwelling amphibian. *Molecular Ecology Resources*, **14**, 109–116.

Piñol J, Mir G, Gomez-Polo P, Agustí N (2015) Universal and blocking primer mismatches limit the use of high-throughput DNA sequencing for the quantitative metabarcoding of arthropods. *Molecular Ecology Resources*, **15**, 819–830.

Pochon X, Wood SA, Keeley NB *et al.* (2015a) Accurate assessment of the impact of salmon farming on benthic sediment enrichment using foraminiferal metabarcoding. *Marine Pollution Bulletin*.

Pochon X, Zaiko A, Hopkins GA, Banks JC, Wood SA (2015b) Early detection of eukaryotic communities from marine biofilm using high-throughput sequencing: an assessment of different sampling devices. *Biofouling*, **31**, 241–251.

Poikane S, van den Berg M, Hellsten S *et al.* (2011) Lake ecological assessment systems and intercalibration for the European Water Framework Directive: Aims, achievements and further challenges. *Procedia Environmental Sciences*, **9**, 153–168.

Poikane S, Johnson RK, Sandin L *et al.* (2016) Benthic macroinvertebrates in lake ecological assessment: A review of methods, intercalibration and practical recommendations. *The Science of the Total Environment*, **543**, 123–134.

Polidoro BA, Comeros-Raynal MT, Cahill T, Clement C (2017) Land-based sources of marine pollution: Pesticides, PAHs and phthalates in coastal stream water, and heavy metals in coastal stream sediments in American Samoa. *Marine Pollution Bulletin*, **116**, 501–507.

Potapov V, Ong JL (2017) Examining Sources of Error in PCR by Single-Molecule Sequencing. *PLOS ONE*, **12**, e0169774.

Potapova M, Charles DF (2005) Choice of substrate in algae-based water-quality assessment. *Journal of the North American Benthological Society*, **24**, 415–427.

Pothoven SA, Fahnenstiel GL, Vanderploeg HA, Nalepa TF (2016) Changes in water quality variables at a mid-depth site after proliferation of dreissenid mussels in

southeastern Lake Michigan. *Fundamental and Applied Limnology / Archiv f??r Hydrobiologie*, **188**, 233–244.

Poulíčková A, Duchoslav M, Dokulil M (2004) Littoral diatom assemblages as bioindicators of lake trophic status: A case study from perialpine lakes in Austria. *European Journal of Phycology*, **39**, 143–152.

Prokopowich CD, Gregory TR, Crease TJ (2003) The correlation between rDNA copy number and genome size in eukaryotes. *Genome / National Research Council Canada = Génome / Conseil National De Recherches Canada*, **46**, 48–50.

Prosser SWJ, Hebert PDN (2017) Rapid identification of the botanical and entomological sources of honey using DNA metabarcoding. *Food Chemistry*, **214**, 183–191.

Protist Information Server (2016) http://protist.i.hosei.ac.jp/PDB/Images/Sarcodina/Allelogromia/index.html.

Prygiel J, Carpentier P, Almeida S *et al.* (2002) Determination of the biological diatom index (IBD NF T 90–354): results of an intercomparison exercise. *Journal of Applied Phycology*, **14**, 27–39.

R Core Team (2013) R: A language and environment for statistical computing.

Ratnasingham S, Hebert PDN (2007) bold: The Barcode of Life Data System (http://www.barcodinglife.org). *Molecular Ecology Notes*, **7**, 355–364.

Rees HC, Bishop K, Middleditch DJ *et al.* (2014) The application of eDNA for monitoring of the Great Crested Newt in the UK. *Ecology and Evolution*, **4**, 4023–4032.

Reichardt E (1999) *Zur Revision der Gattung Gomphonema: Die Arten um G. affine/insigne, G. angustatum/micropus, G. acuminatum sowie gomphonemoide Diatomeen aus dem Oberoligozän in Böhmen*. A.R.G. Gantner.

Reid T, VanMensel D, Droppo IG, Weisener CG (2016) The symbiotic relationship of sediment and biofilm dynamics at the sediment water interface of oil sands industrial tailings ponds. *Water Research*, **100**, 337–347.

Reiss Z, Hottinger L (1984) *The Gulf of Aqaba: Ecological Micropaleontology*. Berlin.

Ren X, Zhu X, Warndorff M, Bucheli P, Shu Q (2006) DNA extraction and fingerprinting of commercial rice cereal products. *Food Research International*, **39**, 433–439.

Resende DV, Pedrosa AL, Correia D *et al.* (2011) Polymorphisms in the 18S rDNA gene of Cystoisospora belli and clinical features of cystoisosporosis in HIV-infected patients. *Parasitology Research*, **108**, 679–685.

Reuter JA, Spacek D, Snyder MP (2015) High-Throughput Sequencing Technologies. *Molecular cell*, **58**, 586–597.

Revello CG (2015) https://www.youtube.com/watch?v=nUl6JfpJZHw.

Reynoldson TB, Bailey RC, Day KE, Norris RH (1995) Biological guidelines for freshwater sediment based on BEnthic Assessment of SedimenT (the BEAST) using a multivariate approach for predicting biological state. *Australian Journal of Ecology*, **20**, 198–219.

Rhumbler L (1904) Systematische Zusammenstellung der recenten Reticulosa. *Archiv fur Protistenhunde*, 181–294.

Rimet F (2012) Recent views on river pollution and diatoms. *Hydrobiologia*, **683**, 1–24.

Rimet F, Bouchez A (2012) Biomonitoring river diatoms: Implications of taxonomic resolution. *Ecological Indicators*, **15**, 92–99.

Rimet F, Chaumeil P, Keck F *et al.* (2016) R-Syst::diatom: an open-access and curated barcode database for diatoms and freshwater monitoring. *Database: The Journal of Biological Databases and Curation*, **2016**.

Rimet F, Trobajo R, Mann DG *et al.* (2014) When is Sampling Complete? The Effects of Geographical Range and Marker Choice on Perceived Diversity in Nitzschia palea (Bacillariophyta). *Protist*, **165**, 245–259.

Robinson CK, Brotman RM, Ravel J (2016) Intricacies of assessing the human microbiome in epidemiologic studies. *Annals of Epidemiology*, **26**, 311–321.

Roe HM, Patterson RT (2014) Arcellacea (testate amoebae) as bio-indicators of road salt contamination in lakes. *Microbial Ecology*, **68**, 299–313.

Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, **19**, 1572–1574.

Rooney AP (2004) Mechanisms underlying the evolution and maintenance of functionally heterogeneous 18S rRNA genes in Apicomplexans. *Molecular Biology and Evolution*, **21**, 1704–1711.

Rooney AP, Ward TJ (2005) Evolution of a large ribosomal RNA multigene family in filamentous fungi: birth and death of a concerted evolution paradigm. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 5084–5089.

Rosenberg R, Blomqvist M, C Nilsson H, Cederwall H, Dimming A (2004) Marine quality assessment by use of benthic species-abundance distributions: a proposed new protocol within the European Union Water Framework Directive. *Marine Pollution Bulletin*, **49**, 728–739.

Rovira L, Trobajo R, Sato S, Ibáñez C, Mann DG (2015) Genetic and Physiological Diversity in the Diatom Nitzschia inconspicua. *The Journal of Eukaryotic Microbiology*, **62**, 815–832.

Rueckert S, Simdyanov TG, Aleoshin VV, Leander BS (2011) Identification of a divergent environmental DNA sequence clade using the phylogeny of gregarine parasites (Apicomplexa) from crustacean hosts. *PloS One*, **6**, e18163.

Ruse L (2010) Classification of nutrient impact on lakes using the chironomid pupal exuvial technique. *Ecological Indicators*, **10**, 594–601.

Sabbatini A, Pawlowski J, Gooday AJ *et al.* (2004) Vellaria zucchellii sp. nov. a new monothalamous foraminifer from Terra Nova Bay, Antarctica. *Antarctic Science*, **16**, 307–312.

de Saedeleer H (1934) *Beitrag zur kenntnis der rhizopoden: morphologische und systematisch euntersuchungen und ein klassifikationsversuch*. Musée royal d'histoire naturelle de Belgique.

Sagouis A, Jabot F, Argillier C (2016) Taxonomic versus functional diversity metrics: how do fish communities respond to anthropogenic stressors in reservoirs? *Ecology of Freshwater Fish*, n/a-n/a.

Sánchez-Quiles D, Marbà N, Tovar-Sánchez A (2017) Trace metal accumulation in marine macrophytes: Hotspots of coastal contamination worldwide. *Science of The Total Environment*, **576**, 520–527.

Sandin L, Schartau AK, Aroviita J *et al.* (2014) Water Framework Directive Intercalibration Technical Report: Northern Lake Benthic invertebrate ecological assessment methods - EU Science Hub - European Commission. *EU Science Hub*.

Schaumburg J, Schranz C, Foerster J *et al.* (2004) Ecological classification of macrophytes and phytobenthos for rivers in Germany according to the water framework directive. *Limnologica - Ecology and Management of Inland Waters*, **34**, 283–301.

Schirmer M, Ijaz UZ, D'Amore R *et al.* (2015) Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform. *Nucleic Acids Research*, **43**, e37.

Schloss PD, Westcott SL, Ryabin T *et al.* (2009) Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. *Applied and Environmental Microbiology*, **75**, 7537–7541.

Schneider SC, Lindstrøm E-A (2011) The periphyton index of trophic status PIT: a new eutrophication metric based on non-diatomaceous benthic algae in Nordic rivers. *Hydrobiologia*, **665**, 143–155.

Schneider S, Melzer A (2003) The Trophic Index of Macrophytes (TIM) – a New Tool for Indicating the Trophic State of Running Waters. *International Review of Hydrobiology*, **88**, 49–67.

Schoch CL, Seifert KA, Huhndorf S *et al.* (2012) Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, 6241–6246.

Schönfeld J, Alve E, Geslin E *et al.* (2012) The FOBIMO (FOraminiferal BIo-MOnitoring) initiative—Towards a standardised protocol for soft-bottom benthic foraminiferal monitoring studies. *Marine Micropaleontology*, **94–95**, 1–13.

Schwing PT, Romero IC, Brooks GR *et al.* (2015) A Decline in Benthic Foraminifera following the Deepwater Horizon Event in the Northeastern Gulf of Mexico. *PLoS ONE*, **10**.

Seenivasan R, Sausen N, Medlin LK, Melkonian M (2013) Picomonas judraskeda gen. et sp. nov.: the first identified member of the Picozoa phylum nov., a widespread group of picoeukaryotes, formerly known as "picobiliphytes." *PloS One*, **8**, e59565.

Shalchian-Tabrizi K, Bråte J, Logares R *et al.* (2008) Diversification of unicellular eukaryotes: cryptomonad colonizations of marine and fresh waters inferred from revised 18S rRNA phylogeny. *Environmental Microbiology*, **10**, 2635–2644.

Shimomura O, Johnson FH, Saiga Y (1962) Extraction, Purification and Properties of Aequorin, a Bioluminescent Protein from the Luminous Hydromedusan, Aequorea. *Journal of Cellular and Comparative Physiology*, **59**, 223–239.

Shokralla S, Hellberg RS, Handy SM, King I, Hajibabaei M (2015) A DNA Mini-Barcoding System for Authentication of Processed Fish Products. *Scientific Reports*, **5**, 15894.

Siemensma FJ (1982) Schaalamoeben. *Natura*, **KNNV Hoogwoud**, 95–106.

Sigsgaard EE, Carl H, Møller PR, Thomsen PF (2015) Monitoring the near-extinct European weather loach in Denmark based on environmental DNA from water samples. *Biological Conservation*, **183**, 46–52.

Simboura N, Zenetos A (2002) Benthic indicators to use in Ecological Quality classification of Mediterranean soft bottom marine ecosystems, including a new Biotic Index. *Mediterranean Marine Science*, **3**, 77–111.

Simon M, López-García P, Deschamps P *et al.* (2015) Marked seasonality and high spatial variability of protist communities in shallow freshwater systems. *The ISME journal*, **9**, 1941–1953.

Simon UK, Weiß M (2008) Intragenomic Variation of Fungal Ribosomal Genes Is Higher than Previously Thought. *Molecular Biology and Evolution*, **25**, 2251–2254.

Sinniger F, Lecroq B, Majewski W (2008) *Bowseria arctowskii gen. et sp. nov., new monothalamous foraminiferan from the Southern Ocean. Polish Polar Research 29, 5-15. Skowronski, R.S.P. Distribuiçâo espacial e variaçâo temporal da meiofauna, com ênfase para o grupo Nematoda, na enseada Martel (*.

de Smet WH (2006) Rotifers inhabiting shells of testate amoebae. *Book of Abstracts, International Symposium on Testate Amoebae*, 44–45.

Smith MJ, Kay WR, Edward DHD *et al.* (1999) AusRivAS: using macroinvertebrates to assess ecological condition of rivers in Western Australia. *Freshwater Biology*, **41**, 269–282.

Smith KF, Kohli GS, Murray SA, Rhodes LL (2017) Assessment of the metabarcoding approach for community analysis of benthic-epiphytic dinoflagellates using mock communities. *New Zealand Journal of Marine and Freshwater Research*, **0**, 1–22.

Smith MB, Rocha AM, Smillie CS *et al.* (2015) Natural bacterial communities serve as quantitative geochemical biosensors. *mBio*, **6**, e00326-315.

Smith MA, Rodriguez JJ, Whitfield JB *et al.* (2008) Extreme diversity of tropical parasitoid wasps exposed by iterative integration of natural history, DNA barcoding, morphology, and collections. *Proceedings of the National Academy of Sciences*, **105**, 12359–12364.

Sola A, Serrano S, Guinea A (1996) Influence of environmental characteristics on the distribution of ciliates in the River Henares (Central Spain). *Hydrobiologia*, **324**, 237–252.

Stamatakis A (2014) RAxML Version 8: A tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies. *Bioinformatics*, btu033.

Stamatakis A, Hoover P, Rougemont J (2008) A Rapid Bootstrap Algorithm for the RAxML Web Servers. *Systematic Biology*, **57**, 758–771.

Stein ED, Martinez MC, Stiles S, Miller PE, Zakharov EV (2014) Is DNA barcoding actually cheaper and faster than traditional morphological methods: results from a survey of freshwater bioassessment efforts in the United States? *PloS One*, **9**, e95525.

Steinberger RE, Holden PA (2005) Extracellular DNA in Single- and Multiple-Species Unsaturated Biofilms. *Applied and Environmental Microbiology*, **71**, 5404–5410.

Stevenson RJ, Pan YD, Dam H van (2010) Assessing environmental conditions in rivers and streams with diatoms. , 57–85.

Stoeck T, Breiner H-W, Filker S *et al.* (2014) A morphogenetic survey on ciliate plankton from a mountain lake pinpoints the necessity of lineage-specific barcode markers in microbial ecology. *Environmental Microbiology*, **16**, 430–444.

Stubbington R, Wood PJ, Reid I (2011) Spatial variability in the hyporheic zone refugium of temporary streams. *Aquatic Sciences*, **73**, 499–511.

Stucki P (2010) Methoden zur Untersuchung und Beurteilung der Fliessgewässer. Makrozoobenthos Stufe F. Bundesamt für Umwelt. *Etat de l'environnement n° 1026. Office fédéral de l'environnement, Berne*.

Svobodová J, Douda K, Fischer D, Lapšanská N, Vlach P (2017) Toxic and heavy metals as a cause of crayfish mass mortality from acidified headwater streams. *Ecotoxicology*, **26**, 261–270.

Swiss Federal Council (1998) *Waters Protection Ordinance, https://www.admin.ch/opc/en/classified-compilation/19983281/index.html*.

Taberlet P, Coissac E, Pompanon F, Brochmann C, Willerslev E (2012) Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology*, **21**, 2045–2050.

Tamura K, Peterson D, Peterson N *et al.* (2011) MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Molecular Biology and Evolution*, **28**, 2731–2739.

Tang YZ, Dobbs FC (2007) Green Autofluorescence in Dinoflagellates, Diatoms, and Other Microalgae and Its Implications for Vital Staining and Morphological Studies. *Applied and Environmental Microbiology*, **73**, 2306–2313.

Tarkowska-Kukuryk M, Mieczan T (2017) Submerged macrophytes as bioindicators of environmental conditions in shallow lakes in eastern Poland. *Annales de Limnologie - International Journal of Limnology*, **53**, 27–34.

Techen N, Parveen I, Pan Z, Khan IA (2014) DNA barcoding of medicinal plant material for identification. *Current Opinion in Biotechnology*, **25**, 103–110.

Teklu BM, Hailu A, Wiegant DA, Scholten BS, Brink PJV den (2016) Impacts of nutrients and pesticides from small- and large-scale agriculture on the water quality of Lake Ziway, Ethiopia. *Environmental Science and Pollution Research*, 1–10.

Teta R, Romano V, Sala GD *et al.* (2017) Cyanobacteria as indicators of water quality in Campania coasts, Italy: a monitoring strategy combining remote/proximal sensing and in situ data. *Environmental Research Letters*, **12**, 24001.

Thiéry O, Vasar M, Jairus T *et al.* (2016) Sequence variation in nuclear ribosomal small subunit, internal transcribed spacer and large subunit regions of Rhizophagus irregularis and Gigaspora margarita is high and isolate-dependent. *Molecular Ecology*, **25**, 2816–2832.

Thomas R (1961) Note sur quelques Rhizopodes de France. *Cahiers des Naturalistes, Bulletin des Naturalistes Parisiens*, **17**, 74–80.

Tindall KR, Kunkel TA (1988) Fidelity of DNA synthesis by the Thermus aquaticus DNA polymerase. *Biochemistry*, **27**, 6008–6013.

Tolotti M, Thies H, Cantonati M, Hansen CME, Thaler B (2003) Flagellate algae (Chrysophyceae, Dinophyceae, Cryptophyceae) in 48 high mountain lakes of the Northern and Southern slope of the Eastern Alps: biodiversity, taxa distribution and their driving variables. In: *Phytoplankton and Equilibrium Concept: The Ecology of Steady-State Assemblages*, pp. 331–348. Springer, Dordrecht.

Torti A, Lever MA, Jørgensen BB (2015) Origin, dynamics, and implications of extracellular DNA pools in marine sediments. *Marine Genomics*, **24 Pt 3**, 185–196.

Tsai YL, Olson BH (1992) Rapid method for separation of bacterial DNA from humic substances in sediments for polymerase chain reaction. *Applied and Environmental Microbiology*, **58**, 2292–2295.

Tsuchiya M, Gooday AJ, Nomaki H, Oguri K, Kitazato H (2013) Genetic Diversity and Environmental Preferences of Monothalamous Foraminifers Revealed through Clone Analysis of Environmental Small-Subunit Ribosomal DNA Sequences. *The Journal of Foraminiferal Research*, **43**, 3–13.

Turner TE, Swindles GT (2012) Ecology of Testate Amoebae in Moorland with a Complex Fire History: Implications for Ecosystem Monitoring and Sustainable Land Management. *Protist*, **163**, 844–855.

Turner CR, Uy KL, Everhart RC (2015) Fish environmental DNA is more concentrated in aquatic sediments than surface water. *Biological Conservation*, **183**, 93–102.

Urbanič G (2014) A Littoral Fauna Index for assessing the impact of lakeshore alterations in Alpine lakes. *Ecohydrology*, **7**, 703–716.

Valentine J, Davis SR, Kirby JR, Wilkinson DM (2013) The use of testate amoebae in monitoring peatland restoration management: case studies from North West England and Ireland. *Acta Protozoologica*, **3**.

Valentini A, Taberlet P, Miaud C *et al.* (2015) Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. *Molecular Ecology*.

Van den Wyngaert S, Möst M, Freimann R, Ibelings BW, Spaak P (2015) Hidden diversity in the freshwater planktonic diatom Asterionella formosa. *Molecular Ecology*, **24**, 2955–2972.

de Vargas C, Audic S, Henry N *et al.* (2015) Eukaryotic plankton diversity in the sunlit ocean. *Science*, **348**, 1261605.

Vartak VR, Narasimmalu R, Annam PK, Singh DP, Lakra WS (2015) DNA barcoding detected improper labelling and supersession of crab food served by restaurants in India. *Journal of the Science of Food and Agriculture*, **95**, 359–366.

Vasselon V, Domaizon I, Rimet F, Kahlert M, Bouchez A (2017) Application of high-throughput sequencing (HTS) metabarcoding to diatom biomonitoring: Do DNA extraction methods matter? *Freshwater Science*, **36**, 162–177.

Veach AM, Dodds WK, Jumpponen A (2015) Woody plant encroachment, and its removal, impact bacterial and fungal communities across stream and terrestrial habitats in a tallgrass prairie ecosystem. *FEMS microbiology ecology*, **91**.

Vermeij GJ, Dudley R (2000) Why are there so few evolutionary transitions between aquatic and terrestrial ecosystems? *Biological Journal of the Linnean Society*, **70**, 541–554.

Verneaux V, Verneaux J, Schmitt A, Lovy C, Lambert JC (2004) The Lake Biotic Index (LBI): an applied method for assessing the biological quality of lakes using macrobenthos; the Lake Châlain (French Jura) as an example. *Annales de Limnologie - International Journal of Limnology*, **40**, 1–9.

Vidovic J, Cosovic V, Juracic M, Petricioli D (2009) Impact of fish farming on foraminiferal community, Drvenik Veliki Island, Adriatic Sea, Croatia. *Marine Pollution Bulletin*.

Vidovic J, Dolenec M, Dolenec T, Karamarko V, Žvab Rožič P (2014) Benthic foraminifera assemblages as elemental pollution bioindicator in marine sediments around fish farm (Vrgada Island, Central Adriatic, Croatia). *Marine Pollution Bulletin*, **83**, 198–213.

Vilain S, Pretorius JM, Theron J, Brözel VS (2009) DNA as an Adhesin: Bacillus cereus Requires Extracellular DNA To Form Biofilms. *Applied and Environmental Microbiology*, **75**, 2861–2868.

Visco J, Apothéloz-Perret-Gentil L, Cordonier A *et al.* (2015) Environmental Monitoring: Inferring the Diatom Index from Next-Generation Sequencing Data. *Environmental Science & Technology*, **49**, 7597–7605.

Vivien R, Ferrari BJD, Pawlowski J (2016a) DNA barcoding of formalin-fixed aquatic oligochaetes for biomonitoring. *BMC research notes*, **9**, 342.

Vivien R, Lejzerowicz F, Pawlowski J (2016b) Next-Generation Sequencing of Aquatic Oligochaetes: Comparison of Experimental Communities. *PLOS ONE*, **11**, e0148644.

Voltski I, Korsun S, Pawlowski J (2014) Toxisarcon taimyr sp. nov., a new large monothalamous foraminifer from the Kara Sea inner shelf. *Marine Biodiversity*, **44**, 213–221.

Wailes J (1915) *The British Freshwater Rhizopoda and Heliozoa*. Printed for the Ray Society.

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**, 323–330.

Weber AA-T, Pawlowski J (2013) Can Abundance of Protists Be Inferred from Sequence Data: A Case Study of Foraminifera (P López-García, Ed,). *PLoS ONE*, **8**, e56739.

Weber AA-T, Pawlowski J (2014) Wide Occurrence of SSU rDNA Intragenomic Polymorphism in Foraminifera and its Implications for Molecular Species Identification. *Protist*, **165**, 645–661.

Weller C, Wu M (2015) A generation-time effect on the rate of molecular evolution in bacteria. *Evolution; International Journal of Organic Evolution*, **69**, 643–652.

Whitchurch CB, Tolker-Nielsen T, Ragas PC, Mattick JS (2002a) Extracellular DNA Required for Bacterial Biofilm Formation. *Science*, **295**, 1487–1487.

Whitchurch CB, Tolker-Nielsen T, Ragas PC, Mattick JS (2002b) Extracellular DNA required for bacterial biofilm formation. *Science (New York, N.Y.)*, **295**, 1487.

Whitton BA, Rott E, Friedrich G (1991) Use of algae for monitoring rivers. *Journal of Applied Phycology*, **3**, 287–287.

Wilcox TM, Schwartz MK, McKelvey KS, Young MK, Lowe WH (2014) A blocking primer increases specificity in environmental DNA detection of bull trout (Salvelinus confluentus). *Conservation Genetics Resources*, **6**, 283–284.

Wood SA, Smith KF, Banks JC *et al.* (2013) Molecular genetic tools for environmental monitoring of New Zealand's aquatic habitats, past, present and the future. *New Zealand Journal of Marine and Freshwater Research*, **47**, 90–119.

Woodward G, Gray C, Baird DJ (2013) Biomonitoring for the 21st century: New perspectives in an age of globalisation and emerging environmental threats. *Limnetica*, **32**, 159–174.

Word JQ (1979) The Infaunal Trophic Index. *Annual Report 1978*, 19–39.

Wright JF, Sutcliffe DW, Furse MT (2000) Assessing the biological quality of freshwaters. *RIVPACS and other techniques. Freshwater Biological Association, Ambleside, England*.

Wu Z-W, Wang Q-M, Liu X-Z, Bai F-Y (2016) Intragenomic polymorphism and intergenomic recombination in the ribosomal RNA genes of strains belonging to a yeast species Pichia membranifaciens. *Mycology*, **7**, 102–111.

Wylezich C, Kaufmann D, Marcuse M, Hülsmann N (2014) Dracomyxa pallida gen. et sp. nov.: A New Giant Freshwater Foraminifer, with Remarks on the Taxon Reticulomyxidae (emend.). *Protist*, **165**, 854–869.

Xu H, Zhang W, Jiang Y, Yang EJ (2014) Use of biofilm-dwelling ciliate communities to determine environmental quality status of coastal waters. *The Science of the Total Environment*, **470**–**471**, 511–518.

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**, 321–329.

Yerubandi RR, Boegman L, Bolkhari H, Hiriart-Baer V (2016) Physical processes affecting water quality in Hamilton Harbour. *Aquatic Ecosystem Health & Management*, **19**, 114–123.

Yilmaz P, Parfrey LW, Yarza P *et al.* (2014) The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. *Nucleic Acids Research*, **42**, D643–D648.

Yoon T-H, Kang H-E, Kang C-K *et al.* (2016) Development of a cost-effective metabarcoding strategy for analysis of the marine phytoplankton community. *PeerJ*, **4**, e2115.

Yu DW, Ji Y, Emerson BC *et al.* (2012) Biodiversity soup: metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. *Methods in Ecology and Evolution*, **3**, 613–623.

Zaets IY, Podolich OV, Reva ON, Kozyrovska NO (2016) DNA metabarcoding of microbial communities for healthcare. *Biopolymers and Cell*, **32**, 3–8.

Zaiko A, Martinez JL, Ardura A *et al.* (2015) Detecting nuisance species using NGST: Methodology shortcomings and possible application in ballast water monitoring. *Marine Environmental Research*, **112**, **Part B**, 64–72.

Zelinka M, Marvan P (1961) *Zur Präzisierung der biologischen Klassifikation der Reinheit fliessender Gewässer*.

Zhou X, Li Y, Liu S *et al.* (2013) Ultra-deep sequencing enables high-fidelity recovery of biodiversity for bulk arthropod samples without PCR amplification. *GigaScience*, **2**, 4.

Zhou J, Wu L, Deng Y *et al.* (2011) Reproducibility and quantitation of amplicon sequencing-based detection. *The ISME journal*, **5**, 1303–1313.

Zhu F, Massana R, Not F, Marie D, Vaulot D (2005) Mapping of picoeucaryotes in marine ecosystems with quantitative PCR of the 18S rRNA gene. *FEMS microbiology ecology*, **52**, 79–92.

Zimmermann J, Abarca N, Enke N *et al.* (2014) Taxonomic reference libraries for environmental barcoding: a best practice example from diatom research. *PloS One*, **9**, e108793.

Zimmermann J, Glöckner G, Jahn R, Enke N, Gemeinholzer B (2015) Metabarcoding vs. morphological identification to assess diatom diversity in environmental studies. *Molecular Ecology Resources*, **15**, 526–542.

Zimmermann J, Jahn R, Gemeinholzer B (2011) Barcoding diatoms: evaluation of the V4 subregion on the 18S rRNA gene, including new primers and protocols. *Organisms Diversity & Evolution*, **11**, 173–192.

Zlatogursky VV, Kudryavtsev A, Udalov IA *et al.* (2016) Genetic structure of a morphological species within the amoeba genus Korotnevella (Amoebozoa: Discosea), revealed by the analysis of two genes. *European Journal of Protistology*, **56**, 102–111.

Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research*, **31**, 3406–3415.

ANNEXES

CERTIFICATE OF THE DOCTORAL PROGRAM IN ECOLOGY & EVOLUTION

PRESS RELEASE ABOUT THE CHAPTER 8 - TAXONOMY-FREE MOLECULAR DIATOM INDEX FOR HIGH-THROUGHPUT EDNA BIOMONITORING

JOURNAL-FORMATTED COPIES OF THE PUBLISHED CHAPTERS

CUSO CONFÉRENCE UNIVERSITAIRE DE SUISSE OCCIDENTALE

Laure APOTHÉLOZ-PERRET-GENTIL

(University of Geneva)

completed the

CUSO Doctoral Program in Ecology and Evolution

Courses	Credits
An Introduction to R, 17-20th October 2011, University of Lausanne (CH) *Doctoral program in Population Genomics* *CUSO doctoral program in Ecology & Evolution*	2
Computer Skill for Biological Research, 22-26th October 2012, University of Geneva (CH) *CUSO doctoral program in Ecology & Evolution*	2
Phylogeny and Molecular Evolution, 11-15th February 2013, University of Geneva (CH) *CUSO doctoral program in Ecology & Evolution*	3

Workshops	Credits
Fundamental and Applied Protistology, 15-16th April 2014, University of Neuchâtel (CH) *Doctoral program in Organismal Biology*	1.5
Eukaryotic –Omics: Exploring and testing with next-generation sequencing, 24-25th April 2014, University of Geneva (CH) *CUSO doctoral program in Ecology & Evolution*	1

Congresses & Symposia	
VI European Congress of Protistology, 25-29th July 2011, Berlin (DE)	1
DGP 2013, the 32nd Meeting of the German Society for Protozoology, 27th February – 2nd March 2013, Kartause Ittigen (CH)	1
Biology14, the Swiss Conference on Organismal Biology, 13-14th February 2014, University of Geneva (CH)	0.5
Biology15, the Swiss Conference on Organismal Biology, 13-14th February 2015, Dübendorf (CH)	1
SEFS9 2015, 9th Symposium for European Freshwater Sciences, 5-10th July 2015, University of Geneva (CH)	1.25

UNIVERSITÄT BERN · UNIVERSITAS FRIBURGENSIS · UNIVERSITÉ DE GENÈVE · UNIL | Université de Lausanne · Université de Neuchâtel · unine

CUSO CONFÉRENCE UNIVERSITAIRE DE SUISSE OCCIDENTALE

Presentation in another CUSO university

Talk at 2016 general assembly of SwissBOL: "Environmental monitoring: inferring swiss diatom index from NGS data", 1
19th May 2016, Natural History Museum of Neuchâtel (CH)

Total ECTS 15.25

Lausanne, 23rd January 2017

Prof. Ian Sanders
Head of the CUSO
Doctoral Program in Ecology and Evolution
University of Lausanne
CH-1015 Lausanne

UNIVERSITÄT BERN · UNIVERSITAS FRIBURGENSIS · UNIVERSITÉ DE GENÈVE · UNIL | Université de Lausanne · Université de Neuchâtel unine

UNIVERSITÉ DE GENÈVE

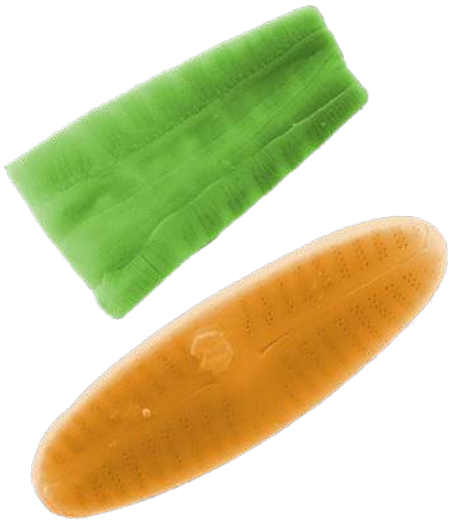
COMMUNIQUÉ DE PRESSE

Genève | 6 avril 2017

ATTENTION: sous embargo jusqu'au 13 avril 2017, 7h30, heure locale

Traquer la pollution grâce à l'ADN des algues

Un outil révolutionnaire permet de traiter des échantillons prélevés en rivière en un temps réduit et à moindre coût.



Diatomées comprises entre 0.01 et 0.02 mm, constituées d'une seule cellule entourée d'un squelette de silice coloré artificiellement. L'algue en vert est présente dans les milieux propres, tandis que celle en orange vit dans de l'eau plus polluée. © Laure Apothéloz-Perret-Gentil, UNIGE.

Les diatomées, un groupe diversifié d'algues unicellulaires, sont particulièrement sensibles aux changements qui affectent leur milieu aquatique. C'est pourquoi elles sont utilisées comme bio-indicateurs pour le suivi biologique de la qualité des eaux. Mais leur identification au microscope à partir des échantillons prélevés en rivière requiert beaucoup de temps et des compétences pointues. Des biologistes de l'Université de Genève (UNIGE) sont parvenus à établir un indice de la qualité de l'eau basé uniquement sur les séquences d'ADN des diatomées présentes dans les échantillons, sans qu'il soit nécessaire d'en identifier visuellement chaque espèce. Cette étude, publiée dans la revue *Molecular Ecology Resources*, présente un outil révolutionnaire permettant de traiter un très grand nombre d'échantillons à la fois, avec une couverture plus étendue du réseau de surveillance en un temps réduit et à moindre coût.

Le degré de pollution des cours d'eau résultant des activités humaines est évalué à l'aide de différents indices biotiques. Ceux-ci reflètent la quantité et la diversité, dans un échantillon prélevé en rivière, d'organismes choisis comme bio-indicateurs en raison de leurs préférences écologiques et de leur tolérance à la pollution. C'est le cas des diatomées, des algues constituées d'une cellule unique entourée d'un squelette de silice, que l'Union Européenne et la Suisse recommandent comme l'un des bio-indicateurs idéaux pour les cours d'eau.

La qualité de nos rivières est déterminée à l'aide de l'indice suisse des diatomées (DI-CH), dont la valeur définit le statut écologique. «L'identification morphologique des différentes espèces présentes dans chaque échantillon ne répond toutefois plus aux directives actuelles qui renforcent les mesures de protection des milieux aquatiques. C'est pourquoi nous avons tenté de mettre au point une nouvelle méthode», explique Jan Pawlowski, professeur au Département de génétique et évolution de la Faculté des sciences de l'UNIGE.

Des séquences d'ADN bio-indicatrices

En collaboration avec le Service de l'écologie de l'eau (SECOE) de Genève et le bureau PhycoEco de La Chaux-de-Fonds, les chercheurs ont analysé les quelque 90 prélèvements qu'ils ont effectués dans différentes rivières en Suisse et déterminé leur statut écologique à l'aide du DI-CH. Ils ont ainsi établi un système de référence, en vue de valider l'indice moléculaire en développement. Ce dernier est basé sur les séquences d'ADN caractéristiques de toutes les espèces de diatomées pouvant être présentes dans ces échantillons.

«L'ensemble des séquences d'ADN révélées dans chaque échantillon correspond à un indice de qualité DI-CH spécifique. Par ailleurs, chaque séquence identifiée a une répartition différente et est détectée en quantités variables d'un prélèvement à l'autre. En intégrant l'ensemble de ces données, nous avons pu calculer une valeur écologique pour chaque séquence, sans devoir identifier l'espèce qui lui correspond», détaille Laure Apothéloz-Perret-Gentil, membre du groupe genevois et première auteure de l'étude.

Un indice moléculaire à l'écoute de l'environnement

Cette approche permet de déterminer la qualité de l'eau en utilisant l'ensemble de ces valeurs écologiques. «Notre évaluation était correcte pour près de 80% des prélèvements, ce qui est très encourageant. L'augmentation du nombre et de la diversité des échantillons permettra de calibrer notre méthode en vue d'effectuer des analyses de routine à grande échelle», note Jan Pawlowski.

Le traitement synchrone de très nombreux prélèvements en un temps record et à coût réduit n'est pas le seul avantage de ce nouvel outil. L'indice moléculaire mis au point par les biologistes de l'UNIGE pourrait en effet facilement être adapté à d'autres groupes de bioindicateurs unicellulaires : un atout de taille pour la surveillance de différents écosystèmes aquatiques.

contact

Jan Pawlowski
+41 22 379 30 69
Jan.Pawlowski@unige.ch

UNIVERSITÉ DE GENÈVE
Service de communication
24 rue du Général-Dufour
CH-1211 Genève 4

Tél. +41 22 379 77 17

media@unige.ch
www.unige.ch

Available online at www.sciencedirect.com

SciVerse ScienceDirect

European Journal of Protistology 49 (2013) 210–216

European Journal of PROTISTOLOGY

www.elsevier.com/locate/ejop

Arnoldiellina fluorescens gen. et sp. nov. – A new green autofluorescent foraminifer from the Gulf of Eilat (Israel)

Laure Apothéloz-Perret-Gentil, Maria Holzmann, Jan Pawlowski*

Department of Genetics and Evolution, University of Geneva, 1211 Geneva 4, Switzerland

Received 23 March 2012; received in revised form 13 August 2012; accepted 13 August 2012
Available online 19 September 2012

Abstract

A new monothalamous (single-chambered) soft-walled foraminiferal species, *Arnoldiellina fluorescens* gen. et sp. nov., was isolated from samples collected in the Gulf of Eilat, Israel. The species is characterized by a small elongate organic theca with a single aperture of allogromiids. It is characterized by the emission of green autofluorescence (GAF) that has so far not been reported from foraminifera. Phylogenetic analysis of a fragment of the 18S rDNA indicates that the species is related to a group of monothalamous foraminiferans classified as clade I. Although the morphology of the new species is very different compared to the other members of this clade, a specific helix in 18S rRNA secondary structure strongly supports this position.
© 2012 Elsevier GmbH. All rights reserved.

Keywords: Foraminifera; Allogromiids; 18S rDNA; Green autofluorescence; *Arnoldiellina fluorescens*

Introduction

Foraminifera are a large and diverse group of protists well known from marine environments (Murray 2006) but also found in terrestrial and freshwater habitats (Meisterfeld et al. 2001; Lejzerowicz et al. 2010). Most foraminiferans produce either 'single chambered' (monothalamous) or 'multi-chambered' (polythalamous) tests, with organic, agglutinated or calcareous walls while some of them lack a test at all (athalamids). Foraminiferal research has focused largely on polythalamous calcareous species, whose hard-walled shells are well preserved in the fossil record (Haynes 1981; Murray 2006). The diversity of soft-walled monothalamous foraminifera, also called allogromiids, remains largely unknown as they are poorly preserved. The interest in this group increased recently, due to their abundance in the deep-sea and polar regions (Gooday 2002; Gooday et al. 2005) and their application in genomic studies (Habura et al. 2005; Parfrey and Katz 2010). Many new monothalamous species have been described in the last decade (Altin et al. 2009; Gooday et al. 2004, 2010; Gooday and Pawlowski 2004; Pawlowski and Majewski 2011; Sabbatini et al. 2004).

The study of monothalamous foraminiferans was also prompted by the development of molecular systematics, which greatly facilitated the identification of their morphologically rather featureless tests. Molecular studies completely changed our view of their phylogenetic relationships and led to the discovery of a huge diversity in this group (Pawlowski et al. 2002a,b, 2003). A new dimension of monothalamiid diversity was revealed by environmental DNA surveys of foraminiferal assemblages (Habura et al. 2004, 2008; Pawlowski et al. 2011; Lecroq et al. 2011).

The new monothalamid species described here was discovered in a culture dish containing sediment and algal debris from the Gulf of Eilat (Israel). Molecular analysis of three specimens showed that they all belong to the same species that is genetically well distinguished from other monothalamids, resulting in a description of a new species and new genus.

*Corresponding author.
E-mail address: Jan.Pawlowski@unige.ch (J. Pawlowski).

0932-4739/$ – see front matter © 2012 Elsevier GmbH. All rights reserved.
http://dx.doi.org/10.1016/j.ejop.2012.08.005

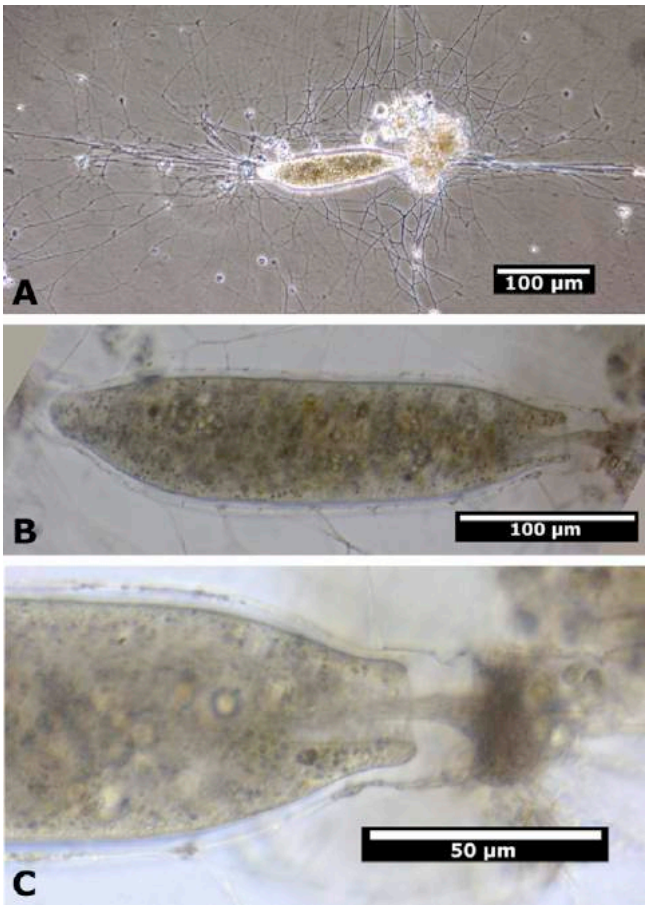


Fig. 1. Living specimens of *Arnoldiellina fluorescens*, gen. and sp. nov. Overview of granuloreticulopodial network (A). View of a specimen (B) with close up of the terminal aperture (C). Pictures were taken with differential interference contrast.

Material and Methods

Isolation and culture

Specimens were isolated from surface sediment samples collected by SCUBA diving at 5–10 m in front of the Inter-university Institute for Marine Sciences (IUI), near Eilat, Israel, on January 2011.

The sediment was distributed in two Petri dishes and cultured in Erdschreiber medium (5% soil extract, 1 mM $NaNO_3$, 0.07 mM Na_2HPO_4, 10 mM Tris, pH = 8, filled up with sterile seawater) and filtered seawater. A few drops of heat killed *Dunaliella salina* (Chlorophyceae) were added for nutrition every two weeks. Specimens with extended pseudopodia were first observed in culture dishes 6 months after collection. The specimens were abundant during a period of 3 months, but later disappeared from the dish and have not been observed again.

Fixation and colouration

Cultured specimens were transferred by means of a pipette to a 10% formalin solution. They were fixed for 1 h at room temperature and afterwards washed briefly in PBS (Phosphate Buffered Saline, 137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, and 2 mM KH_2PO_4, adjusted pH to 7.4). A final immersion lasting 30 min was carried out in a dark room at ambient temperature using 4′,6′-diamidino-2-phénylindole (DAPI) at 5.10E−4 mg/ml to stain and subsequently identify nuclei.

Morphological studies

Living and fixed specimens were observed with an inverted microscope (Nikon Eclipse Ti) and a fluorescence microscope (Nikon Eclipse E200). Photographs were taken with Leica DFC 420C and Nikon Digital DXM 1200 cameras. Videos were made with the Imaging Source DFK 41AF02 camera. They are available in the online version of this article and at http://forambarcoding.unige.ch/movies.

Molecular analyses

DNA from 13 specimens was extracted in guanidine lysis buffer (Pawlowski 2000), each extraction was performed with a single specimen. PCR amplifications of a fragment of the 18S rDNA were performed using the primer pair s14F3 (5′ACG CA(AC) GTG TGA AAC TTG) and 20R (5′GAC GGG CGG TGT GTA CAA). PCR products were re-amplified using the nested primer s14F1 (5′AAG GGC ACC ACA AGA ACG C) and 20R. PCR amplifications for a shorter fragment of the 18S rDNA were performed using the primer pair s14F3 and s17 (5′CGG TCA CGT TCG TTG C). PCR products were re-amplified using the nested primer s14F1 and s17. The amplified PCR products were purified using High Pure PCR Purification Kit (Roche Diagnostics) and sequenced directly. Sequencing reactions were performed using the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) and analysed on a 3130XL Genetic Analyser (Applied Biosystems). The 3 new sequences reported in this paper were deposited in the EMBL/GenBank data base (accession numbers HE775247–HE775249). The secondary structure was created using the RNAfold program from the University of Vienna (Gruber et al. 2008).

The obtained sequences were aligned to 57 other foraminiferans using Seaview v. 4.3.3. software (Gouy et al. 2010). After elimination of the highly variable regions, 869 sites were left for analysis. The phylogenetic tree was constructed using maximum likelihood method based on the GTR + G model, using RAxML BlackBox (Stamatakis et al. 2008).

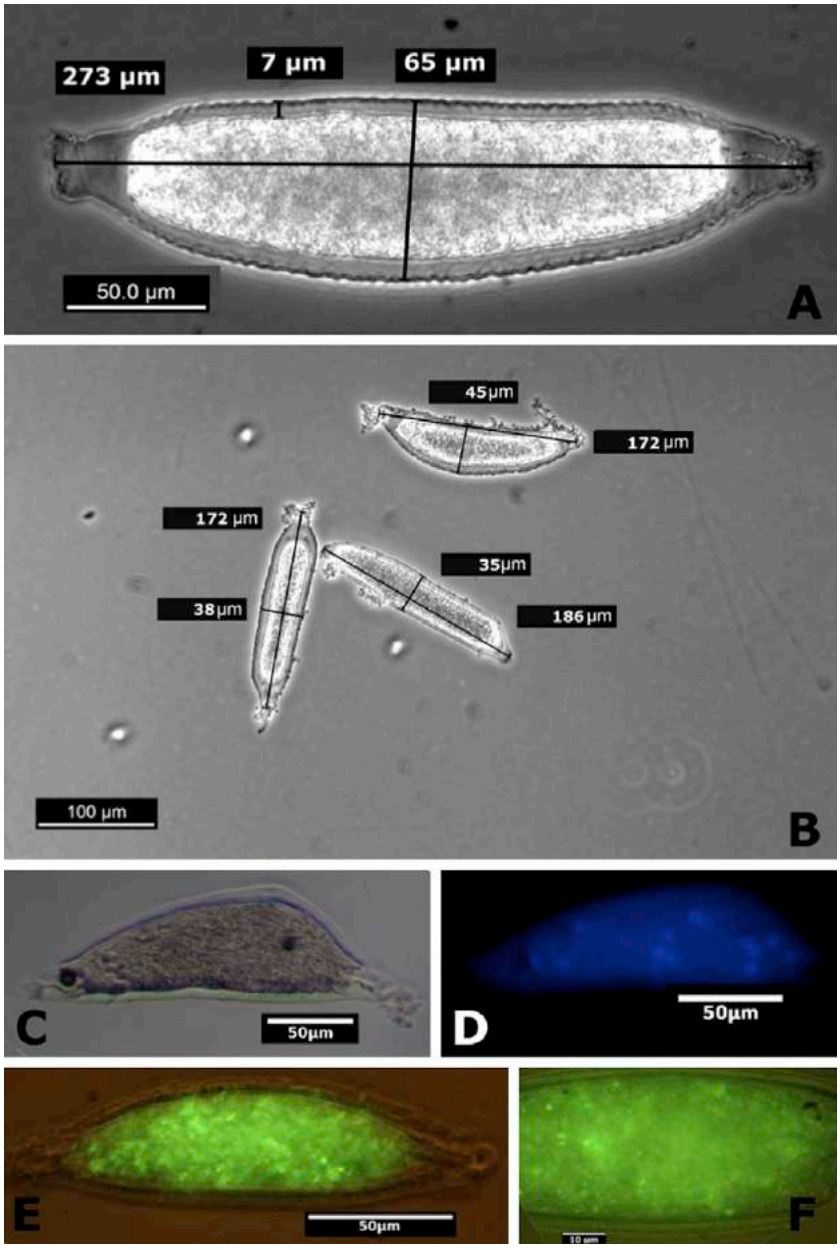


Fig. 2. Fixed and stained specimens of *Arnoldiellina fluorescens* gen. and sp. nov. Fixed holotype (A) and paratypes (B) indicating their respective size. Specimen stained with DAPI (blue) viewed with differential interference contrast (C) or UV light excitation (D). With DAPI staining, the multiple nuclei show up as light-blue coloured rounded spots. Living specimen showing the green autofluorescence viewed with differential interference contrast (E) and UV light excitation (460–500 nm) (F). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of the article.)

Results

Systematics

Supergroup RHIZARIA Cavalier-Smith, 2002
Phylum FORAMINIFERA D'Orbigny, 1826
Genus *Arnoldiellina* gen. nov.

Type species: *Arnoldiellina fluorescens* sp. nov.

Etymology: The genus was named in honour of Zach Arnold, Professor Emeritus of Palaeontology at the University of California, Berkeley who described several monothalamous foraminiferans and studied their life cycles and evolution.

Diagnosis: Test free, monothalamous, fusiform, <300 µm in length and <70 µm in width; organic wall transparent from 2 to 7 µm in width, thicker around the aperture. The single aperture is funnel-shape with a tubular internal extension. Multinucleate cytoplasm (up to 11 nuclei); granular, in constant rapid movement. Reticulopodes very active with rapidly forming reticulopodial network and fast moving granules. Specimens emit GAF, which disappeared with fixation.

Table 1. Measurements of 16 specimens of *Arnoldiellina fluorescens*.

#	Length (μm)	Width (μm)	Ratio length/width	Remarks
1	273	65	4.2	Holotype: Fig. 2A
2	172	44	3.9	Paratype: Fig. 2B
3	172	37	4.6	Paratype: Fig. 2B
4	186	35	5.3	Paratype: Fig. 2B
5	167	54	3.1	
6	160	38	4.2	
7	173	35	4.9	
8	154	36	4.3	
9	152	34	4.5	
10	205	70	2.9	Fig. 2(C, D)
11	166	47	3.5	Fig. 2(E, F)
12	170	38	4.5	Fig. 1A
13	300	68	4.4	Fig. 1B
14	253	69	3.7	
15	244	63	3.9	
16	174	54	3.2	

Remarks: The new genus was introduced because the species is morphologically very different from previously described genera and our phylogenetic analyses do not show any close relationship with other sequenced monothalamous species.

Arnoldiellina fluorescens sp. nov.

Holotype: MHNG INVE 82002.

Type material: A specimen preserved in formalin was selected as holotype and deposited at the Museum of Natural History in Geneva (MHNG) together with 7 paratypes (MHNG INVE 82003).

Type locality: Gulf of Eilat, Israel.

Other material examined: 35 additional specimens were either extracted in guanidine (13 specimens), preserved in formalin (8 specimens), fixed for DAPI staining (4 specimens) or observed and photographed alive (10 specimens). The rapid streaming of protoplasm can be observed in Videos S1 and S2. The multiple nuclei in *Arnoldiellina* are shown in Fig. 2D.

Etymology: The species name is based on the ability of this foraminifer to emit green autofluorescence.

Diagnosis: As for genus.

Description: Measurement of length and width of 16 different specimens are shown in the Table 1. All specimens were fusiform; however, the ratio length/width may vary between the specimens. The length is 3–5 times the width. The most compressed specimens is the one shown in Fig. 2(C, D); one of the paratypes shown in Fig. 2B was the most elongated.

Description of the holotype: Test free, monothalamous, fusiform, 270 μm in length and 65 μm in width, organic wall transparent of 6.6 μm width; the wall increases in thickness around the single terminal aperture. The aperture is funnel-shaped with a tubular extension inside the theca.

Remarks: Compared to the other species assembled in clade I, *Arnoldiellina* differs in its morphology by its small size and organic wall. All other members of clade I are characterized by the presence of cell bodies and agglutinated tests surrounding them.

Molecular characterization: A total of 13 single-cell DNA extracts were obtained. PCR amplification of the 18S rDNA fragment produced positive results for seven DNA extractions. Three sequences were obtained for the fragment s14F1-20r and four additional sequences were obtained for a shorter fragment (s14F1-s17). All obtained sequences were nearly identical. Only the longer fragments (s14F1-s20r) were used for the following analysis.

The three sequences of *A. fluorescens* were aligned to 57 sequences of monothalamous foraminifera selected from our database. We arbitrarily used environmental clades (Pawlowski et al. 2011) as an outgroup (Fig. 3). Our analysis shows that the *Arnoldiellina* sequences branch within clade I (Pawlowski et al. 2002b), as sister group to *Pelosina* and *Astrammina*, but this relationship is weakly supported (59%). Higher bootstrap value (85%) was obtained for the whole clade I, including *Armorella*, *Saccodendron* and *Pelosinella* (Fig. 3). Interestingly, an insertion of about 50 nucleotides characteristic for clade I is also present in *Arnoldiellina* (Video S1). Analysis of the secondary structure shows that this insertion forms a helix situated between helices 45 and 47, absent in other foraminiferans, except some lineages of clade C (Fig. 3A).

Discussion

Our study is the first report of GAF in foraminifera, but the phenomenon is relatively well known in protists. It seems

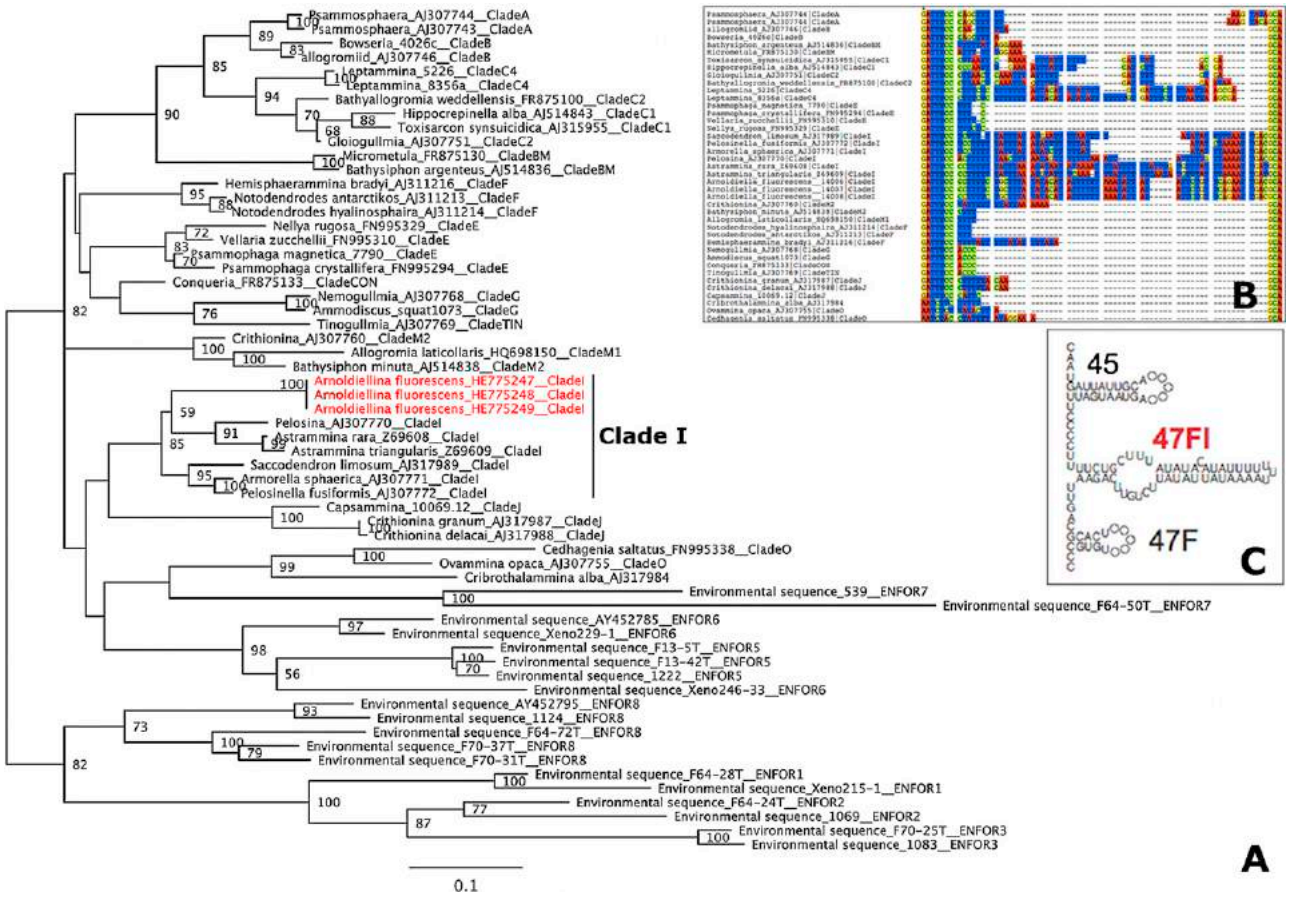


Fig. 3. (A) Phylogenetic tree of monothalamous foraminifera based on partial 18S rDNA sequences, showing the position of *Arnoldiellina fluorescens* gen. and sp. nov. Support values are given as RaxML bootstrap; only values ≥50 are shown. (B) Alignment of the region of the 18S rDNA between the helix 45 and 47, showing the insertion specific to clade I. (C) Secondary structure of the insertion in *Arnoldiellina fluorescens*.

to be a common feature in heterotrophic and autotrophic dinoflagellates (Carpenter and Chang 1991; Tang and Dobbs 2007), considered sometimes as a useful taxonomic character (Elbrächter 1994). Its presence in all life-history stages of the parasitic dinoflagellate *Amoebophrya* (Chambouvet et al. 2011) is commonly used to detect infection of phytoplankton (Coats and Bockstahler 1994; Park et al. 2004). The GAF was also found in diatoms, chlorophytes, raphidophytes, and other microalgae (Tang and Dobbs 2007). Among heterotrophic protists other than dinoflagellates, GAF was only observed in ciliates (Laval-Peuto and Rassoulzadegan 1988).

The case of *Arnoldiellina* confirms that the presence of GAF is not specifically linked to autotrophic activity. Although in many algae GAF is found in association with chloroplasts, its localisation is often very different, for example near the dinoflagellate stigma (Tang and Dobbs 2007) or in the flagellum of brown and golden algae (Coleman 1988). In *Arnoldiellina*, the GAF is evenly distributed throughout the cytoplasm, suggesting the presence of a fluorescent compound produced by the cell. The nature of this compound is unknown, but it might be similar to luciferase or the green fluorescent protein present in many organisms (Gould et al. 1988; Shimomura et al. 1962), or else the flavoprotein found in the posterium flagellum of brown algae (Fujita et al. 2005).

The evolutionary importance of GAF in foraminifera is questionable. *Arnoldiellina* is the first well documented case of a foraminiferan that emits green autofluorescence. However, this property might occur more often among foraminifera as assumed so far. Some unpublished observations suggest GAF activity in other foraminiferal species (Sam Bowser, Ivan Volsky, pers. commun.). In fact, until now very few foraminiferans have been examined using epifluorescence microscopy. A systematic use of this technique in foraminiferal research may reveal other cases of natural green autofluorescence in this group and possibly also in other protists.

Acknowledgements

This study was supported by Swiss National Foundation Grant 31003A-125372 (J.P.) and G & L Claraz Donation.

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.2012.08.005.

References

Altin, D.Z., Habura, A., Goldstein, S.T., 2009. A new allogromiid foraminifer *Niveus flexilis* nov. gen., nov. sp., from Coastal Georgia, USA: fine structure and gametogenesis. J. Foraminiferal Res. 39, 73–86.

Carpenter, E.J., Chang, J., Shapiro, L.P., 1991. Green and blue fluorescing dinoflagellates in Bahamian waters. Mar. Biol. 108, 145–149.

Chambouvet, A., Alves-de-Souza, C., Cueff, V., Marie, D., Karpov, S., Guillou, L., 2011. Interplay between the parasite *Amoebophrya* sp. (Alveolata) and the cyst formation of the red tide dinoflagellate *Scrippsiella trochoidea*. Protist 162, 637–649.

Coats, D.W., Bockstahler, K.R., 1994. Occurrence of the parasitic dinoflagellate *Amoebophrya ceratii* in Chesapeake Bay populations of *Gymnodinium sanguineum*. J. Eukaryot. Microbiol. 41, 586–593.

Coleman, A.W., 1988. The auto-fluorescent flagellum – a new phylogenetic enigma. J. Phycol. 24, 118–120.

Elbrächter, M., 1994. Green autofluorescence – a new taxonomic feature for living dinoflagellate cysts and vegetative cells. Rev. Palaeobot. Palynol. 84, 101–105.

Fujita, S., Iseki, M., Yoshikawa, S., Makino, Y., Watanabe, M., Motomura, T., Kawai, H., Murakami, A., 2005. Identification and characterization of a fluorescent flagellar protein from the brown alga *Scytosiphon lomentaria* (Scytosiphonales Phaeophyceae): a flavoprotein homologous to Old Yellow Enzyme. Eur. J. Phycol. 40, 159–167.

Gooday, A.J., 2002. Organic-walled allogromiids: aspects of their occurrence, diversity and ecology in marine habitats. J. Foraminiferal Res. 32, 384–399.

Gooday, A.J., Holzmann, M., Guiard, J., Cornelius, N., Pawlowski, J., 2004. A new monothalamous foraminiferan from 1000 to 6300 m water depth in the Weddell Sea: morphological and molecular characterisation. Deep-Sea Res. II 51, 1603–1616.

Gooday, A.J., Pawlowski, J., 2004. *Conqueria laevis* gen. and sp nov, a new soft-walled, monothalamous foraminiferan from the deep Weddell Sea. J. Mar. Biol. Assoc. UK 84, 919–924.

Gooday, A.J., Bowser, S.S., Cedhagen, T., Cornelius, N., Hald, M., Korsun, S., Pawlowski, J., 2005. Monothalamous foraminiferans and gromiids (Protista) from western Svalbard: a preliminary survey. Mar. Biol. Res. 1, 290–312.

Gooday, A.J., da Silva, A.A., Koho, K.A., Lecroq, B., Pearce, R.B., 2010. The 'mica sandwich'; a remarkable new genus of Foraminifera (Protista, Rhizaria) from the Nazare Canyon (Portuguese margin, NE Atlantic). Micropaleontology 56, 345–357.

Gould, S.J., Subramani, S., 1988. Firefly luciferase as a tool in molecular and cell biology. Anal. Biochem. 175, 5–13.

Gouy, M., Guindon, S., Gascuel, O., 2010. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol. Biol. Evol. 27, 221–224.

Gruber, A.R., Lorenz, R., Bernhart, S.H., Neuboock, R., Hofacker, I.L., 2008. The Vienna RNA websuite. Nucl. Acids Res. 36, W70–W74.

Habura, A., Pawlowski, J., Hanes, S.D., Bowser, S.S., 2004. Unexpected foraminiferal diversity revealed by small-subunit rDNA analysis of Antarctic sediment. J. Eukaryot. Microbiol. 51, 173–179.

Habura, A., Wegener, L., Travis, J.L., Bowser, S.S., 2005. Structural and functional implications of an unusual foraminiferal beta-tubulin. Mol. Biol. Evol. 22, 2000–2009.

Habura, A., Goldstein, S.T., Broderick, S., Bowser, S.S., 2008. A bush, not a tree: the extraordinary diversity of cold-water basal foraminiferans extends to warm-water environments. Limnol. Oceanogr. 53, 1339–1351.

Haynes, J.R., 1981. Foraminifera. Macmillan Publ., London, p. 433.

Lavalpeuto, M., Rassoulzadegan, F., 1988. Autofluorescence of marine planktonic oligotrichina and other ciliates. Hydrobiologia 159, 99–110.

Lecroq, B., Lejzerowicz, F., Bachar, D., Christen, R., Esling, P., Baerlocher, L., Osteras, M., Farinelli, L., Pawlowski, J., 2011. Ultra-deep sequencing of foraminiferal microbarcodes unveils hidden richness of early monothalamous lineages in deep-sea sediments. Proc. Natl. Acad. Sci. U.S.A. 108, 13177–13182.

Lejzerowicz, F., Pawlowski, J., Fraissinet-Tachet, L., Marmeisse, R., 2010. Molecular evidence for widespread occurrence of foraminifera in soils. Environ. Microbiol. 12, 2518–2525.

Meisterfeld, R., Holzmann, M., Pawlowski, J., 2001. Morphological and molecular characterization of a new terrestrial allogromiid species: *Edaphoallogromia australica* gen. et spec. nov. (Foraminifera) from northern Queensland (Australia). Protist 152, 185–192.

Murray, J.W., 2006. Ecology and Applications of Benthic Foraminifera. Cambridge University Press, UK, p. 440.

Parfrey, L.W., Katz, L.A., 2010. Genome dynamics are influenced by food source in allogromia laticollaris strain CSH (foraminifera). Genome Biol. Evol. 2, 678–685.

Park, M.G., Yih, W., Coats, D.W., 2004. Parasites and phytoplankton, with special emphasis on dinoflagellate infections. J. Eukaryot. Microbiol. 51, 145–155.

Pawlowski, J., 2000. Introduction to the molecular systematics of foraminifera. Micropaleontology 46, 1–12.

Pawlowski, J., Fahrni, J.F., Brykczynska, U., Habura, A., Bowser, S.S., 2002a. Molecular data reveal high taxonomic diversity of allogromiid Foraminifera in Explorers Cove (McMurdo Sound, Antarctica). Polar Biol. 25, 96–105.

Pawlowski, J., Holzmann, M., Berney, C., Fahrni, J., Cedhagen, T., Bowser, S.S., 2002b. Phylogeny of allogromiid foraminifera inferred from SSU rRNA gene sequences. J. Foraminiferal Res. 32, 334–343.

Pawlowski, J., Holzmann, M., Berney, C., Fahrni, J., Gooday, A.J., Cedhagen, T., Habura, A., Bowser, S.S., 2003. The evolution of early foraminifera. Proc. Natl. Acad. Sci. U.S.A. 100, 11494–11498.

Pawlowski, J., Fontaine, D., da Silva, A.A., Guiard, J., 2011. Novel lineages of Southern Ocean deep-sea foraminifera revealed by environmental DNA sequencing. Deep-Sea Res. II 58, 1996–2003.

Pawlowski, J., Majewski, W., 2011. Magnetite-bearing foraminifera from Admiralty Bay West Antarctica, with description of *Psammophaga magnetica*, sp. nov. J. Foraminiferal Res. 41, 3–13.

Sabbatini, A., Pawlowski, J., Gooday, A.J., Piraino, S., Bowser, S.S., Morigi, C., Negri, A., 2004. *Vellaria zucchellii* sp. nov. a new monothalamous foraminifer from Terra Nova Bay, Antarctica. Antarct. Sci. 16, 307–312.

Shimomura, O., Johnson, F.H., Saiga, Y., 1962. Extraction, purification and properties of aequorin, a bioluminescent protein from luminous hydromedusan, aequorea. J. Cell Compar. Physl. 59, 223–239.

Stamatakis, A., Hoover, P., Rougemont, J., 2008. A rapid bootstrap algorithm for the RAxML web servers. Syst. Biol. 57, 758–771.

Tang, Y.Z., Dobbs, F.C., 2007. Green autofluorescence in dinoflagellates, diatoms, and other microalgae and its implications for vital staining and morphological studies. Appl. Environ. Microb. 73, 2306–2313.

The Journal of
Eukaryotic Microbiology

Published by the International Society of Protistologists

Journal of Eukaryotic Microbiology ISSN 1066-5234

ORIGINAL ARTICLE

Molecular Phylogeny and Morphology of *Leannia veloxifera* n. gen. et sp. Unveils a New Lineage of Monothalamous Foraminifera

Laure Apothéloz-Perret-Gentil & Jan Pawlowski

Department of Genetics and Evolution, University of Geneva, 1211, Geneva 4, Switzerland

Keywords
Actin; allogromiids; recent; ß-tubulin; SSU rDNA.

Correspondence
L. Apothéloz-Perret-Gentil, Department of Genetics and Evolution, University of Geneva, 1211, Geneva 4, Switzerland
Telephone number: +41-22-379-30-84;
FAX number: +41-22-379-67-70;
e-mail: laure.perret-gentil@unige.ch

Received: 30 July 2014; revised 22 September 2014; accepted September 28, 2014.

doi:10.1111/jeu.12190

ABSTRACT

Monothalamous (single-chambered) foraminifera have long been considered as the "poor cousins" of multichambered species, which calcareous and agglutinated tests dominate in the fossil record. This view is currently changing with environmental DNA surveys showing that the monothalamids may be as diverse as hard-shelled foraminifera. Yet, the majority of numerous molecular lineages revealed by eDNA studies remain anonymous. Here, we describe a new monothalamous species and genus isolated from the sample of sea grass collected in Gulf of Eilat (Red Sea). This new species, named *Leannia veloxifera*, is characterized by a tiny ovoid theca (about 50–100 μm) composed of thin organic wall, with two opposite apertures. The examined individuals are multinucleated and show very active reticulopodial movement. Phylogenetic analyses of SSU rDNA, actin, and beta-tubulin (ß-tubulin) show that the species represents a novel lineage branching separately from other monothalamous foraminifera. Interestingly, the SSU rDNA sequence of the new species is very similar to an environmental foraminiferal sequence from Bahamas, suggesting that the novel lineage may represent a group of shallow-water tropical allogromiids, poorly studied until now.

RECENT development of high-throughput sequencing technology tremendously speeds up the process of the discovery of new environmental lineages of protists. Several high-rank taxonomy groups composed mainly of environmental sequences have been proposed, such as MAST 1-11 (Logares et al. 2012; Massana et al. 2014). Some of these groups could not be assigned to any supergroup and have no morphologically characterized representatives, e.g. Rappemonads (Kim et al. 2011). The interpretation of others has changed after a microscopic examination of cultivated isolates, e.g. Picozoa, formely Picobiliphytes (Seenivasan et al. 2013). The integrated taxonomy of protists based on morphological and molecular study appears as a necessity (Moreira and López-García 2014). Indeed, few studies combining the single DNA-barcoding with morphological and ultrastructural data have been very successful in identifying the enigmatic environmental lineages (Rueckert et al. 2011). However, such studies are time-consuming and require a good taxonomic expertise. Therefore, they are rare and can hardly fill the taxonomic gap in some poorly known groups such as monothalamous foraminifera.

Monothalamids are a heterogeneous assemblage of diverse foraminiferal lineages characterized by organic-walled or agglutinated single-chambered tests, called allogromiids or astrorhizids, respectively (Pawlowski et al. 2002). Because their tests are poorly preserved in dried samples routinely studied by foram specialists, the diversity of monothalamids has never been extensively examined. It is well known that the group dominates in some marine habitats, especially in the deep-sea and high-latitude regions (Gooday 2002; Gooday et al. 2005), but they are also common in warm water environments (Habura et al. 2008) and in freshwater (Dellinger et al. 2014; Holzmann et al. 2003). Many new monothalamous species have been described in the last decade (Altin et al. 2009; Apothéloz-Perret-Gentil et al. 2013; Gooday and Pawlowski 2004; Gooday et al. 2004, 2010; Pawlowski and Majewski 2011; Sabbatini et al. 2004; Voltski et al. 2014). Yet, as suggested by large number of undetermined

© 2014 The Author(s) Journal of Eukaryotic Microbiology © 2014 International Society of Protistologists

Table 1. Description of 23 specimens of *Leannia veloxifera* n. gen. et sp.

	Specimens	Length (μm)	Width (μm)	Ratio length/width	Figure	Sequences		
						18S	ß-tubulin	Actin
1	DNA (17004)	100	54	1.9		LM994876	LM994880	LM994879
2	DNA (17005)	108	64	1.7		LM994877		
3	DNA (17006)	88	60	1.5	Fig. 1C	LM994878		
4	Holotype	111	64	1.7	Fig. 1A, E)			
5	Paratype	100	76	1.3	Fig. 1B			
6	Paratype	106	77	1.4	Fig. 1B			
7	Paratype	94	63	1.5	Fig. 1B			
8	Paratype	102	62	1.7	Fig. 1B			
9	Paratype	95	64	1.5	Fig. 1B			
10	DAPI	74	69	1.1				
11	DAPI	83	62	1.3				
12	Formaline	87	84	1.0				
13	Formaline	82	79	1.0				
14	Formaline	83	46	1.8				
15	Formaline	66	46	1.4	Fig. 1F			
16	Formaline	68	41	1.7				
17	RNA	106	70	1.5				
18	RNA	106	67	1.6				
19	RNA	103	57	1.8				
20	RNA	72	60	1.2				
21	RNA	102	61	1.7				
22	RNA	95	61	1.6				
23	RNA	113	62	1.8	Fig. 1D			

species found in monothalamids diversity surveys (Gooday et al. 2005; Majewski 2005; Majewski et al. 2007), our knowledge of the group is still very fragmentary.

The immense diversity of monothalamids was confirmed by environmental DNA (eDNA) studies. The sequences assigned to monothalamous lineages dominate in all eDNA surveys of foraminiferal communities, both those that used clonal approach (Bernhard et al. 2013; Habura et al. 2004, 2008; Pawlowski et al. 2011a; Tsuchiya et al. 2013) and those using next-generation sequencing technology (Lecroq et al. 2011; Lejzerowicz et al. 2013; Pawlowski et al. 2011b, 2014). In some deep-sea samples, the proportion of monothalamids reaches up to 74% and may be even higher if we consider that most of unassigned OTUs also belong to this group (Lecroq et al. 2011). Most of the monothalamous sequences retrieved from deep-sea samples group within eight large clades defined as ENFOR 1-8 (Pawlowski et al. 2011a), but many represent independent lineages comprising usually one or few sequences. Remarkably, none of these environmental lineages comprises morphologically described species, what makes them even more enigmatic.

To know more about monothalamid diversity, we started to systematically examine the morphology and obtain genetic data for all monothalamous species that appeared in our samples. Previously, we described a new fluorescent allogromiid from Gulf of Eilat (Apothéloz-Perret-Gentil et al. 2013). Here, we report another new species from the same locality. Phylogenetic study of this species shows that it represents a novel lineage of monothalamids, which also comprises the environmental sequence of an uncultured foraminifer from Bahamas.

MATERIALS AND METHODS

Isolation

Samples of the *Halophila* leaves were collected by SCUBA diving at 15 m in front of the Interuniversity Institute for Marine Sciences (IUI), in Eilat, Israel, on December 2012. The coordinates of the sampling spot are: 29.51482 N 34.92674 E. Large benthic foraminifera of the genus *Amphisorus* were detached by hand from the sea grass and transfered to Petri dishes filled with filtered seawater to which few drops of Erdschreiber medium (5% soil extract, 1 mM $NaNO_3$, 0.07 mM Na_2HPO_4, 10 mM Tris pH = 8, filled up with sterile seawater) were added. The specimens of small-sized allogromiid foraminifera that are described in this paper appeared in the culture dishes several days after placing *Amphisorus* there. They flourish in culture dishes for few weeks, probably in result of asexual reproduction of few individuals, and then rapidly disappeared.

Morphology and cytology

Three living specimens were incubated 5 min at ambient temperature using 4′,6′-diamidino-2-phénylindole (DAPI) at 5.10E-4 mg/ml to stain and identify nuclei. The procedure was carried out in a dark room. Five specimens were fixed in a 10% solution of formalin. Living and fixed specimens

© 2014 The Author(s) Journal of Eukaryotic Microbiology © 2014 International Society of Protistologists

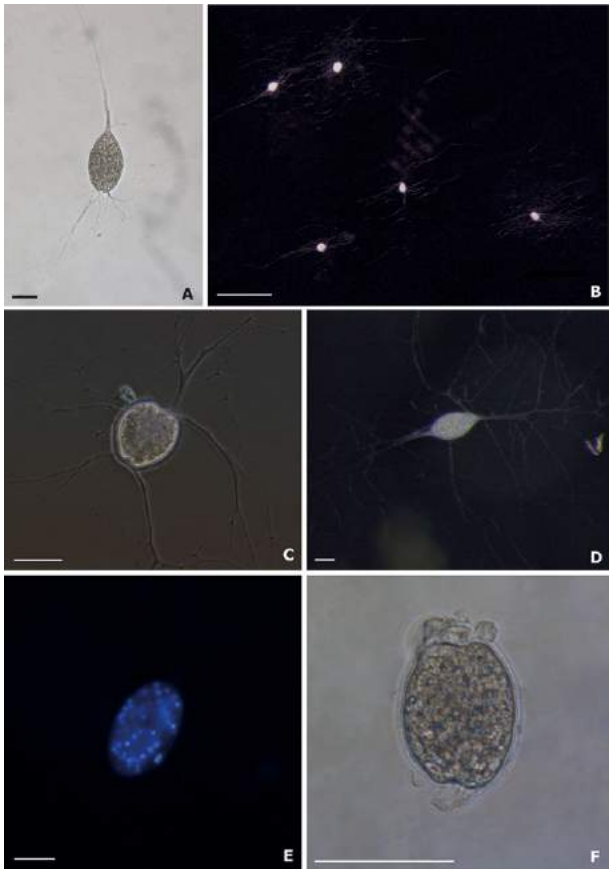


Figure 1 Specimens of *Leannia veloxifera* n. gen. et sp. **A.** Living holotype. **B.** Paratypes with their expensive granuloreticulopodia's web. **C, D.** Two living specimens. **E.** Holotype stained with DAPI (blue) viewed with UV light excitation (460–500 nm). **F.** Fixed specimen. Scale bar (A, C–F) correspond to 50 μm and scale bar (B) correspond to 500 μm.

were observed with an inverted microscope (Nikon Eclipse Ti, Nikon Instruments Europe, Amsterdam, Netherlands), a fluorescence microscope (Nikon Eclipse E200) and a stereoscopic one (Leica M205C, Leica, Hamburg, Germany). Photographs were taken with a Leica DFC 420C, an Imaging Source DFK 41AF02 camera, and a Leica DFC 450C, respectively. Movies were made on the fluorescent microscope and the inverted microscope with the same camera. They are available at: http://forambarcoding.unige.ch/movies

DNA/RNA extraction, amplification, cloning, and sequencing

DNA from three specimens was extracted in guanidine lysis buffer (Pawlowski 2000), each extraction was performed with a single specimen. RNA extraction was performed with seven specimens using the NucleoSpin RNA XS kit (Macherey-Nagel, Düren, Germany). Afterwards cDNA was synthetised using the iScript Select cDNA synthesis Kit (BioRad, Hercules, CA) with random primers. PCR amplifications of the complete SSU rDNA were performed in three steps. The first fragment was amplified using the primer pair s14F3 (5′ACG CA(AC) GTG TGA AAC TTG) and B (5′TGA TCC TTC TGC AGG TTC ACC TAC). PCR products were re-amplified using the nested primer s14F1 (5′AAG GGC ACC ACA AGA ACG C). The second fragment was amplified using the primer pair 6F (5′CCG CGG TAA TAC CAG CTC) and 17 (5′CGG TCA CGT TCG TTG C). PCR products were re-amplified using the nested primer 15A (5′CTA AGA ACG GCC ATG CAC CAC C). The third fragment was amplified using the primer pair A10 (5′CTC AAA GAT TAA GCC ATG CAA GTG G) and 12R (5′G(GT)T AGT CTT (AG)(AC)(ACT) AGG GTC A). PCR products were re-amplified using the nested primer 7R (5′CTG (AG)TT TGT TCA CAG T(AG)T TG). The sequenced fragments have been assembled to retrieve the complete SSU rDNA.

PCR amplifications of a fragment of the actine gene were performed using the primer pair ActN2 (5′ACC TGG GA(CT) GA(CT) ATG GA) and 1354R (5′GGA CCA GAT TCA TCA TA(CT) TC). PCR products were re-amplified using the nested primer ActF1 (5′CNG A(AG)G C(AGT)C CAT T(AG)A A(CT)C), as described in Flakowski et al. (2005).

PCR amplifications of a fragment of the ß-tubulin gene were performed using the primer pair BtubF1 (5′CAA TGT GGT AAC CAA ATT GC) and BtubR1 (5′CAT CTT GTT TGT CTT GAT ATT CAG T). PCR products were re-amplified using the nested primer BtubF2 (5′AAT TGG GCA AAA GGA CAT TA), as described in Habura et al. (2005).

The amplified PCR products were purified using High Pure PCR Purification Kit (Roche Diagnostics, Hoffmann-La Roche AG, Basel, Switzerland) and cloned with the TOPO10 kit from Invitrogen (Thermo Fisher Scientific, Waltham, MA). Between two and four clones were sequenced per PCR. Sequencing reactions were performed using the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Thermo Fisher Scientific) and analysed on a 3130XL Genetic Analyser (Applied Biosystems). The five new sequences reported in this paper were deposited in the EMBL/GenBank data base (LM994876–LM994880).

Sequence alignments and phylogenetic analysis

The gene coding sequences were translated into amino acid sequences using Seaview vs 4.3.3. software (Gouy et al. 2010). All the sequences were aligned using the same program.

The SSU rDNA sequences were aligned to 28 foraminiferan sequences and 1,943 sites of the alignment were used for the analysis using GTR+G+I model. For the short fragment, 37 environmental sequences were added to 25 foraminferan ones and the whole alignment of 2,016 sites was used with GTR+G+I as model of evolution. Actin sequences were aligned to 35 sequences of Retaria (27 foraminiferans and 8 radiolarians used as outgroup) and 274 sites were used for the analysis using WAG+G model. ß-tubulin sequences were aligned to 28 sequences of Retaria (21

© 2014 The Author(s) Journal of Eukaryotic Microbiology © 2014 International Society of Protistologists

foraminiferans and 7 radiolarians used as outgroup) and 262 sites were used for the analysis using the WAG+G model. In addition, we performed a concatenated analysis of the actin and ß-tubulin genes with 35 sequences of Retaria; for 19 species no ß-tubulin gene data were available.

Best models for all analyses were calculated using Mega5 (Tamura et al. 2011). Phylogenetic trees were constructed using maximum likelihood program RAxML BlackBox (Stamatakis et al. 2008). In addition, Bayesian analyses were performed for all gene trees using MrBayes 3.2.1 (Ronquist and Huelsenbeck 2003) with four chains running in parallel for 10,000,000 generations. For each analysis, a burnin of 20% was carried out to construct the best tree and calculate posterior probabilities.

RESULTS

Morphologic description

The new species is a monothalamid without test. Specimens present an ovoid shape (ratio length/width between 1 and 2) between 72 and 113 μm in length and 41 and 84 μm in width. The measurement of each specimens observed is recorded in Table 1. Their organic wall is transparent and measure from 1 to 3 μm in width. They possess two opposite apertures, funnel-shaped with a tubular internal extension. Cytoplasm is multinucleate (Fig. 1E) and granular, with rapid movement (Movie S1). However, the multinucleate nature may represent only a stage of the life cycle. Reticulopodes are very active. They rapidly form large reticulopodial network and fast moving granules inside (Movie S2).

Seventeen additional specimens were used either for DNA or RNA extraction and subsequent amplification, fixed in formalin or observed and photographed alive. Description of the used specimens is summarised in Table 1.

Molecular phylogeny (SSU rDNA, actin, ß-tubulin)

To investigate the phylogenetic position of the new species, we performed an analysis of complete SSU rRNA gene sequence (total length 3,033 bp, GC content 32%). Three sequences were aligned to 25 sequences of foraminifera from our database and phylogenetic trees were built using ML and BI methods (Fig. 2). The tree is rooted at the clade I according to the ß-tubulin phylogeny (Fig. S4) and Hou et al. (2013). The new allogromiid sequences form a very long branch (reduced 50% in Fig. 2) not related to any of the previously described monothalamous clades (Pawlowski et al. 2002). Its position at the base of a clade formed by eight globothalamean species and few monothalamids belonging to clades A, BM, and C is relatively well supported (0.95 PP, 74% BV). Relationships between other monothalamid clades, including *Capsammina patelliformis*, *Allogromia* sp., *Nemogullmia* sp., the clade E and the freshwater foraminifer *Reticulomyxa filosa* are not resolved. The topology of the ML tree differs from the BI tree in the position of *C. patelliformis*, which branches at the base of the tree.

To further refine the phylogenetic position we analysed actin and ß-tubulin genes. In the actin tree (Fig. S1), its branch is very long compared to other foraminiferans. The new species groups in the unresolved clade formed by six tubothalameans, *Bathysiphon flexilis* and *R. filosa*. This clade is sister to monothalamous clade M, composed of *Allogromia*, *Edaphoallogromia* and *Bathysiphon* sp. Both clades form a sister group to Globothalamea, which are well supported in Bayesian analyses (1 PP) but not in ML analysis (53% BV). The topology of the ML tree differs slightly, with monothalameous clade M branching at the base of Globothalamea and the clade, to which belongs the new allogromiid species. However, this topology is not supported (less than 25% BV).

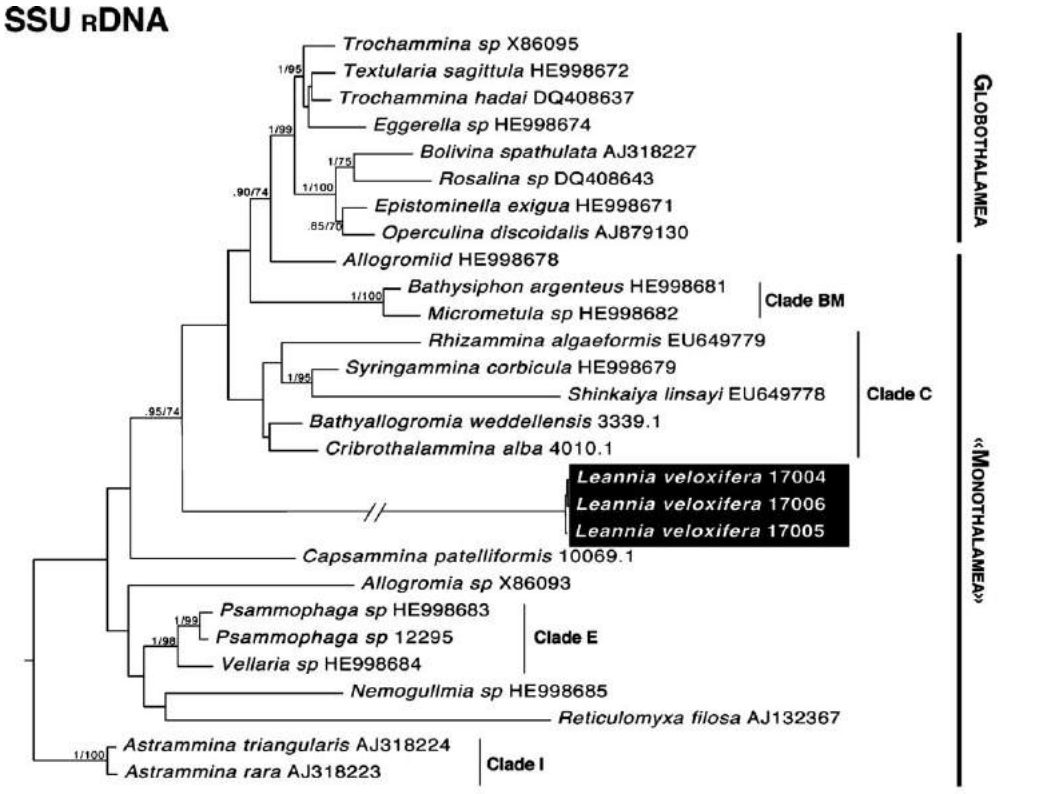
In the ß-tubulin tree (Fig. S2), the amino acid sequence of the new allogromiid branches as sister to Globothalamea. This relation is strongly supported in Bayesian analysis but not in ML analysis. The topology of foraminiferal tree is characterized by strong support for Globothalamea (1 PP, 93% BV), and paraphyly of monothalamids and tubothalameans. A monothalamid *Astrammina rara* branches at the base of the tree, followed by a clade of *Allogromia* and *Crithionina delacai*. However, none of these branching patterns is strongly supported.

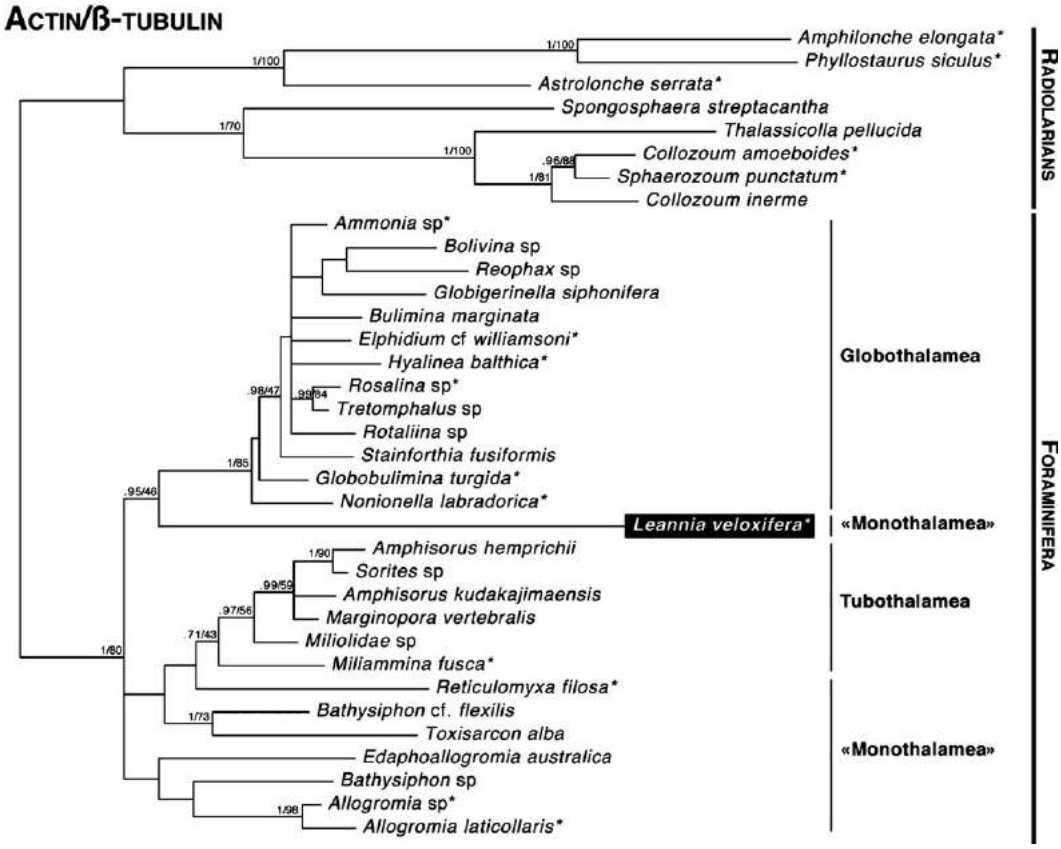
A final analysis was carried out by using the concatenated ß-tubulin and actin genes (Fig. 3) with radiolarians as outgroup. Within foraminifera, Globothalamea form a distinct group (1 PP, 85% BV) with the new species branching at their base (0.95 PP, 46% BV). The other monothalamids form unsupported branches with Tubothalamea branching within them (0.71 PP, 43% BV). The monophyly of foraminifera is relatively well supported (1 PP, 80% BV).

In addition to phylogenetic analyses of complete SSU, actin, and ß-tubulin sequences, we also analysed a short fragment of the SSU rDNA, commonly used as foraminiferal barcode (Pawlowski and Holzmann 2014), and for which many environmental sequences are available. In Fig. 3, we present a tree with 62 selected sequences representing previously described environmental clades (ENFOR), unique environmental lineages (ENV), undetermined

Figure 2 Phylogenetic tree of 34 sequences of foraminifera based on complete SSU rDNA sequences, showing the position of *Leannia veloxifera* n. gen. et sp. in a black frame. Support values are given as MrBayes posterior probabilities/RaxML bootstrap; only values superior or equal to 0.85 for posterior probabilities and 70 for bootstrap values are shown. Concatenated phylogeny of actin and ß-tubulin genes. The analysis was done with 27 sequences of foraminifera sequences with eight sequences of radiolarian used as outgroup. Both genes were retrieved for species with a asterisk (*), ß-tubulin gene are missing for the other. Support values are given as MrBayes posterior probabilities/RaxML bootstrap; only values superior or equal to 0.50 for posterior probabilities and 40 for bootstrap values are shown. The position of *Leannia veloxifera* n. gen. et sp. is shown in a black frame.

© 2014 The Author(s) Journal of Eukaryotic Microbiology © 2014 International Society of Protistologists





© 2014 The Author(s) Journal of Eukaryotic Microbiology © 2014 International Society of Protistologists

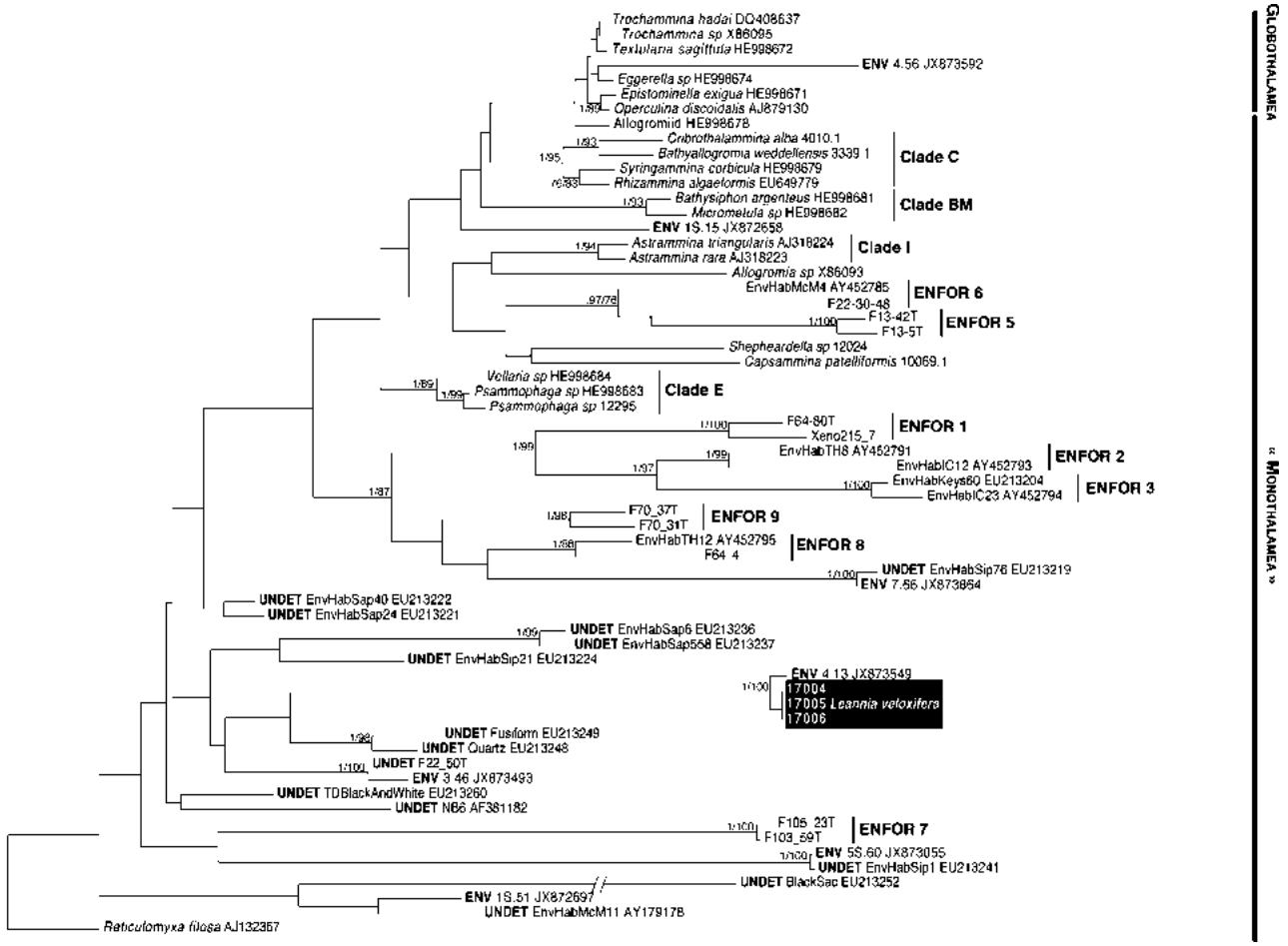


Figure 3 Phylogenetic tree of 62 sequences of foraminifera including 37 environmental sequences. *Leannia veloxifera* n. gen. et sp. sequences are framed in black. Support values are given as MrBayes posterior probabilities/RaxML bootstrap; only values superior or equal to 0.75 for posterior probabilities and 75 for bootstrap values are shown.

monothalamous morphotypes (UNDET) and identified morphospecies. The new allogromiid species did not branch with any of the previously described environmental clades (ENFOR 1-9, Pawlowski et al. 2011a). However, it branches with the unique environmental sequence of "uncultured foraminifera" from the Highborne Cay in Bahamas (Bernhard et al. 2013). Both sequences differ by only 8% and their relation is highly supported (1 PP, 100% BV). A sequence of another uncultured foraminifera from Sippewissett marshes in Massachusetts (Habura et al. 2008) branches at the base of this group.

We also looked for environmental sequences in the large dataset of environmental sequences provided by next-generation sequencing. We found two sequences, both from Marlborough Sounds in New Zealand, which are related to the new clade. One is exactly identical to *Leannia* sequences (100% identity and 100% coverage) and the other is close (98% identity and 90% coverage) to the sequence from Highborne Cay in Bahamas. However, those sequences are very short (53 and 57 bp respectively) and correspond only to one hypervariable region (37F) of the SSU rDNA. Therefore, we did not add them to the tree on Fig. 3.

DISCUSSION

The species described here is the second new allogromiid, after *Arnoldiellina fluorescens* (Apothéloz-Perret-Gentil et al. 2013), reported from the same locality in Gulf of Eilat during the last 3 yr. This may sounds surprising given an extensive foraminiferal research that has been conducted in this area over the years and which conducted to an impressive number of publications about Gulf of Eilat foraminifera (reviewed in Hottinger et al. 1993; Lee and Anderson 1991; Reiss and Hottinger 1984). However, all these classical work focused on large benthic foraminifera and does not care about the small-sized species. One of us (JP) showed many years ago that the poorly known community of calcareous microforaminifera flourish on the *Halophila* leaves and coral rubble in the Gulf of Eilat (Pawlowski and Lee 1991, 1992). At that time, however, our attention was focused on tiny calcareous species, which

© 2014 The Author(s) Journal of Eukaryotic Microbiology © 2014 International Society of Protistologists

could be identified either directly on dried leaves or in the fine fraction of sediment samples. Two new genera and eight new species of microforaminifera belonging to the families Glabratellidae and Rotaliellidae have been described (Pawlowski and Lee 1991, 1992).

Compared to this work on hard-shelled foraminifera, the isolation and description of new allogromiid species is much more challenging. The organic-walled foraminifera are not preserved in dried samples and can be isolated only from laboratory cultures or formalin-fixed samples. The cultivation approach traditionally used in protistology is seldom applied to foraminiferal species, because they are difficult to maintain in laboratory cultures and their description has to be done rapidly after they have been observed. The allogromiids usually flourish in culture dishes for few weeks, probably in result of asexual reproduction of one or two individuals, and then rapidly disappeared. Only few species adapt to culture conditions and can be maintained for longer periods of time, like for example, *Allogromia laticollaris* or other species of this genus (McEnery and Lee 1976; Parfrey and Katz 2010).

Despite these difficulties, our study shows that the cultivation, even for short periods of time, is essential for taxonomic study of this group. Hundreds of novel lineages have been revealed by eDNA and RNA studies (Bernhard et al. 2013; Pawlowski et al. 2011a; Tsuchiya et al. 2013; reviewed in Pawlowski J., Lejzerowicz F., Esling, P., unpubl. data), but most of them remained microscopically undocumented. The fact that *Leannia veloxifera* branches with one of these enigmatic lineages confirms that at least some of them can be assigned to tiny allogromiids, which possibly form a rich community in shallow tropical waters. Their inconspicuous presence may also explain the immense diversity of environmental lineages observed at the deep-sea bottom (Lecroq et al. 2011; Lejzerowicz et al. 2013; Pawlowski et al. 2011a). Many of these undetermined sequences have been amplified from samples of xenophyphoreans or other large deep-sea benthic foraminifera, which tests could provide a suitable habitat for tiny allogromiids (Lecroq et al. 2009). More extensive cultivation efforts coupled with a detailed microscopic study could lift the veil on these mysterious "eDNA" foraminiferans.

TAXONOMIC SUMMARY

Supergroup RHIZARIA Cavalier-Smith, 2002
Phylum FORAMINIFERA D'Orbigny, 1826
Class "Monothalamea" Pawlowski et al. 2003

Leannia n. gen. Apothéloz-Perret-Gentil et Pawlowski 2014

Description. Test free, monothalamous, ovoid shape (ratio length/width between 1 and 2), < 115 μm in length and < 85 μm in width; organic wall transparent from 1 to 3 μm in width. Two opposite apertures, funnel-shaped with a tubular internal extension. Cytoplasm multinucleate (Fig. 1E) at least in this stage of its life cycle; granular, with rapid movement (Movie S1). Reticulopodes very active with rapidly forming large reticulopodial network and fast moving granules (Movie S2).
Type species. *Leannia veloxifera* n. sp. Apothéloz-Perret-Gentil et Pawlowski 2014
Etymology. The genus was named in honour of first author's daughter.

Leannia veloxifera n. sp. Apothéloz-Perret-Gentil et Pawlowski 2014

Description. Same as for genus.
DNA/Amino acids sequences. SSU rDNA sequences, Actin and ß-tubulin proteins (GenBank LM994876–LM994880)
Type locality. Gulf of Eilat, Red Sea (*Halophila* sea grass meadow in front of the IUI, Eilat, Israel).
Type habitat. Marine
Type material. A specimen preserved in formalin was selected as holotype (MHNG INVE 89252) and deposited at the Museum of Natural History in Geneva (MHNG) together with five paratypes (MHNG INVE 89253).
Etymology. The species was named for the extreme rapidity to form its granuloreticulopodial network.
Remarks. *Leannia veloxifera* is morphologically similar to *Arnoldiellina fluorescens*, another allogromiid described from the Gulf of Eilat (Apothéloz-Perret-Gentil et al. 2013). However, *Leannia* had two apertures, while *Arnoldiellina* possesses only one. Moreover, the later species shows green autofluorescence when observed under UV light.

ACKNOWLEDGMENTS

We are thankful to Maria Holzmann, Emmanuela Reo and Shai Oron for help in collecting samples and to Sigal Abramovich and the members of the staff for hosting us in the IUI station in Eilat. Maria Holzmann helps revising the manuscript.

This study was supported by the Swiss National Foundation grant 31003A-125372 (JP) and G & L Claraz Donation.

LITERATURE CITED

Altin, D. Z., Habura, A. & Goldstein, S. T. 2009. A new allogromiid foraminifer *Niveus Flexilis* nov. gen., nov. sp., from Coastal Georgia, USA: fine structure and gametogenesis. *J. Foramin. Res.*, 39:73–86.

Apothéloz-Perret-Gentil, L., Holzmann, M. & Pawlowski, J. 2013. *Arnoldiellina fluorescens* gen. et sp. nov. – a new green autofluorescent foraminifer from the Gulf of Eilat (Israel). *Eur. J. Protistol.*, 49:210–216.

Bernhard, J. M., Edgcomb, V. P., Visscher, P. T., McIntyre-Wressnig, A., Summons, R. E., Bouxsein, M. L., Louis, L. & Jeglinski, M. 2013. Insights into foraminiferal influences on microfabrics of microbialites at Highborne Cay, Bahamas. *PNAS*, 110:9830–9834.

Dellinger, M., Labat, A., Perrouault, L. & Grellier, P. 2014. *Haplomyxa saranae* gen. nov. et sp. nov., a new naked freshwater foraminifer. *Protist*, 165:317–329.

© 2014 The Author(s) Journal of Eukaryotic Microbiology © 2014 International Society of Protistologists

Flakowski, J., Bolivar, I., Fahrni, J. & Pawlowski, J. 2005. Actin phylogeny of foraminifera. *J. Foramin. Res.*, 35:93–102.

Gooday, A. J., Bowser, S. S., Cedhagen, T., Cornelius, N., Hald, M., Korsun, S. & Pawlowski, J. 2005. Monothalamous foraminiferans and gromiids (Protista) from western Svalbard: a preliminary survey published in collaboration with the University of Bergen and the Institute of Marine Research, Norway, and the Marine Biological Laboratory, University of Copenhagen, Denmark. *Mar. Biol. Res.*, 1:290–312.

Gooday, A. J., Holzmann, M., Guiard, J., Cornelius, N. & Pawlowski, J. 2004. A new monothalamous foraminiferan from 1000 to 6300 m water depth in the Weddell Sea: morphological and molecular characterisation. *Deep-Sea Res. Pt II*, 51:1603–1616.

Gooday, A. J. & Pawlowski, J. 2004. *Conqueria laevis* gen. and sp. nov., a new soft-walled, monothalamous foraminiferan from the deep Weddell Sea. *J. Mar. Biol. Assoc. U.K.*, 84:919–924.

Gooday, A. J., da Silva, A. A., Koho, K. A., Lecroq, B. & Pearce, R. B. 2010. The "mica sandwich"; a remarkable new genus of Foraminifera (Protista, Rhizaria) from the Nazare Canyon (Portuguese margin, NE Atlantic). *Micropaleontology*, 56:345–357.

Gooday, A. J. 2002. Organic-walled allogromiids: aspects of their occurrence, diversity and ecology in marine habitats. *J. Foramin. Res.*, 32:384–399.

Gouy, M., Guindon, S. & Gascuel, O. 2010. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.*, 27:221–224.

Habura, A., Goldstein, S. T., Broderick, S. & Bowser, S. S. 2008. A bush, not a tree: the extraordinary diversity of cold-water basal foraminiferans extends to warm-water environments. *Limnol. Oceanogr.*, 53:1339–1351.

Habura, A., Pawlowski, J., Hanes, S. D. & Bowser, S. S. 2004. Unexpected foraminiferal diversity revealed by small-subunit rDNA analysis of Antarctic sediment. *J. Eukaryot. Microbiol.*, 51:173–179.

Habura, A., Wegener, L., Travis, J. L. & Bowser, S. S. 2005. Structural and functional implications of an unusual foraminiferal β-tubulin. *Mol. Biol. Evol.*, 22:2000–2009.

Holzmann, M., Habura, A., Giles, H., Bowser, S. S. & Pawlowski, J. 2003. Freshwater foraminiferans revealed by analysis of environmental DNA samples. *J. Eukaryot. Microbiol.*, 50:135–139.

Hottinger, L., Halicz, E. & Reiss, Z. 1993. Recent foraminiferida from the Gulf of Aqaba, Red Sea. Opera Sazu, Ljubljana.

Hou, Y., Sierra, R., Bassen, D., Banavali, N. K., Habura, A., Pawlowski, J. & Bowser, S. S. 2013. Molecular evidence for β-tubulin neofunctionalization in retaria (Foraminifera and radiolarians). *Mol. Biol. Evol.*, 30:2487–2493.

Kim, E., Harrison, J. W., Sudek, S., Jones, M. D. M., Wilcox, H. M., Richards, T. A., Worden, A. Z. & Archibald, J. M. 2011. Newly identified and diverse plastid-bearing branch on the eukaryotic tree of life. *Proc. Natl Acad. Sci. USA*, 108:1496–1500.

Lecroq, B., Gooday, A. J., Cedhagen, T., Sabbatini, A. & Pawlowski, J. 2009. Molecular analyses reveal high levels of eukaryotic richness associated with enigmatic deep-sea protists (Komokiacea). *Mar. Biodiv.*, 39:45–55.

Lecroq, B., Lejzerowicz, F., Bachar, D., Christen, R., Esling, P., Baerlocher, L., Østerås, M., Farinelli, L. & Pawlowski, J. 2011. Ultra-deep sequencing of foraminiferal microbarcodes unveils hidden richness of early monothalamous lineages in deep-sea sediments. *PNAS*, 108:13177–13182.

Lee, J. J., Anderson, O. R. 1991. Symbiosis in foraminifera. *In*: Academic Press, Biology of Foraminifera. Academic Press, London. p. 157–220.

Lejzerowicz, F., Voltsky, I. & Pawlowski, J. 2013. Identifying active foraminifera in the Sea of Japan using metatranscriptomic approach. *Deep-Sea Res. Pt II*, 86–87:214–220.

Logares, R., Audic, S., Santini, S., Pernice, M. C., de Vargas, C. & Massana, R. 2012. Diversity patterns and activity of uncultured marine heterotrophic flagellates unveiled with pyrosequencing. *ISME J.*, 6:1823–1833.

Majewski, W., Lecroq, B., Sinniger, F. & Pawlowski, J. 2007. Monothalamous foraminifera from Admiralty Bay, King George Island, West Antarctica. *Pol. Polar Res.*, 28:187–210.

Majewski, W. 2005. Benthic foraminiferal communities: distribution and ecology in Admiralty Bay, King George Island, West Antarctica. *Pol. Polar Res.*, 26:159–214.

Massana, R., del Campo, J., Sieracki, M. E., Audic, S. & Logares, R. 2014. Exploring the uncultured microeukaryote majority in the oceans: reevaluation of ribogroups within stramenopiles. *ISME J.*, 8:854–866.

McEnery, M. & Lee, J. J. 1976. *Allogromia laticollaris*: a foraminiferan with an unusual apogamic metagenic fife cycles. *J. Protozool.*, 23:94–108.

Moreira, D. & López-García, P. 2014. The rise and fall of Picobiliphytes: how assumed autotrophs turned out to be heterotrophs. *BioEssays*, 36:468–474.

Parfrey, L. W. & Katz, L. A. 2010. Genome dynamics are influenced by food source in *Allogromia laticollaris* Strain CSH (Foraminifera). *Genome Biol. Evol.*, 2:678–685.

Pawlowski, J., Christen, R., Lecroq, B., Bachar, D., Shahbazkia, H. R., Amaral-Zettler, L. & Guillou, L. 2011a. Eukaryotic richness in the Abyss: insights from pyrotag sequencing. *PLoS ONE*, 6: e18169.

Pawlowski, J., Esling, P., Lejzerowicz, F., Cedhagen, T. & Wilding, T. A. 2014. Environmental monitoring through protist next-generation sequencing metabarcoding: assessing the impact of fish farming on benthic foraminifera communities. *Mol. Ecol. Resour.*, 14:1129–1140.

Pawlowski, J., Fontaine, D., da Silva, A. A. & Guiard, J. 2011b. Novel lineages of Southern Ocean deep-sea foraminifera revealed by environmental DNA sequencing. *Deep-Sea Res. Pt II*, 58:1996–2003.

Pawlowski, J., Holzmann, M., Berney, C., Fahrni, J., Cedhagen, T. & Bowser, S. S. 2002. Phylogeny of allogromiid Foraminifera inferred from SSU rRNA gene sequences. *J. Foramin. Res.*, 32:334–343.

Pawlowski, J. & Holzmann, M. 2014. A plea for DNA barcoding of Foraminifera. *J. Foramin. Res.*, 44:62–67.

Pawlowski, J. & Lee, J. J. 1991. Taxonomic notes on some tiny, shallow water foraminifera from the Northern Gulf of Elat (Red Sea). *Micropaleontology*, 37:149.

Pawlowski, J. & Lee, J. J. 1992. The life cycle of *Rotaliella elatiana* n. sp.: a tiny Macroalgavorous Foraminifer from the Gulf of Elat. *J. Protozool.*, 39:131–143.

Pawlowski, J. & Majewski, W. 2011. Magnetite-bearing foraminifera from Admiralty Bay, West Antarctica, with description of *Psammophaga Magnetica*, sp. nov.. *J. Foramin. Res.*, 41: 3–13.

Pawlowski, J. 2000. Introduction to the molecular systematics of foraminifera. *Micropaleontology*, 46:1–12.

Reiss, Z. & Hottinger, L. 1984. The Gulf of Aqaba: Ecological Micropaleontology. Springer-Verlag, Berlin.

Ronquist, F. & Huelsenbeck, J. P. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, 19:1572–1574.

Rueckert, S., Simdyanov, T. G., Aleoshin, V. V. & Leander, B. S. 2011. Identification of a divergent environmental DNA sequence

© 2014 The Author(s) Journal of Eukaryotic Microbiology © 2014 International Society of Protistologists

clade using the phylogeny of gregarine parasites (Apicomplexa) from crustacean hosts. *PLoS ONE*, 6:e18163.
Sabbatini, A., Pawlowski, J., Gooday, A. J., Piraino, S., Bowser, S. S., Morigi, C. & Negri, A. 2004. *Vellaria zucchellii* sp. nov. a new monothalamous foraminifer from Terra Nova Bay, Antarctica. *Antarct. Sci.*, 16:307–312.
Seenivasan, R., Sausen, N., Medlin, L. K. & Melkonian, M. 2013. Picomonas judraskeda gen. et sp. nov.: the first identified member of the Picozoa phylum nov., a widespread group of picoeukaryotes, formerly known as "picobiliphytes". *PLoS ONE*, 8:e59565.
Stamatakis, A., Hoover, P. & Rougemont, J. 2008. A rapid bootstrap algorithm for the RAxML web servers. *Syst. Biol.*, 57:758–771.
Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. & Kumar, S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.*, 28:2731–2739.
Tsuchiya, M., Gooday, A. J., Nomaki, H., Oguri, K. & Kitazato, H. 2013. Genetic diversity and environmental preferences of monothalamous foraminifers revealed through clone analysis of environmental small-subunit ribosomal DNA sequences. *J. Foramin. Res.*, 43:3–13.
Voltski, I., Korsun, S. & Pawlowski, J. 2014. *Toxisarcon taimyr* sp. nov., a new large monothalamous foraminifer from the Kara Sea inner shelf. *Mar. Biodiv.*, 44:213–221.

SUPPORTING INFORMATION

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

Figure S1. Actin gene phylogeny of 27 sequences of foraminifera with eight sequences of radiolarian used as outgroup. Support values are given as MrBayes posterior probabilities/RaxML bootstrap; only values superior or equal to 0.50 for posterior probabilities and 40 for bootstrap values are shown. The position of *Leannia veloxifera* n. gen. et sp. is shown in a black frame.
Figure S2. ß-tubulin gene phylogeny of 21 sequences of foraminifera with seven sequences of radiolarian are used as outgroup. Support values are given as MrBayes posterior probabilities/RaxML bootstrap; only values superior or equal to 0.50 for posterior probabilities and 40 for bootstrap values are shown. The position of *Leannia veloxifera* n. gen. et sp. is shown in a black frame.
Movie S1. Living specimen of *Leannia veloxifera* n. gen. et sp. Scale bar: 50 μm.
Movie S2. Magnification of the granuloreticulopodia's web on a living specimen of *Leannia veloxifera* n. gen. et sp. Scale bar: 50 μm.

© 2014 The Author(s) Journal of Eukaryotic Microbiology © 2014 International Society of Protistologists

Available online at www.sciencedirect.com

ScienceDirect

European Journal of Protistology 60 (2017) 28–44

European Journal of PROTISTOLOGY

www.elsevier.com/locate/ejop

Taxonomic revision of freshwater foraminifera with the description of two new agglutinated species and genera

Ferry Siemensma[a,*], Laure Apothéloz-Perret-Gentil[b], Maria Holzmann[b], Steffen Clauss[c], Eckhard Völcker[c], Jan Pawlowski[b]

[a] *Julianaweg 10, 1241VW Kortenhoef, Netherlands*
[b] *Dept. of Genetics and Evolution, University of Geneva, Quai Ernest Ansermet 30, CH-1211 Geneva 4, Switzerland*
[c] *Penard Labs, Karwendelstrasse 25, 12203 Berlin, Germany*

Received 12 April 2017; received in revised form 12 May 2017; accepted 16 May 2017
Available online 31 May 2017

Abstract

Most foraminifera inhabit marine habitats, but some species of monothalamids have been described from freshwater environments, mainly from Swiss water bodies over 100 years ago. Recent environmental DNA surveys revealed the presence of four major phylogenetic clades of freshwater foraminifera. However, until now only one of them (clade 2) has been associated to a morphologically described taxon—the family Reticulomyxidae. Here, we present morphological and molecular data for the genera representing the three remaining clades. We describe two new agglutinated freshwater genera from China and the Netherlands, *Lacogromia* and *Limnogromia*, which represent clades 3 and 4, respectively. We also report the first ribosomal DNA sequences of the genus *Lieberkuehnia*, which place this genus within clade 1. Our study provides the first morphotaxonomic documentation of molecular clades of freshwater foraminifera, showing that the environmental DNA sequences correspond to the agglutinated monothalamous species, morphologically similar to those described 100 years ago.
© 2017 Elsevier GmbH. All rights reserved.

Keywords: Freshwater foraminifera; *Lacogromia*; *Lieberkuehnia*; *Limnogromia*; Morphology; Phylogeny

Introduction

Foraminifera are unicellular eukaryotes characterized by the presence of granuloreticulopodia and the possession of a membranous, agglutinated, or calcareous test, which is either monothalamous (single-chambered) or polythalamous (multi-chambered) (Loeblich and Tappan 1987). Within monothalamids some species like *Reticulomyxa filosa* are amoeboid naked forms. Until 1859, foraminifera were only known from marine habitats, but that year Claparède and Lachmann described a monothalamid foraminifer, *Lieberkuehnia wageneri*, sampled from an unknown water body in Berlin. It had a smooth flexible test with an entosolenian tube that separated the main cytoplasm mass from the pseudopodial peduncle.

In 1886 Henri Blanc, a Swiss scientist, described another freshwater foraminifer, *Gromia brunneri*, which he had collected from the bottom of Lake Geneva. This single-chambered species had an agglutinated test, an organic layer covered and/or embedded with foreign, mainly non-organic,

*Corresponding author at: Julianaweg 10, 1241VW Kortenhoef, Netherlands.
E-mail address: ferry@arcella.nl (F. Siemensma).

http://dx.doi.org/10.1016/j.ejop.2017.05.006
0932-4739/© 2017 Elsevier GmbH. All rights reserved.

Table 1. Classifications of agglutinated freshwater allogromiids.

Rhumbler (1904)	De Saedeleer (1934)	Deflandre (1953)	Loeblich and Tappan (1960)
Rhynchogromia	***Allelogromia***	***Allelogromia***	***Saedeleeria***
- *linearis*	- *brunneri*	- *brunneri*	- *gemma*
- *nigricans*	- *nigricans*	- *nigricans*	
- *squamosa*	- *squamosa*	- *squamosa*	
	- *linearis*		
Diplogromia	***Diplogromia*** *emend.*	***Diplogromia***	***Diplogromia***
- *brunneri*	- *gemma*	- *gemma*	- *brunneri*
- *gemma*			- *squamosa*
			- *nigricans*
		Penardogromia	***Penardogromia***
		- *linearis*	- *linearis*
			- *palustris* (1961)
	G. saxicola[a]	*G. saxicola*[a]	*G. saxicola*[a]

[a]Not mentioned.

particles. In subsequent years, Eugène Penard, another Swiss protozoologist, described four similar species *Gromia gemma* and *G. squamosa* (1899), *G. linearis* (1902) and *G. saxicola* (1905) from the same lake. He also described *G. nigricans* (1902), which he found not far from Lake Geneva in Mategnin and a marsh near Rouelbeau. Penard made permanent preparations of these foraminifera, which are still preserved and available in the Penard Collection of the Natural History Museum of Geneva (Switzerland).

In 1904, Ludwig Rhumbler erected the subfamily Allogromiinae for monothalamous foraminifera characterized by a more or less flexible organic test wall commonly with one or rarely two terminal apertures at either end of the test. He included all described freshwater species in this taxon. In a recent higher ranked classification of foraminifera based on molecular phylogenies (Pawlowski et al. 2013), monothalamous foraminifera were considered as a paraphyletic group that contains agglutinated and organic walled species as well as "naked" amoeboid species and environmental clades with unknown morphological affinities.

Traditionally the organic-walled foraminifera are called allogromiids. Most of them are distributed over a wide range of marine and brackish habitats (Gooday 2002). Freshwater allogromiids with an agglutinated test were originally placed in the genus *Gromia* by their discoverers, but as its type species *G. oviformis* is a filose marine species, Rhumbler (1904) transferred three species (*G. squamosa*, *G. nigricans* and *G. linearis*) to *Rhynchogromia* Rhumbler 1894. He further erected a new genus, *Diplogromia*, for the other two species having a double test wall: *G. brunneri* and *G. gemma*, although without designing a type species for the genus (Table 1).

De Saedeleer (1934) revised Rhumbler's classification leaving *D. gemma* in its genus and creating a new genus *Allelogromia* for the *Rhynchogromia* species with *G. brunneri* as type species. Deflandre (1953) erected the genus *Penardogromia* for *G. linearis*, with the argument that it had a homogenous agglutinated test with calcareous particles. Loeblich and Tappan (1960) argued that the classification of De Saedeleer was unacceptable, because *G. brunneri* had been fixed as the type of *Diplogromia* by subsequent designation of Cushman (1928). They created the genus *Saedeleeria* for *G. gemma*, transferring *G. squamosa* and *G. nigricans* also to *Diplogromia*, but without giving any supporting explanations. Another agglutinated allogromiid, *Penardogromia palustris*, was described by Thomas (1961) from a freshwater marsh near Bordeaux (France).

Beside these descriptions there have been some scattered records of agglutinated freshwater allogromiids over the years (Grospietsch 1958; Hoogenraad and De Groot, 1940; Siemensma 1982; Wailes 1915; Meisterfeld pers. comm.; Clauss, unpublished) and some photomicrographs available online (Revello 2015; Protist Information Server 2016).

Leidy (1879) was the first who described an allogromiid foraminifer, *Gromia terricola*, from a terrestrial habit. He found this non-agglutinated species "among moist moss in the crevices of pavements, in shaded places, in the city of Philadelphia". A similar terrestrial organic walled allogromiid *Edaphoallogromia australica* has been described by Meisterfeld et al. (2001).

Apart from these agglutinated and organic-walled species, some naked amoeboid freshwater species belonging to the family Reticulomyxidae have been described. The best known of these species is *Reticulomyxa filosa* (Nauss 1949), long time considered as an amoebozoan, until its foraminiferal affinity was demonstrated by molecular study (Pawlowski et al. 1999). Since then two new species of Reticulomyxidae were described: *Haplomyxa saranae* (Dellinger et al. 2014) and *Dracomyxa pallida* (Wylezich et al. 2014).

In an attempt to rediscover the allogromiids described by Penard and Blanc, Holzmann and Pawlowski (2002) examined samples from Lake Geneva. They did not succeed in finding any specimens by microscopic observations. However, several foraminiferal DNA sequences were obtained from the same sediment samples that built a monophyletic clade with the marine genera *Ovammina* and *Cribrothalam-*

mina at its base. In a later report, numerous environmental rDNA sequences revealed the existence of a large number of freshwater monothalamids branching in several clades. However, none of these clades (except clade 2 that comprises the family Reticulomyxidae) could be linked to known freshwater allogromiids (Holzmann et al. 2003). Further studies based on environmental DNA surveys showed that foraminifera are also a ubiquitous component of soil samples (Geisen et al. 2015; Lejzerowicz et al. 2010).

Here, we describe two new agglutinated freshwater species (*Lacogromia cassipara* gen. nov., sp. nov. and *Limnogromia sinensis* gen. nov., sp. nov.). *Lacogromia cassipara* is commonly encountered in mesotrophic water bodies in the Netherlands. We collected specimens from different locations and found two morphotypes. The other species, *Limnogromia sinensis*, is an isolate from China. We compare both new species with those described by Blanc, Penard and Thomas, with reference to the slides of the Penard Collection in Geneva. In addition, we describe a *Lieberkuehnia* species based on cultured material and report the first DNA data for this species. Based on these data, we revise the taxonomy of agglutinated freshwater foraminifera and discuss their phylogeny and ecology.

Material and Methods

Sampling

Sediment samples containing morphotype A of *Lacogromia cassipara* collected weekly from March to May 2016 were taken from the bottom of a mesotrophic pond in the natural reserve Crailoo, 52°14′54.2″N 5°09′57.3″E (The Netherlands). A wide mouth pipette with an internal opening of 5 mm was used to collect the upper layer of the sediment from a depth of 30–40 cm. Every time a wide mouthed bottle was filled with 5 cm of sediment, transported to the lab and kept at room temperature on a windowsill on the north side. Small amounts of sediment were transported to 60 mm Petri dishes and examined with an inverted microscope. A Petri dish contained on average two specimens. 13 specimens were isolated with a micro pipette and kept in RNAlater® and over 220 specimens were isolated to be examined, measured and photographed with an upright microscope. A small number were kept in wet mounts in moisture chambers for observations.

One sample of morphotype B of *Lacogromia cassipara* was taken in April 2014 from a mesotrophic ditch in the natural reserve of Laegieskamp, 52°16′39.0″N 5°08′24.7″E (The Netherlands). The ditch had a thick layer of organic sediment. The upper layer of the sediment was collected from a depth of *c*. 20 cm also using a wide mouth pipette. 7 specimens were isolated and preserved in guanidine for subsequent DNA extraction. Over 100 specimens were examined, measured and photographed with an upright microscope.

A small sediment aliquot, <1 cc, with specimens of *Limnogromia sinensis*, was taken from sediment of a shallow pond in the city park of Yangshuo (China) on October 2015 (24°46′48.5″N 110°29′07.1″E) and kept for three weeks in a closed mini tube. We found 11 specimens, 7 of them were isolated, photographed and fixed in RNAlater®, one was prepared as type specimen and the others were used for light microscopic study.

The cultured specimens of *Lieberkuehnia* sp. came from the river Havel in Berlin (Germany).

Morphological analyses

Living specimens of *Limnogromia* and *Lacogromia* were filmed and photographed with a Canon D70 camera using an Olympus BX51 microscope with following objectives: 10XAPLN, 20 × 0.75 APO, 60 × 0.90 APO with correction collar and 100 × 1.30 oil, all with DIC. This equipment was also used for the slides of the Penard Collection. Adobe Photoshop was used for processing and measuring. For searching samples and for isolating specimens of both new allogromiid species a Leitz Diavert inverted microscope was used.

Living cells of *Lieberkuehnia* sp. were filmed and photographed with a Nikon TE2000U inverse microscope and a Jenaval microscope with DIC.

DNA extraction, amplification, cloning and sequencing

DNA was extracted using guanidine lysis buffer (Pawlowski 2000) for 22 specimens of *L. cassipara*, 4 specimens of *L. sinensis* and 16 specimens of *Lieberkuehnia* sp. DNA isolate numbers and accession numbers are given in Table 2. Semi-nested PCR amplifications of the 5′ terminal barcoding fragment of small-subunit (SSU) rDNA were performed using primer pairs s14F3 (acgcamgtgtgaaacttg)-sB (tgatccttctgcaggttcacctac) and 14F1 (aagggcaccacaagaacgc)-sB.

The amplified PCR products were purified using High pure PCR Purification Kit (Roche Diagnostics) cloned with the TOPO TA Cloning Kit (Invitrogen) following the manufacturer's instructions and transformed into competent *E. coli*. Sequencing reactions were performed using the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) and analyzed on a 3130XL Genetic Analyzer (Applied Biosystems).

Phylogenetic analysis

The obtained sequences were manually aligned to 65 other foraminiferal sequences (43 freshwater sequences and 22 marine sequences) using Seaview software (Gouy et al. 2010). After elimination of the highly variable regions, 721

Table 2. Isolate and accession numbers of sequenced freshwater foraminifera.

Species	Isolate	Accession numbers
Lacogromia cassipara	18849	LT576147–LT56154
	18990	LT576139
	18991	LT576155
	18992	LT576140
	18993	LT576141
	18994	LT576142
	18995	LT576156
	18996	LT576143
	18997	LT576157
	18998	LT576146–LT576158
	18999	LT576144
	19000	LT576159
	19001	LT576160
	19002	LT576145
	19179	LT604807
	19180	LT604808
	19181	LT604809
	19184	LT604813
	19185	LT604810
	19186	LT604811
	19188	LT604812
Limnogromia sinensis	18810	LT222211
	18811	LT222212–LT222213
	18812	LT222214–LT222216
	18813	LT222217–LT222219
Lieberkuehnia sp.	19189	LT604814
	19191	LT604815
	19192	LT604816
	19193	LT604817
	19194	LT604818
	19197	LT604819
	19198	LT604820
	19199	LT604821
	19200	LT604822
	19201	LT604823
	19202	LT604824
	19203	LT604825
	19205	LT604826
	19207	LT604827
	19208	LT604828
	19209	LT604829

sites were left for analysis. The phylogenetic tree was constructed with maximum likelihood method based on the GTR + G model with 1000 bootstrap replicates, using PhyML algorithms as implemented in the Seaview software.

We built a phylogenetic tree based on partial 18S rRNA with marine monothalamous foraminifera from several clades (Pawlowski et al. 2002) and environmental freshwater and soil sequences. Moreover, the sequences from two formerly described freshwater/soil species (*Reticulomyxa filosa* and *Edaphoallogromia australica*) were added to the analysis. The tree was arbitrarily rooted on monothalamous clades A–C.

Results and Discussion

Taxonomic descriptions

Supergroup Rhizaria Cavalier-Smith 2002

Phylum Foraminifera (D'Orbigny 1826)

Monothalamids (Pawlowski et al. 2013)

Clade 3

Lacogromia gen. nov.

Diagnosis: Test elongated to broadly pyriform or lens- or spindle-shaped, with a layer of small siliceous particles and commonly with some organic particles of debris. Test colourless or yellowish to almost black; aperture straight or oblique; test up to 1000 μm long. Generally with 1-8 nuclei, sometimes up to 30. Nuclei spherical, ovular. Peduncle and entosolenian tube asymmetrical.

Etymology: the prefix *Laco*, Latin for "pond", in reference to its freshwater habitat. The suffix—*gromia* refers to its relationship with *Allogromia*.

Type species: *Lacogromia cassipara*

New combinations:

Lacogromia squamosa (Penard, 1899) comb. nov.

Basionym *Gromia squamosa* Penard (1899)

Lacogromia brunneri (Blanc, 1886) comb. nov.

Basionym *Gromia brunneri* Blanc (1886); synonym *Gromia gemma* Penard (1899)

Lacogromia palustris (Thomas, 1961) comb. nov.

Basionym *Penardogromia palustris* Thomas (1961)

Lacogromia cassipara sp. nov. (**Figs. 1–4**)

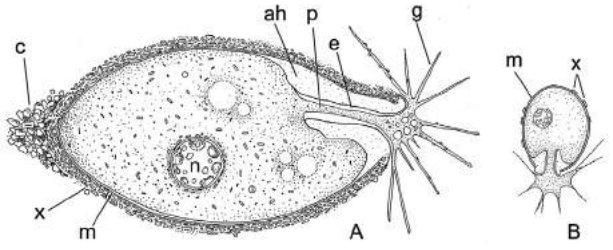


Fig. 1. General morphology of *Lacogromia cassipara*. (**A**) Adult cell. (**B**) Young cell, test *c*. 50 μm. Abbreviations: ah—apertural hyaloplasm; p—peduncle; e—entosolenian tube; g—granuloreticulopodia; n—nucleus; m—membrane; x—particles; c—cap of adhering bunch of particles.

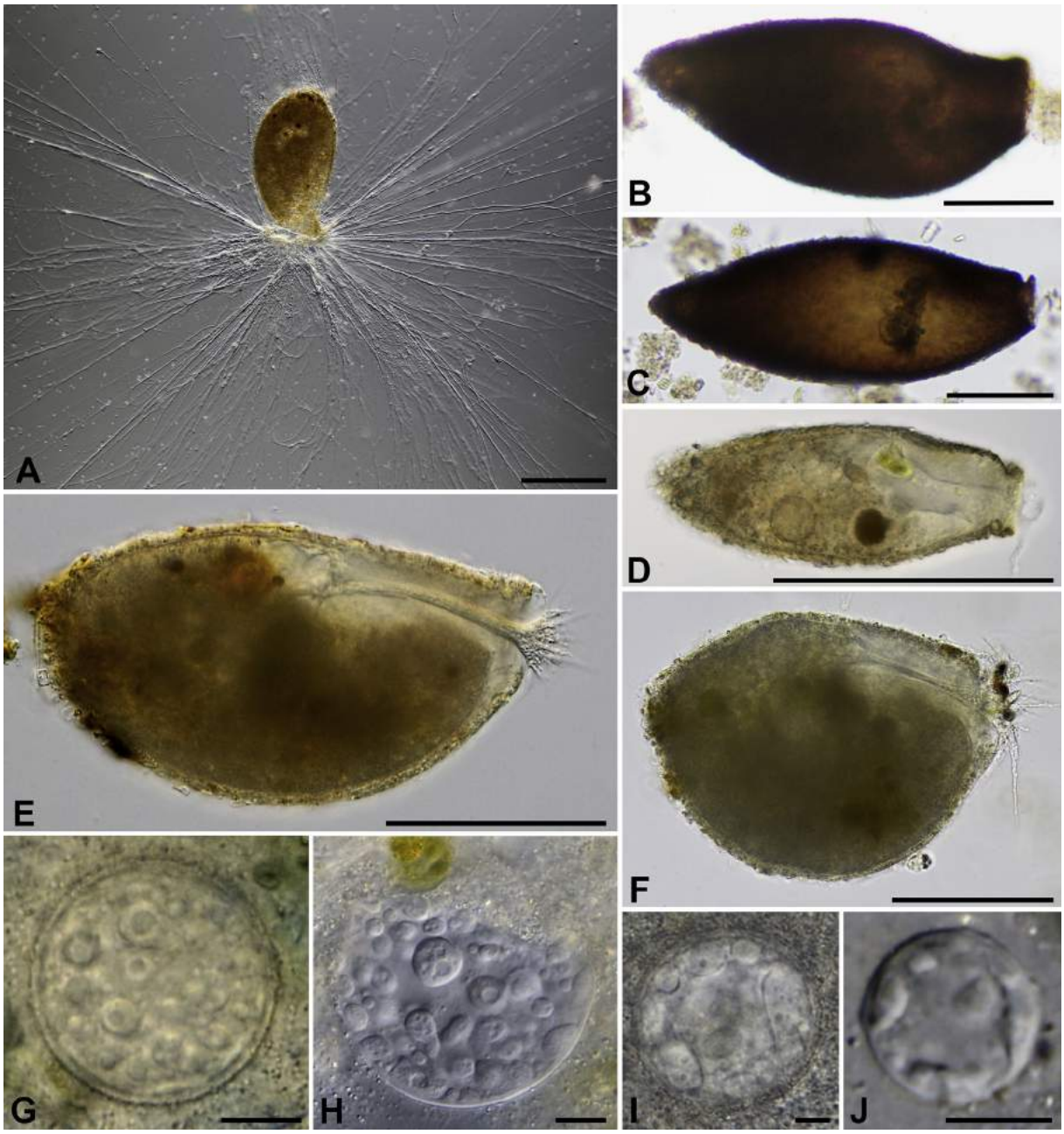


Fig. 2. *Lacogromia cassipara.* (**A**) Cell with fully employed granuloreticulopodium. (**B–D**) Tests of morphotype B. (**E, F**) tests of morphotype A. (**G–J**) Nuclei. Scale bars: (**A**) 200 μm, (**B–F**) 100 μm, (**G–J**) 10 μm.

Diagnosis: Test broadly ovoid to elongated pyriform, sometimes lens- or spindle-shaped, with a layer of small siliceous particles and generally with more or less organic particles from sediment. Test slightly flexible, colourless or light yellow, ochre, brown or almost black, 50–560 μm long; aperture oblique. Some specimens have a double ring around the aperture. Cell usually with 1–8 nuclei, sometimes up to 30. Nuclei spherical, with irregular but rounded pieces distributed throughout the nucleus with slightly more nucleoli in the periphery. No resting stages have been observed.

Etymology: *cassipara* is the Latin epitheton for "making a spider's web" that refers to the large web like granofilose reticulum of this species.

Type locality: Organic sediment, 40 cm deep, freshwater pond in the natural reserve Crailoo in the central area of the Netherlands, located at 52°14′54.2″N 5°09'57.3″E.

Type specimen: The type specimen has been deposited in the Natural History Museum of Geneva (holotype in alcohol nr. MHNG-INVE-97019; 3 paratypes in alcohol, nr. MHNG-INVE-97020 and 5 paratypes in slides, nr. MHNG-INVE-97021, embedded in HYDRO-Matrix®).

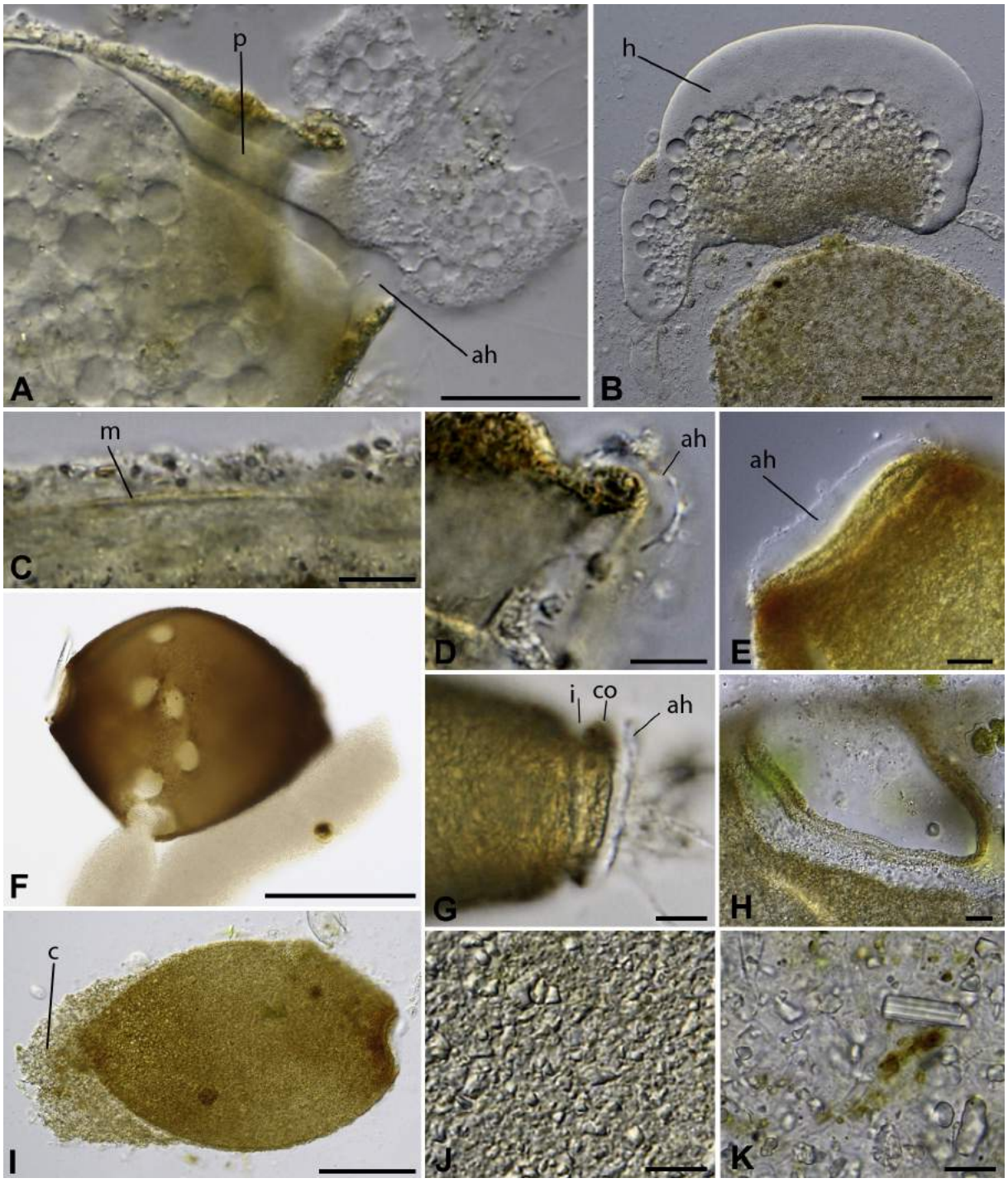


Fig. 3. *Lacogromia cassipara.* (**A**) Apertural region with peduncle and entosolenian tube; specimen strongly flattened, pressed by cover glass. (**B**) Mass of cytoplasm, pressed out of the test. (**C**) Test wall with layer of particles. (**D**) Detail of aperture; optical section with hyaline collar arrowed. (**E**) Apertural hyaline ring and double ring. (**F**) Empty test with holes, probably made by offspring; when pressed, fine granular cytoplasm streamed out. (**G**) Test with constriction behind collar of hyaline material. (**H**) Aperture with double ring, strongly flattened. (**I**) Flattened test with cap of adhering particles and double ring. (**J, K**) Detail of surface of a test. Abbreviations: ah—apertural hyaloplasm; c—cap of adhering particles; co—collar; h—hyaloplasm; i—constriction; p—peduncle; m—membrane. Scale bars: (**A**) 20 μm; (**B, F, I**) 100 μm; all other bars 10 μm.

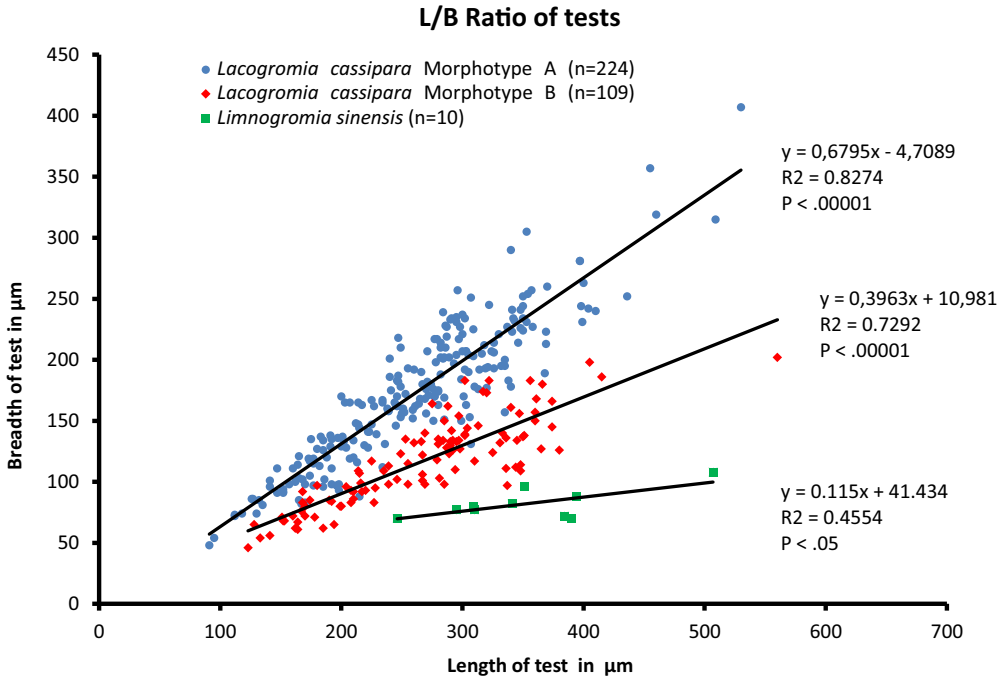


Fig. 4. Biometric analysis of the length/breadth ratio of tests of *Lacogromia cassipara*, morphotype A and B, and *Limnogromia sinensis*.

Description: The general shape and structure and the corresponding terminology of agglutinated allogromiids with *Lacogromia* as an example are summarized in Fig. 1.

The shape of the test is variable, ranging from broadly ovoid to elongated pyriform (Fig. 2A–F). Some, usually larger, specimens are rather subglobular (Fig. 2F), while other large specimens can have a more lens- or spindle-shaped outline (Fig. 2C). Smaller specimens, up to circa 160 µm, are always elongated ovoid (Fig. 2D).

The proximal end can be broadly rounded (Fig. 2A, E, F) or more conical (Fig. 2B, C). All tests are bilaterally symmetrical, usually with one side more curved than the other (Fig. 2B, E, F). Sometimes the less curved side bends slightly upwards towards the aperture (Fig. 2B, C, E). All tests are circular or nearly circular in cross section.

The test wall is a thin membrane, not always visible and usually colourless, more or less flexible and covered with a layer of very small, irregularly shaped, usually flattened, particles, mainly siliceous, but organic material may also be present (Fig. 3C, J–K). The agglutinated layer is about 4–8 µm thick with the proximal area usually being thicker. The size of these particles is variable (*c*. 1–3 µm). Size and density of the particles may vary per specimen (Fig. 3J–K). Some particles could be identified as fragments of diatom shells. All these particles are probably held together by a kind of cement.

Specimens that were kept for many weeks in petri dishes with a small layer of sediment, had a thinner layer of particles than freshly collected specimens, probably because building material became scarce. Because all particles are more or less of the same size, it is likely that the material is selected by the foraminifera.

Morphological variations: We found morphological differences between populations from different locations and consider them as different morphotypes (A and B). Cells of type A (Fig. 2A, E, F) look greyish or brownish grey when observed under transmitted light. The colour depends on the kind of food in the cytoplasm, the number of crystal-like particles and the colour of the agglutinated material in the test wall. Mineral material is commonly colourless, but organic particles are mostly ochre yellow, brown or black. Tests of type B vary in colour, those of younger, smaller specimens are light ochre yellow (Fig. 2D), and tests of older, larger specimens are darker ochre yellow or reddish brown and black (Fig. 2B, C). The colour is not always evenly distributed. Usually the proximal and apertural region are darker (Figs. 2C, 3F).

Another difference between the two types is the covering of the proximal part. Tests of morphotype B have an extra layer of loosely attached particles, resembling a kind of cap, while tests of morphotype A do not have any extra covering. Agglutinated particles of these caps are larger than the regular ones, up to 10 µm. Differences between both morphotypes are summarized in Table 3.

The length of all observed tests, both alive or empty, varied between 91 and 560 µm (mean 264 µm, std. dev. 77, n = 333), with a width of 48–407 µm (mean 154 µm). The average length/breadth ratio is 1.8, with extremes between 1.1 and 3.5. Biometrical analysis showed differences in this ratio between both morphotypes (Table 3, Fig. 4).

Aperture. The test has one circular aperture, commonly at its

Table 3. Morphological differences between morphotypes of *Lacogromia cassipara.*

L. cassipara	Morphotype A (n = 224)	Morphotype B (n = 109)
Aperture	No pronounced collar, smooth	Distinct collar with double ring, often with constriction
Shape	Broadly ovoid-pyriform, proximal end broadly rounded	Elongated ovoid-elongated pyriform, or spindle-shaped; proximal end conical, rounded
Proximal end	Without extra cap of particles	Usually with cap of larger particles
Structure	Particles loosely attached	Particles close to each other
Colour	Colourless or light ocre yellow	Dark brown, ocre yellow or black
L/B ratio	1.1–2.4, mean 1.5	1.7–3.5, mean 2.3
Length	91–530 μm, mean 262 μm	123–560 μm, mean 267 μm
Width	48–407 μm, mean 173 μm	46–202 μm, mean 117 μm
Nuclei, diameter	18–66 μm, mean 38.6 μm	8.7–77 μm, mean 29.0 μm

smallest end, and usually cut obliquely. The diameter of the aperture is highly variable per test, between 9 and 133 μm, mean 42 μm. Specimens of morphotype B have a double ring around the aperture (Fig. 3E, H, I), built of particles and commonly with a more or less clear constriction behind this collar (Figs. 2C, D, 3G). The second ring of this collar is a little broader than the first one and also more pointed in cross section.

The granular cytoplasm is separated from the aperture by an area of extremely hyaline material, resembling a pierced rubber stopper. This hyaline material, which we call here apertural hyaloplasm, is attached to the rim of the aperture (Fig. 3A, D, E, G). It is translucent and only detectable by small granules and bacteria attached to its surface (Fig. 3D). In tests of morphotype B, this hyaloplasm is attached to the second ring and in cross section visible as a clear curl (Fig. 3A, D).

The apertural hyaloplasm surrounds the entosolenian tube, which connects the granuloplasm with the surrounding environment. The narrow stream of cytoplasm flowing through this tube, the peduncle, is usually small in lateral view and broader in dorsal view. Sometimes two or more peduncles are present. The entosolenian tube is located eccentrically, usually on the less curved side, and becomes funnel-shaped towards the aperture, with the peduncle following its shape (Figs. 2E, 3A).

Although the almost featureless apertural hyaloplasm is difficult to detect visually, the presence within it of an entosolenian tube can be detected indirectly when larger particles are pushed through it, e.g. when the cell is pressed by the cover glass. In such a case, we observed that nuclei blocked the opening or passed the tube like a balloon which is pressed through a tube. The flexibility of the entosolenian tube could be observed when large food remnants were exported out of the cell. The same is true for phagocytose. Many cells contained food particles like rotifers and algae that were much larger than the diameter of the aperture and the entosolenian tube, so the cell must widen its aperture and entosolenian tube to engulf these large objects.

Based on our observations a cell can change the shape and amount of its apertural hyaloplasm dynamically. When a cell is disturbed it can decrease the amount of the hyaloplasm rather quickly.

Cytoplasm: The cytoplasm is granular with a large number of yellowish birefringent rod-like particles, probably crystals, about 1.3 μm long. One or more vacuoles of different size are present and smaller ones may fuse. We could not observe any contractile vacuole, probably because of the constantly moving plasm and the opaqueness of the test. When cytoplasm is pressed or squeezed out of the test, a zone of viscous hyaloplasm is formed together with a large number of non-contractile vacuoles (Fig. 3A, B). Pseudopodia are granuloreticulopodia with bidirectional streaming as is characteristic for foraminifera. They emerge from the peduncle.

Nucleus: About 26% of the cells (n = 333) had one nucleus, while the other cells had 2–8 nuclei. Except two cells which had over 20 and 30 nuclei respectively. The nuclei vary in diameter from 8.7–77 μm. Uninucleate cells have the largest nuclei while multinucleate cells have smaller ones, this probably corresponds to different life stages as described for *Allogromia laticollaris* in Parfrey and Katz (2010). The nucleoli are irregularly rounded and distributed throughout the nucleus with slightly more at the periphery (Fig. 2G–J). These nucleoli are about 1.4–14.6 μm in diameter. Large nucleoli may show one or more small lacunae (Fig. 2G–H). The amount of nucleoli in a nucleus may strongly differ per cell.

The nuclei are constantly rotating, with frequent changes in direction. In living cells, nuclei are difficult to observe in detail because of the opaqueness of the agglutinated wall. When nuclei are squeezed out of the test, they usually escape through the smaller entosolenian tube or a tear or rupture in the test and get damaged. Within a minute, the nucleoli disintegrate and a weakly granular nucleus remains.

Reproduction: We could not observe the complete life cycle, but did observe an isolated specimen that divided overnight in two daughter cells. In the past, we have observed schizogony with multiple fissions of a specimen that we now recog-

nize as *L. cassipara*. In this 'medium sized' specimen (about 250 μm long), we observed the large nucleus dividing into over 30 nuclei. The following day, 36 small daughter cells were observed around the empty test (Siemensma 1982). All daughter cells were about 50 μm long. They had a smooth membrane which became covered with particles during the next days (Fig. 1B).

In the recent samples about 28% of all observed tests were empty. Empty tests were on average larger, 317 μm in length, compared with 241 μm for living cells. Most empty tests showed holes in their wall, usually in the median area (Fig. 3F). We assume that these holes were made by offspring when leaving the test.

Phylogenetic position: Based on partial SSU rDNA sequences, *Lacogromia cassipara* branches within group 3 (Fig. 7). This clade is composed exclusively of environmental sequences obtained mainly from samples collected in Geneva basin.

Ecology: The observed population of *L. cassipara*, morphotype A, was present in the surface layer of organic sediment in a shallow mesotrophic freshwater pond. This pond is part of Zanderij Crailoo in the Netherlands, an area were sand had been excavated between 1870 and 1971. The area is fed by ground water and is now a natural reserve. Common amoeboid organisms in the sample were *Pelomyxa flava*, *Difflugia binucleata*, *D. pyriformis* and *Centropyxis ecornis*. Characteristic algae were *Micrasterias americana* and *M. rotata*.

Lacogromia cassipara feeds on diatoms (e.g. *Navicula* spp., *Diatoma vulgarum*, *Tabellaria* spp.), blue and green algae (*Ankistrodesmus* spp., *Phacus triqueter*, *Euglena acus*, *Cosmarium* spp.), filamentous algae (*Hyalotheka* spp.) and fungal spores. We also noticed rotifers and small testate amoebae (*Euglypha rotunda*, *Cryptodifflugia oviformis*) in cells and once a small nematode had been engulfed. Generally speaking one can say that *L. cassipara* feeds on anything it can get; it is omnivorous.

The observed population of morphotype B was isolated from a ditch in the nature reserve Laegieskamp, also in the Netherlands, and about 6 km away from Zanderij Crailoo, with similar environmental conditions. Both populations were discovered in early spring, specimens were abundant in April and disappeared end of May. Other findings of *L. cassipara* also come from shallow mesotrophic water bodies, like ditches in the Hol, Naardermeer and Westbroekse zodden, all old peat bogs in the central area of the Netherlands. It was also found in the flood plain of a small oligotrophic stream near Renkum, the Netherlands. Another location, which is also oligotrophic, is the Diepveen, a fen in the northern part of the Netherlands, 200 km distant from Zanderij Crailoo. However, our findings over the years are very scarce.

Remarks: *Lacogromia cassipara* resembles in its pyriform shape *Gromia brunneri*, but it differs from it in the structure of the nucleus and the much thinner test wall. It differs from *Gromia squamosa* in several aspects. *G. squamosa* is much larger, always spindle-shaped, with a thick layer of particles, and in cross section its test is more elliptical than circular and sometimes strongly compressed. Measured specimens from Penard's permanent slides show that *G. squamosa* has an average L/B ratio of 3.2 vs. 1.5 and 2.4 for morphotypes A and B of *L. cassipara* respectively. The structure of its nucleus is quite different from all other known freshwater allogromiids. It has an internal layer, by which the nucleus resembles "a very thick ring bordered on its inner contour with a clear, dark line (. . .) which consists of small elongated flakes" (Penard 1902), a phenomenon that has never been observed in *L. cassipara*. *Gromia nigricans*, *G. linearis* and *G. saxicola* differ from *L. cassipara* in having more elongated and much more flexible tubular tests. It differs from *P. palustris* in its general shape and the straight aperture of the latter.

Monothalamids (Pawlowski et al. 2013)

Clade 4

Limnogromia gen. nov.

Diagnosis: Test cylindrical tot elongated cylindrical, agglutinated, encrusted with a large number of small siliceous particles. Test very flexible, extendible and pliable. Up to 200 ovular nuclei. Peduncle and entosolenian tube asymmetrical.

Type species: *Limnogromia sinensis*

Etymology: the prefix *limnos* of the genus name refers to the freshwater habitat. The suffix—*gromia* refers to the relationship with *Allogromia*.

New combinations:

Limnogromia saxicola (Penard, 1905) comb. nov.

Basionym *Gromia saxicola* Penard (1905)

Limnogromia nigricans (Penard, 1902) comb. nov.

Basionym *Gromia nigricans* Penard (1902)

Limnogromia linearis (Penard, 1902) comb. nov.

Basionym *Gromia linearis* Penard (1902)

Limnogromia sinensis sp. nov. (**Fig. 5**)

Diagnosis: Test cylindrical, agglutinated, encrusted with a large number of small siliceous particles. Test very flexible, extensible and pliable; neck can bend very strongly and the proximal end can be stretched like a spine. Multinucleate, up to 200 nuclei; nuclei very small, usually spherical but sometimes ovoid with nucleolar material laying close to the nuclear membrane. Test 235–411 μm long (mean 345 μm) and 65–75 μm broad (n = 11); nuclei 6.0–8.2 μm in diameter.

Etymology: *sinensis* is a toponym with suffix—ensis which refers to the country of the type locality, China.

Type locality: 24°46′48.5″N 110°29′07.1″E, city park of Yangshuo, China (October, 2015).

Type material: The type specimen has been deposited

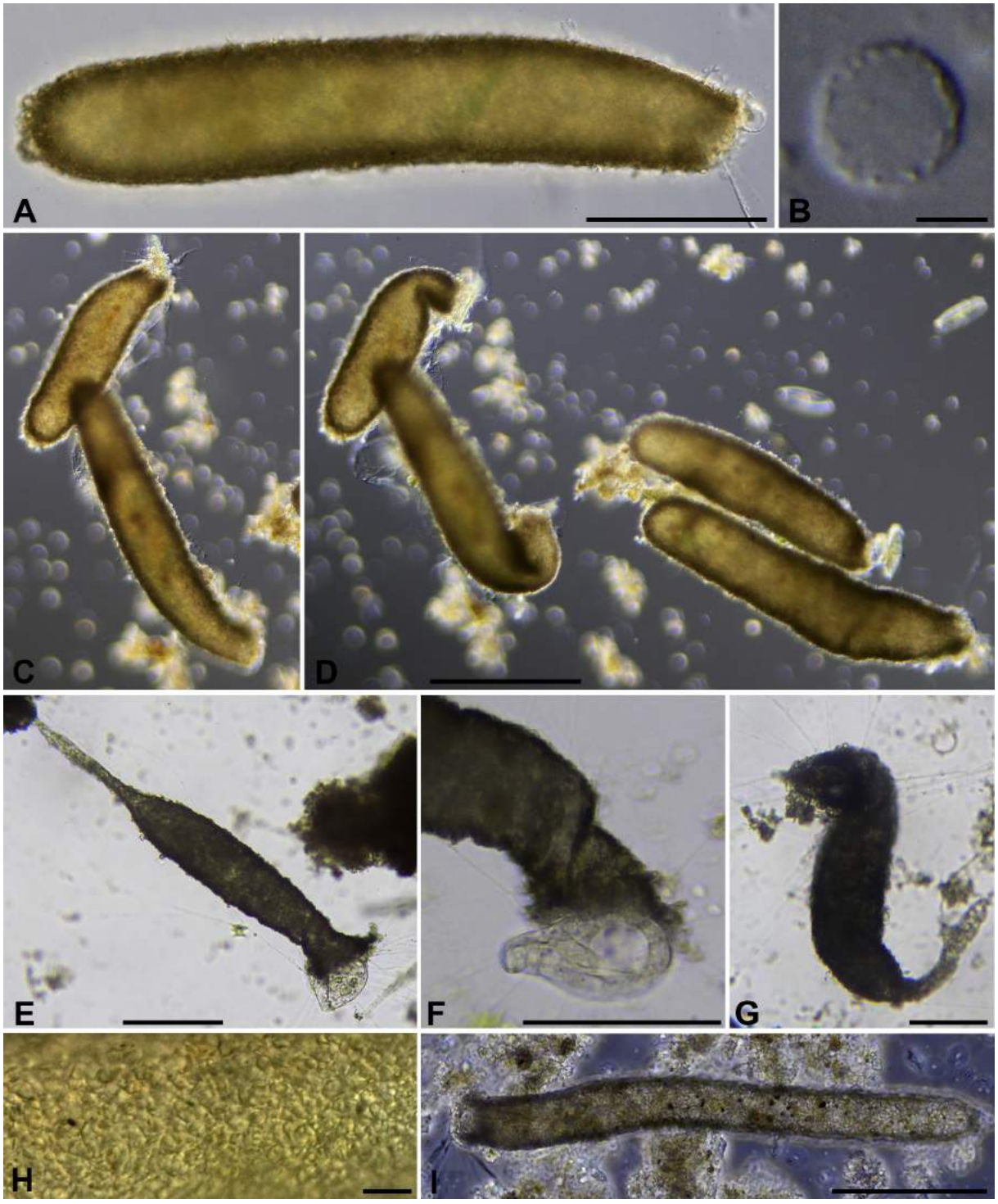


Fig. 5. *Limnogromia sinensis.* (**A**) Common morphology. (**B**) Nucleus. (**C, D**) Micrographs showing the flexibility of the neck. Three-minutes time lapse from C to D. (**E**) Specimen with elongated proximal end. (**F**) Same specimen with twisted and folded anterior part. (**G**) The same specimen, S-shaped. (**I**) Detail of test. (**H**) Unknown agglutinated freshwater species from Uruguay (photomicrograph Revello 2015). Scale bars: (**B**) 5 μm; H—10 μm; all other bars 100 μm.

in the Natural History Museum of Geneva (holotype in alcohol, nr. MHNG-INVE-97022; 3 paratypes in alcohol, nr. MHNG-INVE-97023).

Description: Cells of *L. sinensis* have a cylindrical yellowish to brownish test with an organic wall, encrusted with a large number of very small siliceous particles, lying closely packed together (Fig. 5H). Tests are 235–411 μm long (mean 345 μm) and 65–75 μm broad (n = 11). The L/B ratio

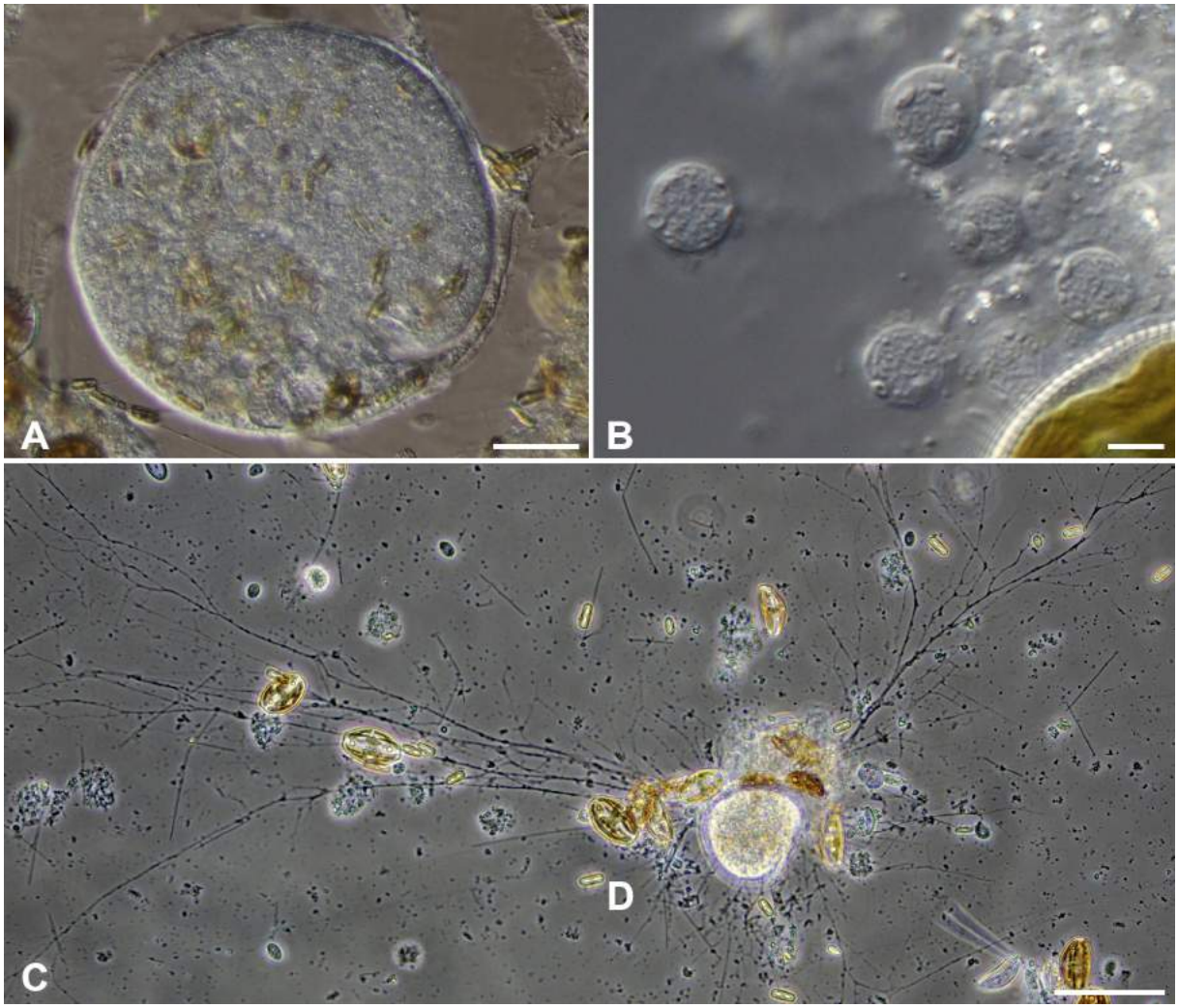


Fig. 6. *Lieberkuehnia* sp. (**A**) Common habitus. (**D**) Nuclei. (**B**) Cell with pseudopodial network. Scale bars: (**A**) 50 μm. (**B**) 10 μm. (**C**) 200 μm.

is 4.3 (3.5–5.6). Though the tests are tubular, they are not of equal width throughout. The area around the aperture is pliable and extensible and the neck can bend very strongly, through nearly 180° (Fig. 5C, D). On one occasion we observed a fold in the neck region indicating that the neck was twisted (Fig. 5F). The proximal end is usually rounded, but a specimen, kept in a petri dish for some weeks, showed an extensible proximal end which was pulled out far, shaped like a spine (Fig. 5E, G). The same specimen could also widen its aperture to resemble a funnel (Fig. 5E, F). One specimen was squeezed between cover and object glass, which caused most nuclei to be ejected. We counted up to 150 nuclei and estimated the total number around 200. The nuclei were 6.0–8.2 μm in diameter, usually spherical or ovoid, with small pieces of nucleolar material laying close to the nuclear membrane (Fig. 5B). No resting stages have been observed.

Phylogenetic position: *Limnogromia sinensis* branches within group 4, close to OTU22, an environmental sequence found in a river located in the Geneva basin (Fig. 7).

Ecology: We have only restricted observations of *L. sinensis* because of the small number of specimens we had. Observed food were diatoms and blue algae, but probably it is omnivorous.

Remarks: Besides the molecular data, its morphology characterizes *L. sinensis* as a new genus, and consequently new species. The overall shape and structure of the cell and the number of small nuclei are quite different from any other described freshwater foraminifer, except *G. saxicola* (Fig. 8G–J). Both species have a tubular but highly flexible test which can stretch, bend and twist and which can form trails with a viscous appearance, and an aperture that can be transformed into a funnel. Both species have up to 200 small nuclei. Nuclei of *L. sinensis* have a diameter of 6.0–8.2 μm, which is comparable to the measurements given by Penard (1905) for nuclei of *G. saxicola* (6–8 μm). However, nuclei in preserved cells of *G. saxicola* (slide 437 of the Penard Collection) are 3.3–4.5 μm in diameter in one cell and

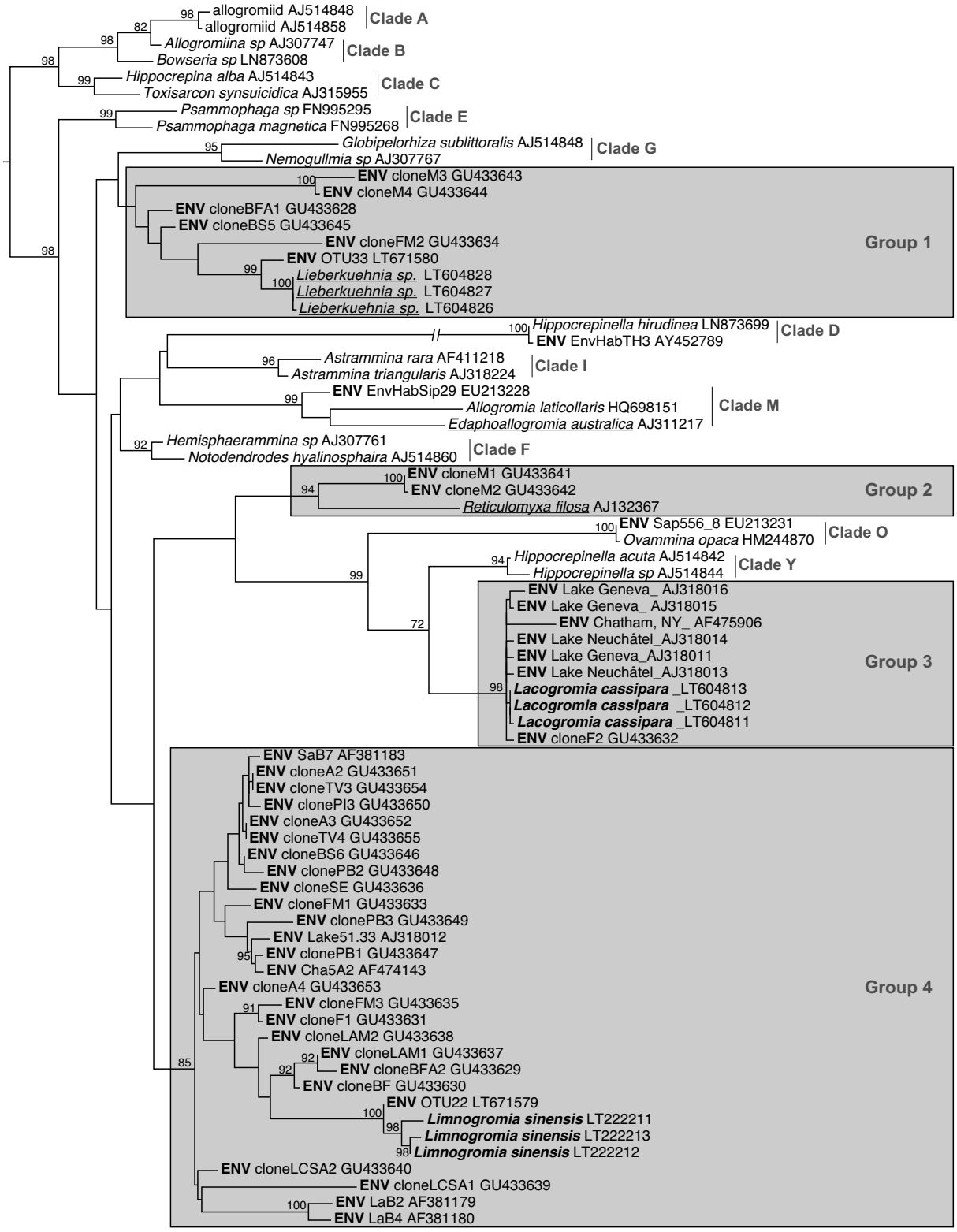


Fig. 7. Phylogenetic tree based on partial 18S rDNA of 74 sequences of foraminifera, including 44 environmental sequences (ENV) and 3 freshwater species (underline). Newly described species are highlighted in bold. Grey boxes correspond to the four freshwater clades. Only bootstrap values greater than 70% are shown.

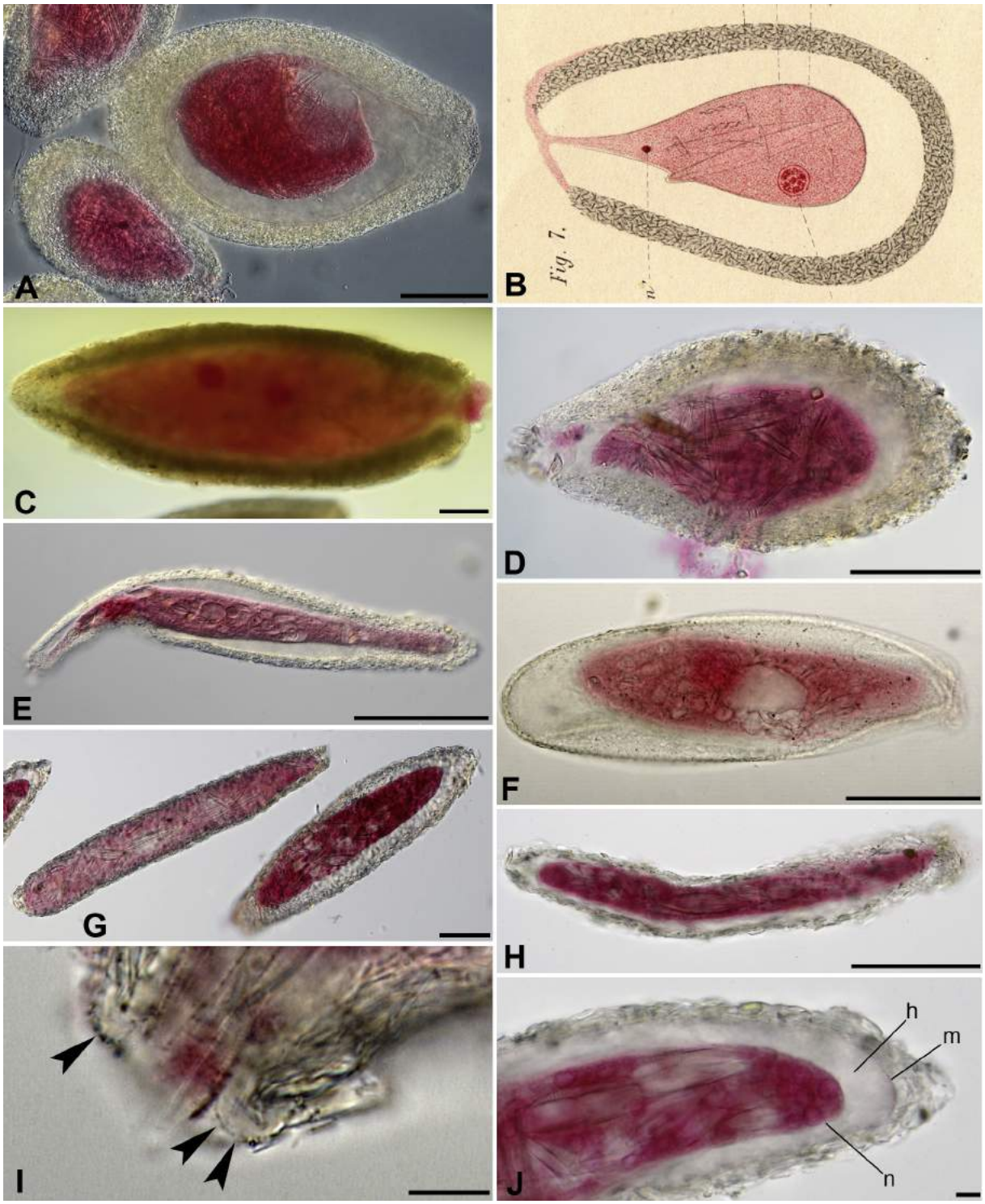


Fig. 8. (**A**) *Gromia brunneri.* (**B**) *Gromia brunneri*, in Blanc 1888. (**C**) *Gromia squamosa.* (**D**) *Gromia gemma.* (**E, F**) *Gromia linearis.* (**G–J**) *Gromia saxicola.* Hyaline collar arrowed. All images, except B, are from the Penard Collection in Geneva. According to Penard all specimens were treated with alcohol, stained with borax-carmine and embedded in Canada balsam. Abbreviations: h = hyaloplasm; m—cell membrane; n—nuclei. Scale bars: (**I, J**) 10 μm; all other bars 100 μm.

4.9–6.2 μm in another cell.

There are other differences. *G. saxicola* has a blackish test, according to Penard resembling a *Difflugia* species, while tests of *L. sinensis* are yellowish and smooth. *G. saxicola* was found at a depth of 20–40 m., while *L. sinensis* was isolated from very shallow water.

Morphologically *Limnogromia* and *Lacogromia* differ mainly in three characters: *Limnogromia* has a tubular test that is very flexible with up to 200 nuclei. *Lacogromia* has a more or less pyriform test, ranging from nearly circular to spindle-shaped, that is more or less flexible, and contains a cell body with rarely more than 30 nuclei.

An interesting observation appeared end of 2015 on YouTube, where Carlos Revello published a video of what seems to be an unknown *Limnogromia* species (Fig. 5I). It was found in a freshwater brook near San José, Uruguay (Revello, pers. comm.). Similar micrographs were published on the Japanese Protist Information Server (2016), showing specimens from the USA that also resemble *Limnogromia*. A flexible test has also been observed in some marine agglutinated monothalamids such as *Cedhagenia saltatus* (Gooday et al. 2010).

Monothalamids (Pawlowski et al. 2013)

Clade 1

Lieberkuehnia sp. (Fig. 6A–C)

Morphology: The genus *Lieberkuehnia* includes foraminiferal specimens with an ovoid or spherical flexible organic-walled test with a single aperture. An entosolenian tube built of hyaloplasm separates the main cytoplasm mass from the pseudopodial peduncle. The pseudopodial peduncle is at the origin of the pseudopodial network as well as a cytoplasm layer which surrounds the cell.

The cytoplasm of *Lieberkuehnia* sp. is colourless, yellowish, brownish or greenish, shows continuous cytoplasm streaming and contains over 100 nuclei and many vacuoles. Nuclei are granular with usually two or three relatively small rounded nucleoli with one or two lacunae each. Well-fed cells are filled with cytoplasm and also have a layer of hyaloplasm completely surrounding the tests. Starving cells often lack from that surrounding cytoplasm and sometimes it does not even fill the test completely. The test of a well-fed cell is not always easy to detect as the inner cytoplasm is difficult to differentiate from the one surrounding the cell. We have observed cells with a test size ranging from 50 μm to 300 μm. The pseudopodial network can be very large, extending some millimetres over the substrate. Sometimes the main cell body is covered by detritus.

Reproduction: We have observed two different modes of reproduction. Most often the main cell body divides into several cells (up to 5). Although each new cell has its own pseudopodial peduncle, young cells often share at least parts of the pseudopodial network. Sometimes a second form of reproduction was observed. Within the pseudopodial network a blob of plasma is formed. This blob then forms a new test and a new peduncle. Initially, the new cell is connected with the pseudopodial network of the old cell.

Phylogenetic position: *Lieberkuehnia* sp. branches within clade 1 in the SSU rDNA phylogenetic tree (Fig. 7). It is closely related to the OTU33 from Geneva basin. Other OTUs present in this clade have been reported from soil samples (Lejzerowicz et al. 2010).

Ecology: In our cultures *Lieberkuehnia* sp. fed mainly on diatoms and green algae.

Remarks: For the moment, we leave this *Lieberkuehnia* species in open nomenclature, because we are not fully convinced that it is identical to *L. wageneri* as described by Claparède and Lachmann (1859) and re-described by Penard (1907) and Mrva (2009). *Lieberkuehnia* sp. resembles *L. wageneri* in several aspects (general shape, entosolenian tube and pseudopodial peduncle), but also shows some different morphological features (structure of the nuclei, shape of the aperture, thickness of the test). We will address the species problem in this genus in a separate paper including additional sequences from different *Lieberkuehnia* strains, which we have in culture.

Taxonomic revision of some historical freshwater foraminiferal species and genera

The known freshwater agglutinated monothalamids share morphological similarities, which lead to some complications when reading the original descriptions and viewing Penard's slides as well as the illustrations made by Blanc (1888), Penard (1899, 1902, 1905) and Thomas (1961). Penard (1899), a careful observer and describer, already recognized the difficulty of differentiating between several species. Therefore the question is: how well defined are those classical species and genera?

Gromia brunneri Blanc 1886. Penard (1899) states that Blanc's *G. brunneri* might in fact represent three different species: *G. brunneri*, *G. gemma* and *G. squamosa*. However, based on the descriptions of Blanc (1886, 1888), we cannot agree with Penard. Blanc described the largest specimens of *G. brunneri* (500–1000 μm) as being ovoid to almost spherical, and the smallest specimens (200 μm) as spindle- or bottle-shaped. This description does not fit the features of *G. squamosa*, which is a large spindle-shaped species (Fig. 8C), up to 1000 μm. Morphotype A of *L. cassipara* is in this respect similar to Blanc's *G. brunneri*, with larger specimens being almost spherical and smaller specimens being spindle-shaped.

Gromia gemma Penard 1899. In the original description of *G. gemma*, Penard (1899) mentioned the thick internal mucous layer as an important character. In a later publication (Penard 1902), he remarked that this layer is not visible in living cells, but only in stained preparations. In 1905 Penard also observed such an internal mucous layer in *G. brunneri*. We were able to repeat his experiment of pressing cells out of their tests, but what Penard described as a mucous layer is, in our opinion, just a layer of viscous hyaloplasm (Fig. 3B). In a later description of *G. gemma*, Penard (1905) did not even mention this mucous layer, which should be so char-

acteristic. In summarizing the main differences between his *G. brunneri* and *G. gemma* he only mentioned the size of the test, the thickness of the test wall and the oblique aperture. According to Penard (1905) further differences concern the test wall that is much thinner in *G. brunneri* than in *G. gemma.* His 1902 illustration of *G. brunneri* shows an extremely thin wall, almost a membrane with some attached particles. However, the numerous specimens in his two slides labeled "*G. brunneri*", have a very thick test wall, between 20 and 77 μm (Fig. 8A). We compared the only specimen of *G. gemma* in the Penard Collection with those of *G. brunneri*, and found no significant difference (Fig. 8A, D). All these specimens are also very similar to the drawings given by Blanc (Fig. 8B). Penard (1905) remarked that *G. brunneri* and *G. gemma* might be one species, as he considered the three main differences mentioned above as not very important. Based on Penard's statement and our observations of his slides we consider *G. gemma* as a junior synonym of *G. brunneri*. Penard also supposed that *G. gemma* is an adult stage of *G. brunneri*, but that seems to be less likely considering the dimensions given by Blanc for *G. brunneri* (200–1000 μm) and those by Penard for *G. gemma* (200–600 μm). The 33 specimens of *G. brunneri* preserved in the Penard Collection measure 160–670 μm.

Gromia squamosa Penard 1899. In our opinion a well described species. Large and robust, spindle-shaped, with typically its broadest part at one third of the test measured from the aperture. Tests in the Penard Collection are 383–783 μm long (Fig. 8C).

Gromia linearis Penard 1902. Penard's description of *G. linearis* seems clear. Slide 433 of the Penard collection contains four specimens, all labeled "*G. linearis*" (Fig. 8E, F), but one of them is very different in shape and structure (Fig. 8F). It has a thin test wall and an elongated ovoid shape. The nuclei of the four specimens have the same structure but they most probably do not belong to the same species.

Gromia nigricans Penard 1902. This species resembles in its general shape *G. squamosa* and smaller specimens of *L. cassipara*, but differs strongly from both species by its highly flexible and pliable test, which resembles those of *G. linearis*, *G. saxicola* and *Limnogromia sinensis. G. nigricans* has also been found by Hoogenraad and De Groot (1940) and their observations correspond to those of Penard. The four specimens observed by Wailes (1915) and labeled *G. nigricans* represent probably a *Lacogromia* species.

Gromia saxicola Penard 1905. In our opinion a well described species, morphologically closely related to *Limnogromia sinensis*.

Penardogromia palustris Thomas 1961. According to Thomas (1961), the test is covered with calcareous particles, but we doubt if this is specific to this species and therefore a distinctive feature. In fact, we find the description of *P. palustris* insufficient to distinguish it from other related species. Though Thomas described the test as elongated tubular, he did not mention anything about the flexibility, extensibility and pliability of the test, which is so characteristic for tubular species. Based on the original drawing (Thomas 1961) the species resembles much more a small *Lacogromia* than a *Limnogromia* species. The test in this drawing (Thomas 1961) also resembles the deviating specimen in slide 433 of the Penard Collection (Fig. 8F).

Rhynchogromia Rhumbler 1894. Rhumbler transferred *G. squamosa*, *G. nigricans* and *G. linearis* to *Rhynchogromia* based on the assumption that the small particles in the test wall of these species are mainly secreted. However, he stated that there is an important difference between his *Rynchogromia variabilis* and the three *Gromia* species, because Penard and Blanc both described the test wall particles as siliceous plates and rods, while the particles of *R. variabilis* are not of siliceous origin. There is no reason to assume that the particles in the test walls of the three *Gromia* species are secreted. Firstly, neither Penard nor Blanc mentioned this option. Secondly, in the preserved *Gromia* specimens from the Penard collection, the small particles were comparable with those in the test wall of *L. cassipara*, including diatom frustules. Because the particles in all examined agglutinated freshwater foraminifera are true xenosomes, these species cannot be assigned to *Rhynchogromia*, as originally defined.

Diplogromia Rhumbler 1904. This genus is characterized by the presence of an internal mucous test wall. Its type species is *G. brunneri*, according to Loeblich and Tappan (1960), but this species does not have such a layer. What Blanc (1888) considered to be a second internal layer, is just the cell membrane, as is clearly visible in his drawings (Fig. 8B). Therefore we reject *Diplogromia* as a legal genus.

Allelogromia De Saedeleer 1934. The genus *Allelogromia* has been rejected by Loeblich and Tappan (1960) as being a junior synonym for *Diplogromia.*

Penardogromia Deflandre 1953. This genus was designed by Deflandre for species with a homogenous agglutinated test with calcareous particles, similar to some tests of agglutinated miliolids. He based the introduction of this new genus on his observations of Penard's slide of *G. linearis* in polarized light, but without giving any additional information. We also observed Penard's slides in polarized light, but did not find any significant difference between the material in the tests of all preserved species. According to Penard (1902) the test of *G. linearis* is comparable with those of *G. squamosa* and *G. brunneri*. We agree with Penard and therefore we do not accept this genus.

Saedeleeria Loeblich and Tappan 1960. This genus was designed for *G. gemma*, but as we consider this species as a junior synonym of *G. brunneri*, it is rejected.

General remarks on morphology, ecology, and taxonomy of freshwater agglutinated foraminifera

Morphology: This is the first time that both morphological and molecular data for agglutinated foraminiferal freshwater species could be acquired and used to revise the taxonomy of this poorly known group. The obtained results allow an increased understanding of the morphological variation within the different freshwater foraminiferal clades. Both new described species closely resemble in their general morphology the classical ones described by Blanc, Penard and Thomas. All species have an agglutinated test with an entosolenian tube and a peduncle. Though an entosolenian tube has only been described for *G. gemma* by Penard (1899), we could also detect it in two stained specimens of Penard's slides: in *G. brunneri* and *G. saxicola* (Fig. 8I), where small particles attached to the surface of the apertural hyaloplasm made the tube visible, just as in *L. cassipara* (Fig. 3D). Because all known species have the same overall structure, we assume that all classical species have such a tube. Penard (1902) described how difficult it is to detect this tube, because the surrounding material is "as clear as water" as we confirmed. He also noted that the tube is only visible in stained preparations and never in living cells. Blanc (1888) remarked that *G. brunneri* does not have such a tube, but that is unlikely, given the presence of a tube in his drawing of this species (Fig. 8B). In the same publication he mentioned the opaqueness of the test that prevented any clear observation and which might be the reason why he was not able to detect a tube.

The function of the entosolenian tube might be to protect the cell against penetration by predators and/or parasites, comparable with the diaphragms and/or narrow apertures in some testate amoebae, e.g. *Lesquereusia*, *Zivkovicia* and *Cucurbitella*, which prevents rotifers from laying their eggs inside (De Smet 2006).

With the exception of *L. sinensis* and *G. saxicola*, all known agglutinated freshwater foraminifera are mononucleate, having one large nucleus, usually 60–77 μm in diameter, or multinucleate, with smaller nuclei, usually 2–8 but sometimes more than 30 in number. Only *L. sinensis* and *G. saxicola* have a large number of small nuclei, up to 200. The number of nuclei in a cell might be related to different life stages as has been described for some other monothalamids (Goldstein and Barker 1990; Parfrey and Katz 2010). Due to the limited number of specimens available for observation, we cannot exclude that *L. sinensis* and *G. saxicola* also possess mononucleated specimens. Comparing the nuclei of the two newly described species with those preserved on slides is also difficult as nuclei disintegrate rapidly once removed from the cytoplasm. Penard squeezed tests to get the nuclei out of it and also stained and observed them, so we do not know if damaged ones have been described.

Differences between both morphotypes in *L. cassipara* could be induced by environmental factors; for example, the amount of iron could affect the colour of the test, as has been described for *Gromia oviformis* (Hedley 1960).

Ecology: Freshwater foraminifera seem to be rare, given the very scarce microscopic records over the years. However, molecular data show a rich diversity of freshwater and soil foraminiferans (Holzmann and Pawlowski 2002; Holzmann et al. 2003; Lejzerowicz et al. 2010). The close relationship between *Lieberkuehnia* sp. and *L. sinensis* with environmental sequences (OTU33 and OTU22) suggests that the same morphotypes might also live in the Geneva basin . Members of clades 3 and 4, represented by *Lacogromia* and *Limnogromia*, respectively, seem to be present in all types of habitats tested molecularly (lake, small and big river, pond, soil) in the Geneva area. Groups 3 and 4 are represented by more sequences than groups 1 and 2 (Apothéloz-Perret-Gentil, unpublished), which suggests that the species described by Penard might still occur in the Geneva basin.

Taxonomy: Based on molecular phylogenetic data, we could place a morphologically described species in each of the major freshwater foraminiferal clades. *Lieberkuehnia* clusters with clade 1, *Reticulomyxa* with clade 2, *Lacogromia* is a member of clade 3 and *Limnogromia* is a representative of clade 4. As none of the classical genera are well established, we transfer the classical species to either *Lacogromia* (*G. brunneri*, *G. squamosa* and *P. palustris*) or *Limnogromia* (*G. linearis*, *G. saxicola* and *G. nigricans*) As criteria, we choose the flexibility and shape of the test. We are aware that our choice is arbitrary, but for the moment it is the only useful morphological character.

Acknowledgements

We thank Norbert Hülsmann for comments and advice and Enrique Lara for taking care of the first isolates of *L. cassipara*. We would also like to thank André Piuz who provided us with several allogromiid slides from the Penard collection of the Museum of Natural History, Geneva. This study was supported by the Swiss National Science Foundation grants 31003A-140766 and 313003A-159709, and the G. and A. Claraz Donation.

References

Blanc, H., 1886. Un nouveau foraminifère de la faune profonde du Lac. Biblio. Universelle (Arch. Sci. Phys. Nat.) ser. 3 16, 362–366.

Blanc, H., 1888. La *Gromia brunnerii* un nouveau foraminifère. Recueil Zoologique Suisse 4, 498–513.

Cavalier-Smith, T., 2002. The phagotrophic origin of eukaryotes and phylogenetic classification of Protozoa. Int. J. Syst. Evol. Microbiol. 52 (2), 297–354.

Claparède, É., Lachmann, J., 1859. Études sur les infusoires et les rhizopodes. Mémoires de l'Institut National Genevois 6, 264–466, Pl. 25.

Cushman, J.A., 1928. Foraminifera, their classification and economic use. Cushman Lab. Foram. Res. Spec. Publ. 1, 1–40.

D'Orbigny, A., 1826. Tableau methodique de la classe des Céphalopodes. Annales des Sciences Naturelles 7, 245–314.

De Smet, W.H., 2006. Rotifers inhabiting shells of testate amoebae. In: Book of Abstracts, International Symposium on Testate Amoebae, Antwerp, pp. 44–45.

Deflandre, G., 1953. Thécamoebiens, In Grassé, P.P., Traité de Zoologie, 1(2), Masson et Cie, Publ. 97–148.

Dellinger, M., Labat, A., Perrouault, L., Grasse Grellier, P., 2014. *Haplomyxa saranae* gen. nov. et sp. nov., a new naked freshwater foraminifer. Protist 165, 317–329.

De Saedeleer, H., 1934. Beitrag zur Kenntnis der Rhizopoden: morphologische und systematische Untersuchungen und ein Klassifikationsversuch. Mém. Mus. Roy. d'Hist. Nat. Belg. (Bruxelles) 60, 1–112.

Geisen, S., Tveit, A.T., Clark, I.M., Richter, A., Svenning, M.M., Bonkowski, M., Urich, T., 2015. Metatranscriptomic census of active protists in soils. ISME J. 9, 2178–2190.

Goldstein, S.T., Barker, W.W., 1990. Gametogenesis in the monothalamous agglutinated foraminifer *Cribrothalammina alba*. J. Protozool. 37, 20–27.

Gooday, A.J., 2002. Organic-walled allogromiids: aspects of their occurence, diversity and ecology in marine habitats. J. Foraminifer. Res. 32, 384–399.

Gooday, A.J., Anikeeva, O.V., Pawlowski, J., 2010. New genera and species of monothalamous foraminifera from Balaclava and Kazach'ya Bays (Crimean Peninsula, Black Seas). Mar. Biodivers. 41 (4), 481–494, http://dx.doi.org/10.1007/s12526-010-0075-7.

Gouy, M., Guindon, S., Gascuel, O., 2010. Seaview version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol. Biol. Evol. 27, 221–224.

Grospietsch, T., 1958. Wechseltierchen (Rhizopoden). Sammlung: Einführung in die Kleinlebewelt. Kosmos, Stuttgart, Tafel III.

Hedley, R.H., 1960. The iron-containing shell of *Gromia oviformis* (Rhizopoda). J. Cell Sci. s3–s101, 279–293.

Hoogenraad, H.R., De Groot, A.A., 1940. Zoetwaterrhizopoden en –Heliozoën. Fauna van Nederland. Sijthoff, Leiden, pp. 198–205.

Holzmann, M., Pawlowski, J., 2002. Freshwater foraminiferans from Lake Geneva: past and present. J. Foraminifer. Res. 32, 344–350.

Holzmann, M., Habura, A., Giles, H., Bowser, S.S., Pawlowski, J., 2003. Freshwater foraminiferans revealed by analysis of environmental DNA samples. J. Eukaryot. Microbiol. 50, 135–139.

Leidy, J., 1879. Freshwater Rhizopods of North America. U.S. Geol. Survey, 12. Washington, pp. 277–280.

Lejzerowicz, F., Pawlowski, J., Fraissinet-Tachet, L., Marmeisse, R., 2010. Molecular evidence for widespread occurence of foraminifera in soils. Environ. Microbiol. 12, 2518–2526.

Loeblich, A.R., Tappan, H., 1960. *Saedeleeria*, new genus of the family Allogromiidae (Foraminifera). Proc. Biol. Soc. Washington 73, 195–196.

Loeblich, A.R., Tappan, H., 1987. Foraminiferal Genera and their Classification, vol. 1–2. Van Nostrand Reinhold, NY, pp. 1694.

Meisterfeld, R., Holzmann, M., Pawlowski, J., 2001. Morphological and molecular characterization of a new terrestrial allogromiid species: *Edaphoallogromia australica* gen. et spec. nov. (Foraminifera) from Northern Queensland (Australia). Protist 152, 185–192.

Mrva, M., 2009. Re-discovery of *Lieberkuehnia wageneri* Claparède *et* Lachmann 1859 (Rhizaria, Foraminifera): taxonomical and morphological studies based on a Slovak population. Acta Protozool. 48, 111–117.

Nauss, R.N., 1949. *Reticulomyxa filosa* gen. et sp. nov., a new primitive plasmodium. Bull. Torrey Bot. Club 76, 161–173.

Parfrey, L.W., Katz, L.A., 2010. Genome dynamics are influenced by food source in *Allogromia laticollaris* strain CSH (Foraminifera). Genome Biol. Evol. 2, 678–685.

Pawlowski, J., Bolivar, I., Fahrni, J.F., De Vargas, C., Bowser, S.S., 1999. Molecular evidence that *Reticulomyxa filosa* is a freshwater naked foraminifer. J. Eukaryot. Microbiol. 46, 612–617.

Pawlowski, J., 2000. Introduction to the molecular systematics of foraminifera. Micropaleontology 46 (Suppl. 1), 1–112.

Pawlowski, J., Holzmann, M., Fahrni, J.F., Cedhagen, T., Bowser, S.S., 2002. Phylogenetic position and diversity of allogromiid Foraminifera inferred from rRNA gene sequences. J. Foraminifer. Res. 32, 334–343.

Pawlowski, J., Holzmann, M., Tyszka, J., 2013. New supraordinal classification of Foraminifera: molecules meet morphology. Mar. Micropaleontol. 100, 1–10.

Penard, E., 1899. Les rhizopodes de faune profonde dans le lac Léman. Rev. Suisse Zool. 7, 82–99.

Penard, E., 1902. Faune Rhizopodique du Basin de Lac Léman. Kündig Publ., Genève, pp. 551–570.

Penard, E., 1905. Les Sarcodinés des grands lacs. Kündig Publ., Genève, pp. 68–81.

Penard, E., 1907. Recherches biologiques sur deux *Lieberkühnia*. Arch. Protistenkd. 8, 225–258.

Protist Information Server, 2016. http://protist.i.hosei.ac.jp/PDB/Images/Sarcodina/Allelogromia/index.html.

Rhumbler, L., 1894. Beiträge zur Kenntnis der Rhizopoden. II. Zeitsch. F. wiss. Zool. 57, 587–616.

Rhumbler, L., 1904. Systematische Zusammenstellung der recenten Reticulosa. Arch. Protistenkd. 3, 181–294.

Revello, C.G., 2015. https://www.youtube.com/watch?v=nUl6JfpJZH

Siemensma, F.J., 1982. Schaalamoeben. Natura, 4. KNNV, Hoogwoud, pp. 95–106.

Thomas, R., 1961. Note sur quelques Rhizopodes de France, Cahiers des Naturalistes. Bull. NP 17, 77–79.

Wailes, G.H., 1915. The British Freshwater Rhizopoda and Heliozoa, vol. III. Ray Society, London, pp. 127–143.

Wylezich, C., Kaufmann, D., Marcuse, M., Hülsmann, N., 2014. *Dracomyxa pallida* gen. et sp. nov.: a new giant freshwater foraminifer, with remarks on the taxon Reticulomyxidae (emend.). Protist 65, 854–869.

Environmental Science & Technology

Article

pubs.acs.org/est

Environmental Monitoring: Inferring the Diatom Index from Next-Generation Sequencing Data

Joana Amorim Visco,[†] Laure Apothéloz-Perret-Gentil,[†] Arielle Cordonier,[‡] Philippe Esling,[†,§] Loïc Pillet,[†,∥] and Jan Pawlowski*[,†]

[†]Department of Genetics and Evolution, University of Geneva, Boulevard d'Yvoy 4, CH 1205 Geneva, Switzerland

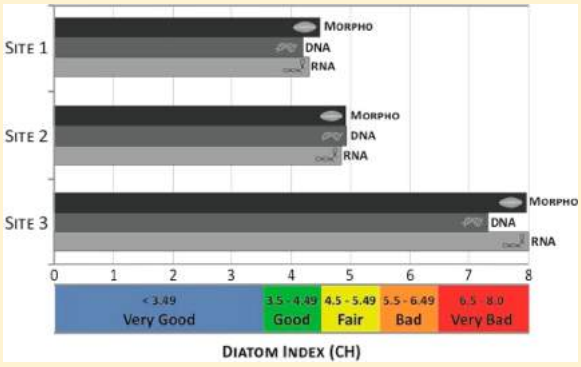
[‡]Water Ecology Service, Department of Environment, Transports and Agriculture, Canton of Geneva, CH 1205 Geneva, Switzerland

[§]IRCAM, UMR 9912, Université Pierre et Marie Curie, 4 place Jussieu, 75005 Paris, France

[∥]CNRS, UMR 7144, Laboratoire Adaptation et Diversité en Milieu Marin, Place Georges Teissier, CS90074, 29688 Roscoff, France

Supporting Information

ABSTRACT: Diatoms are widely used as bioindicators for the assessment of water quality in rivers and streams. Classically, the diatom biotic indices are based on the relative abundance of morphologically identified species weighted by their autoecological value. Obtaining such indices is time-consuming, costly, and requires excellent taxonomic expertise, which is not always available. Here we tested the possibility to overcome these limitations using a next-generation sequencing (NGS) approach to identify and quantify diatoms found in environmental DNA and RNA samples. We analyzed 27 river sites in the Geneva area (Switzerland), in order to compare the values of the Swiss Diatom Index (DI-CH) computed either by microscopic quantification of diatom species or directly from NGS data. Despite gaps in the reference database and variations in relative abundance of analyzed species, the diatom index shows a significant correlation between morphological and molecular data indicating similar biological quality status for the majority of sites. This proof-of-concept study demonstrates the potential of the NGS approach for identification and quantification of diatoms in environmental samples, opening new avenues toward the routine application of genetic tools for bioassessment and biomonitoring of aquatic ecosystems.



INTRODUCTION

Diatoms are phototrophic protists common in all aquatic ecosystems and widely used as bioindicators of environmental conditions, particularly in rivers and streams.[1,2] The applications of diatoms as bioindicators range from routine monitoring of water quality to the assessment of industrial pollution impact.[3−6] Because diatoms are highly sensitive to environmental conditions and grow rapidly, they respond quickly to changes in chemical, physical, or biological factors. Hence, analyzing the composition of their communities provides an easy method to detect environmental changes due to natural or anthropogenic causes.

Various biotic indices have been developed to assess environmental impact using diatoms.[7] Most of these indices are based on the relative frequency of species weighted by their autoecological value and eventually other index-specific factors. In Europe, the Water Framework Directive[8] recommends using diatoms to assess water quality, but the computation of diatom indices vary from one country to another.[2] In Switzerland, the Swiss Diatom Index (DI-CH) was proposed in order to characterize the biological status of rivers and streams using the frequencies and distributions of more than 400 diatom species and morphological varieties.[9] The DI-CH classifies watercourses into 5 categories, corresponding to *very good*, *good*, *average*, *poor*, and *bad* degree of pollution, as established by the Swiss Federal Council in the Waters Protection Ordinance.[10]

The DI-CH is calculated as follows

$$\text{DI-CH} = \frac{\sum_{i=1}^{n} D_i G_i H_i}{\sum_{i=1}^{n} G_i H_i}$$

where D_i is the factor based on the autoecological value for taxon i, G_i is the weighting factor for taxon i, H_i is the relative frequency of taxon i in a studied sample (number of valves found for the taxon i divided by the total number of valves counted), and n is the total number of taxa found in a sample.

The main limitation of all other diatom indices is related to the species identification being based on morphology. Indeed, diatoms constitute one of the most specious groups of protists, with the number of species estimated at nearly 200 000.[11] However, most freshwater diatoms are small (usually <50 um),

Received: December 18, 2014
Revised: June 2, 2015
Accepted: June 8, 2015
Published: June 8, 2015

ACS Publications © 2015 American Chemical Society

DOI: 10.1021/es506158m
Environ. Sci. Technol. 2015, 49, 7597−7605

and their microscopic identification requires special sample preparation methods and expert taxonomic knowledge. The size, shape, and design of diatom valves are the main features used for taxonomic identification of diatom species. Yet, intraspecific variability can be very high, and some morphological characters can become indistinct as a result of size reduction during the life cycle. In some cases, the morphological differences between species are so subtle that even trained taxonomists may come to different conclusions.[12]

Over the past decade, molecular barcoding has become widely recognized as an efficient tool for species identification. This approach is based on the assumption that a short DNA sequence (DNA barcode) contains enough information to distinguish species. The main advantage of using DNA barcodes in applied studies is that standardization and automation of the protocols is easier than that in the traditional morphology-based approach. Several diatom barcoding studies have been performed based mainly on the analysis of five genes: cox1,[13,14] the rbcL gene,[15,16] the ITS region,[17,18] the V4 region of the 18S rDNA,[19,20] and the D2/D3 region of the LSU rRNA gene.[15] Although there is no consensus on the ideal diatom DNA barcode, it has been proposed that some highly discriminating barcodes (ITS, cox1) are more suitable for taxonomic studies, whereas those that are less variable but more universal (18S, rbcL) are more appropriate for applied studies.[12]

Recent developments of next-generation sequencing (NGS) technologies offer the possibility to use molecular barcoding for fast and reliable diversity surveys based on environmental samples. NGS-based environmental monitoring has been proposed as a time and cost-effective alternative to the traditional morphology-based approaches.[21−23] Several experimental studies have been conducted on NGS-based inventories of freshwater benthic macroinvertebrates.[24−26] The major gaps highlighted by these studies include the incompleteness of the database, the technical biases, and the irrelevance of NGS quantitative data as compared to the abundance of specimens. Previous studies focusing specifically on diatoms completed their taxonomic reference database, evaluated different DNA barcodes, and compared the composition of diatom communities inferred from microscopic and NGS data.[27−30] One of these studies also briefly compared the diatom indices computed from morphological and molecular data,[28] although presently this aspect has still not been thoroughly examined.

Here, we test the hypothesis that the use of NGS could lead to a similar assessment of the water quality as the morphological study. To do so, we analyze the diatom communities in 27 watercourses of the Geneva basin, using the hypervariable region V4 of 18S rDNA as the diatom DNA barcode and the Illumina Miseq platform for high-throughput sequencing. Assuming that the RNA provides a better proxy for active cells, we compare the DNA and RNA data for the relative abundance of each taxon in order to test which ones fit better to the morphological data. Finally, we compute the DI-CH values for each site and compare them with the values inferred from microscopic study. We analyze the congruence between NGS and morphological analyses and discuss the current limitations of NGS approach that should be overcome to reduce the divergence between molecular and morphological indices.

MATERIALS AND METHODS

Sampling. The samples were collected in 2013−14 as part of a routine bioassessment campaign performed by the Service of Water Ecology (SECOE) of the Department of Environment, Transport, and Agriculture in Geneva, Switzerland.[31] The biofilm containing epilithic diatoms was collected from 27 sites located in shallow waterways of the Geneva basin following the directives established by the Swiss Federal Office for the Environment[9] (Supporting Information, SI, Table S1). Between three to five stones were selected at each sampling site. The periphyton taken by scratching the stones with diatom-scraping devices was resuspended with freshwater taken from the river and then transferred to sampling bottles. Each sample was homogenized and divided into two subsamples, one for morphological analysis by SECOE and the other for molecular analysis. Morphological samples were preserved in a concentrated (37%) formaldehyde solution, while molecular samples were kept cold (ca. 0 °C) during sampling (max. Four hours). Upon arrival to the laboratory, 1 mL of homogenized periphyton suspension was transferred to 1.5 mL tubes and centrifuged at 8000*g* for 10 min. The supernatant was discarded and the pellets stored at −80 °C until DNA/RNA extractions.

Morphological Analysis. Sample preparation, species identification, counting, and DI-CH calculations were performed as recommended by the Swiss Federal Office for the Environment.[9] Periphyton suspensions were sorted, and undesirable material was discarded. A decarbonation step using hydrochloric acid was performed, followed by the elimination of organic material by calcination combined with a treatment with hydrogen peroxide. Diatoms were then washed and mounted in Naphrax. Diatoms slides were observed using an Olympus light microscope with Nomarski differential interference contrast optics at a magnification of 1000x. Species identification was performed with the bibliographic support of The Flora of Diatoms,[32] Diatoms of Europe,[33] Iconographia Diatomologica,[34,35] and Diatomeen im Süsswasser-Benthos von Mitteleuropa.[36]

DNA/RNA Extraction. DNA and RNA were extracted with PowerBiofilm DNA and RNA isolation kits (MO BIO Laboratories Inc.) following the manufacturer instructions. RNA was purified from carried-over DNA molecules with TURBO DNase kit Ambion (Life Technologies) and cDNA obtained by reverse transcription using SuperScript III Reverse Transcriptase kit (Invitrogen). A total of 27 DNA and 27 cDNA (RNA) samples were obtained for this study.

For the extraction of cultured diatoms, pelleted cells were prepared by centrifuging 1 mL of fresh diatoms cultures at 8000*g* for 10 min. The extractions were then performed with DNeasy Plant Mini Kit (Qiagen) or PowerBiofilm DNA isolation (MO BIO).

Reference Database. We built a reference database of the V4 region composed of 460 unique diatom sequences. First, we downloaded from the GenBank database all sequences corresponding to the species and genera found in the morphological analyses of Geneva samples and also those commonly found in Switzerland.[9] The alignment was performed with the Seaview program.[37] Sequences were analyzed by Maximum Likelihood (ML) phylogenetic inference, and those showing incorrect identification were discarded. A total of 298 unique sequences from GenBank were kept.

To extend our reference database, we sequenced 10 diatom species obtained from culture collections: *Fragilaria pinnata* and

DOI: 10.1021/es506158m
Environ. Sci. Technol. 2015, 49, 7597−7605

Nitzschia ovalis from the CCAP (Culture Collection of Algae and Protozoa, SAMS Research Services Ltd., Scottish Marine Institute, Oban, U.K., http://www.ccap.ac.uk), *Achnanthidium minutissimum*, *Achnanthidium pyrenaicum*, *Achnanthidium straubianum*, *Amphora pediculus*, *Cocconeis placentula*, *Encyonema silesiacum*, *Nitzschia palea*, and *Sellaphora seminulum* from the TCC (Thonon Culture Collection, INRA-UMR Carrtel, Thonon-les-Bains, France, http://www6.inra.fr/carrtel-collection). We also added 152 Sanger sequences from other eDNA analyses of Geneva watercourses. The sequences were submitted to the Genbank database (KR089906-KR090057, KR150668-KR150677).

PCR Amplification, Cloning, and Sanger Sequencing. To complete the reference database and to test the specificity of PCR primers, the diatom cultures and environmental samples cited above were examined. The hypervariable region V4 of the 18S rRNA gene was amplified using primers modified after Zimmermann[19] DIV4for: 5′-GCGGTAATTCCAGCTCCA-ATAG-3′, DIV4rev3:5′-CTCTGACAATGGAATACGAATA-3′. PCR amplifications were performed in a total volume of 25 μL using Taq DNA Polymerase by Roche Applied Science. PCR regime included an initial denaturation at 94 °C for 2 min, then 35 cycles of denaturation at 94 °C for 45 s, annealing at 50 °C for 45 s, elongation at 72 °C for 1 min, and a final elongation at 72 °C for 10 min. PCR amplicons were purified with a High Pure PCR Product Purification kit (Roche Applied Science) and cloned using a TOPO TA Cloning kit for sequencing (Invitrogen). Sequence reactions were performed with BigDye Terminator (Applied Biosystems), and sequences were obtained by Sanger sequencing on ABI PRISM 3130XL Genetic Analyzer System (Applied Biosystems/Hitachi).

PCR Amplification for Next-Generation Sequencing. PCR were performed on DNA and RNA (cDNA) isolated from periphyton samples using unique combinations of forward and reverse tagged primers. Individual tags are composed of 8 nucleotides attached at each primer's 5′-extremity. A total of 20 different forward and reverse tagged primers were designed to enable multiplexing of all PCR products in a unique sequencing library. PCRs were performed as described above. Purified PCR products were quantified by fluorometric method using QuBit HS dsDNA kit (Invitrogen). Concentrations were then calculated and normalized for all samples. Approximately 50 ng of amplicons of each DNA and RNA sample from the SECOE 2013 (DIATOM 2013) and 2014 (DIATOM 2014) campaigns were pooled. An amount of 100 ng of pooled amplicons was used for the Illumina library preparation.

Illumina Library Preparation and Sequencing. Indexed paired-end libraries of pooled amplicons for consecutive cluster generation and DNA sequencing were constructed using an Illumina TruSeq Nano DNA Sample Preparation Kit—Low Throughput. Libraries were prepared following the manufacturer instructions. The fragment sizes of each library were verified by loading 3 μL of the final product in a 1.5% agarose gel with 1x SYBRSafe (Invitrogen) and quantified by a fluorometric method using a QuBit HS dsDNA kit (Invitrogen). An MiSeq Reagent Nano kit v2, with 500 cycles with nano (2 tiles) flow cells was used to run libraries on the MiSeq System. Two 250 cycles were used for an expected output of 500 Mb and an expected number of 1 million reads per library.

NGS Data Analysis. Operational Taxonomic Units (OTUs) were obtained and assigned following the method described in Pawlowski et al.[38] using the diatoms reference database described above. Raw FASTQ reads were quality-filtered by removing any sequence with a mean quality score of 30, and also removing all sequences with ambiguous bases or any mismatch in the tagged primer or contig region. These extremely stringent parameters ensure that we keep only high-quality reads. Then, paired-end reads were assembled by aligning them into a contiguous sequence with highest similarity. In case of mismatching bases, we kept in the final contig the closest base from the read 5′-extremity, based on the fact that the probability of miscalls increases toward the 3′-extremity. These sequences were then demultiplexed (assigned to their corresponding sample) depending on the tagged primers found at each end. Dereplication of the data set obtained after assembly was necessary in order to obtain unique sequences, called Independent Sequence Units (ISUs). An abundance threshold of 10 was used for the minimum number of replicates found for each ISU, and this abundance was recorded for further analyses. Subsequently, ISUs were assigned by performing a pairwise Needleman—Wunsch global alignment against our entire reference database. For the ISUs that were not assigned at the end of this procedure, we relied on a BLAST filtering procedure. We removed the ISUs that did not match any Bacillariophyceae sequences in the NCBI database with at least 99% coverage and 97% identity.

Phylogenetic Analyses. The taxonomic assignment of OTUs was checked by phylogenetic analyses. A tree was built with all the sequences from the database and the OTUs from the NGS analysis. The most abundant ISU was used as the representative sequence for each OTU. The ML phylogeny was constructed using RAxML v.7.4.2,[39] with GTR + G as model of evolution and 1000 replicates for the bootstrap analysis. The OTUs were assigned to the reference morphospecies if they formed a clade supported by bootstrap values >60 (following Zimmermann et al.[29] and references cited therein).

RESULTS

NGS Data Statistics. For DIATOM 2013, we obtained 1 176 424 reads from Illumina sequencing (SI Table S2). The filtering process rejected 169 841 reads with low mean quality, 61 508 reads with low base quality, 2205 reads with not enough matching bases in the contig region and 177 325 reads with errors or mismatches in the primers. Hence, a total of 765 545 reads remained after filtering and were available for further analysis. For DIATOM 2014, we obtained 1 055 387 reads. The filtering process rejected 296 799 reads with low mean quality, 17 095 reads with low base quality, 152 394 reads with not enough matching bases in the contig region, 247 694 reads with errors or mismatches in the primers and 23 222 with insufficient sequence lengths. Hence, a total of 318 183 good reads remained for further analysis.

Morphological Data and DI-CH Calculation. For each sampling site, about 400 valves were observed and identified with light microscopy at SECOE. Morphospecies were counted, and the relative abundance of each taxon was calculated for each site (SI Table S3). A total of 96 species was found by morphological identification. The number of taxa per site varied from 5 (AMB) to 37 (HEB). One species (*Amphora pediculus*) was found at every site and represented the most abundant taxon counted for all sites together. The values of DI-CH were calculated using the formula presented previously. The DI-CH values varied from 3.64 (NAM) to 7.98 (AMB). Highest DI-CH values were obtained for sites with larger numbers of diatoms with high autoecological values, such as *Nitzschia*

DOI: 10.1021/es506158m
Environ. Sci. Technol. 2015, 49, 7597−7605

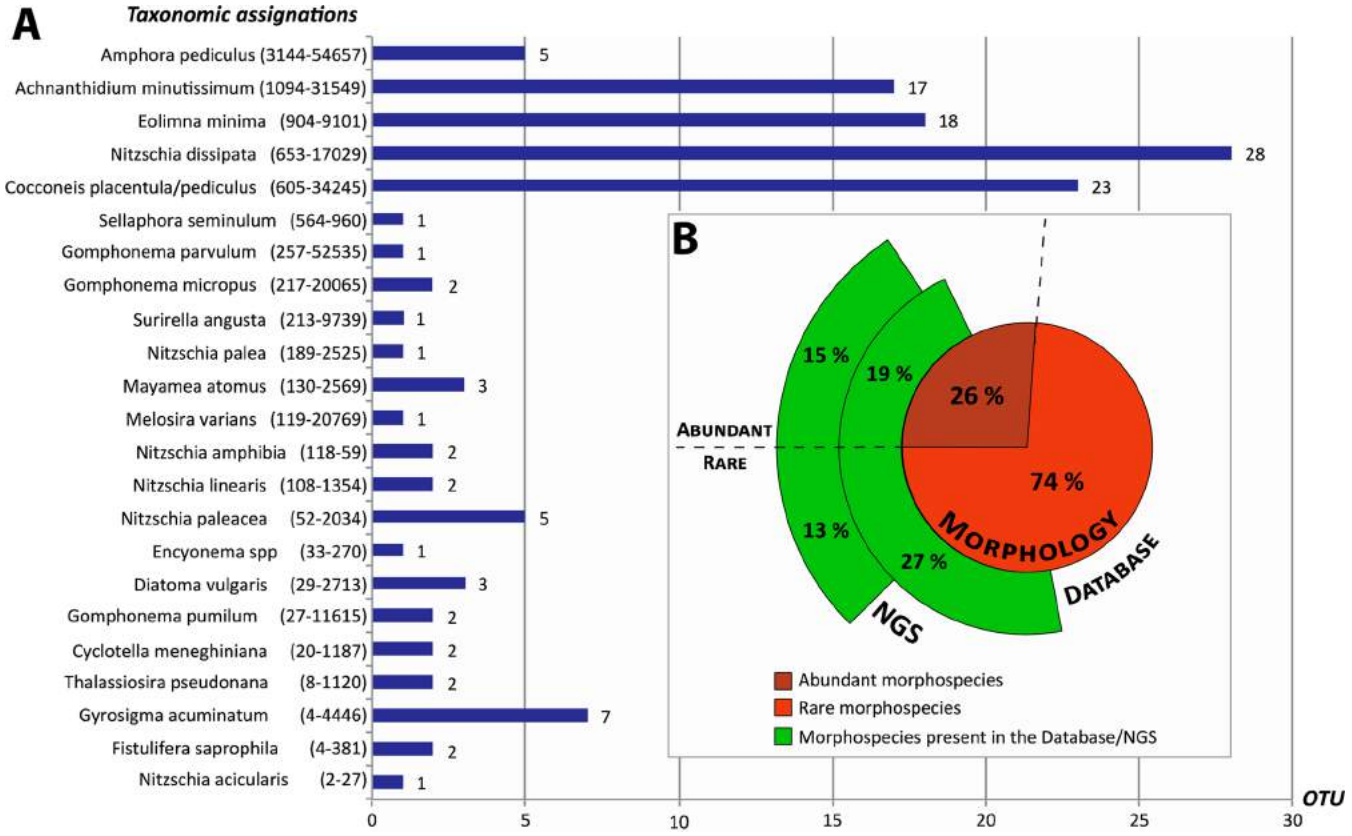


Figure 1. (A) Taxonomic assignments in common with morphospecies sorted by the number of counts in the morphologic analysis (in parentheses). The bar plot represents the number of OTU in each taxonomic assignation. (B) Pie chart of abundant (brown) and rare (orange) morphospecies found in morphologic analysis. Arcs in green represent the morphospecies present in the database (internal one) and in the NGS assignations (external one). Each arc is divided between abundant and rare species by a dashed line.

amphibia, *Sellaphora seminulum*, *Eolimna minima*, *Gomphonema micropus*, *Gomphonema parvulum*, *Eolimna subminuscula*, *Navicula veneta*, and *Nitzschia acicularis*.

Taxonomic Assignment of NGS data. Analysis of the NGS data grouped the reads into 242 OTU for the DIATOM 2013 and 103 for the DIATOM 2014 runs. In order to assign those OTUs to morphological taxa, an ML tree with all OTUs and our reference database was built. After phylogenetic analysis, we removed 128 OTUs for the DIATOM 2013 run and 60 OTUs for the DIATOM 2014 run because they could not be univocally assigned to any morphological clade. In total, 144 OTUs remained and were assigned to 30 taxa. Twenty-three of these taxa corresponded to the morphospecies found in microscopic analyses, while seven matched to species in the reference database that were not evidently found with the morphology-based approach.

Among the 23 assigned species (Figure 1A), 15 were confidently identified, i.e., they formed well-supported clades (BV > 60) including reference sequences assigned to a single morphospecies. *Encyonema* spp. was a special case since the only GenBank reference sequence of the clade was not identified beyond the genus level. Five species formed clades with reference sequences assigned to two different species of the same genus. These species were *Amphora pediculus*, *Achnanthidium minutissimum*, *Cocconeis placentula/pediculus*, *Mayamea atomus*, and *Fistulifera saprophila*.

Two assignments were particularly problematic. The OTUs assigned to *Cyclotella meneghiniana* formed a well-supported clade (BV 78) with 8 other *Cyclotella* species, half of which were marine species. We assigned these OTUs to *C. meneghiniana* because it was the only species present in the morphological list with an autoecological value. In the second case, the two OTUs assigned to the morphospecies *Thalassiosira pseudonana* formed a well-supported clade (BV 88) with 13 other *Thalassiosira* species and with the species *Stephanodiscus minutulus*. As both *S. minutulus* and *T.pseudonana* have the same autoecological value, we kept them together using the name of *T. pseudonana* as in morphological analyses.

In total, the number of morphospecies recognized in the NGS data amount to only 28% of all those identified in this study microscopically. However, it should be noted that the GenBank database only covers 46% of the morphospecies found in microscopic analyses (Figure 1B). The difference between these two percentages is accounted for by morphospecies (i.e., genus *Navicula*) that could not be identified unambiguously due to the lack of resolution of the V4 region. However, it is important to notice that most species not found in NGS were rare (below 100 counts in the morphologic analysis), as shown by Figure 1B. The list of the morphospecies with their count in the morphologic analysis and their presence in the database and in the NGS assignation are reported in SI Table S4.

Abundance of Assigned Species. As the calculation of diatom indices includes the relative abundance of species, we analyzed the variations in morphological counts and the number of reads inferred from DNA and RNA data for each assigned species. As can be seen in the SI (Table S5 and Figure

DOI: 10.1021/es506158m
Environ. Sci. Technol. 2015, 49, 7597−7605

S1), the relative abundance of species per site varies considerably depending on the type of data. In particular, the proportion of a species in DNA samples is often lower than in morphological counts and RNA samples. We checked whether this could be a consequence of the higher abundance of undetermined sequences in the DNA data, by reanalyzing the data with assigned OTUs only. However, the proportions between DNA, RNA, and morphological abundances remain the same in most of the cases.

The correlation between the number of reads and individuals for the most ubiquitous and abundant species is significant for both DNA and RNA of *A. pediculus* and DNA of *A.minutissima* (Figure 2). The relative abundance of some species (*A. pediculus*, *E. minima*) is higher in morphocounts than in NGS data. However, among the assigned morphospecies, there are very few sites where the species was found in microscopic preparations but not in the NGS data. This deviation is more obvious in less common taxa, with species such as *Nitzschia amphibia* being found almost exclusively in morphological analyses, while some species (e.g., *Gyrosigma acuminatum*) or genera (e.g., *Gomphonema*) are overrepresented in NGS data (SI Figure S1).

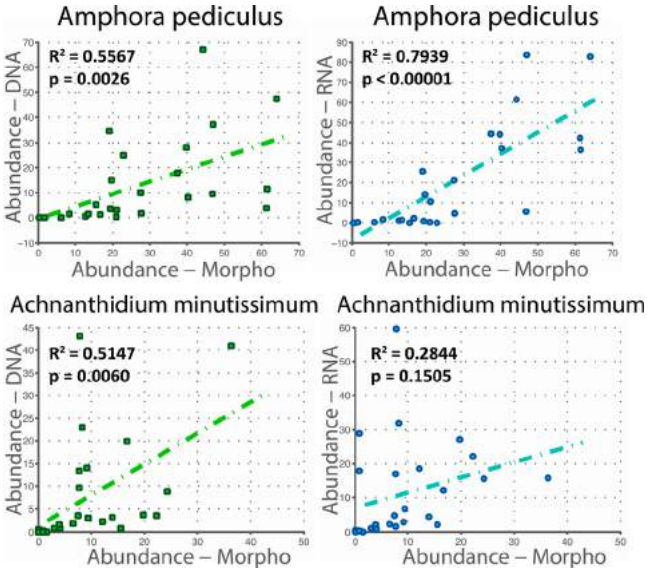


Figure 2. Relationships between the relative abundance of the two most abundant species *Amphora pediculus* (upper) and *Achnanthidium minutissimum* (lower). This information is displayed separately for DNA (left) and RNA (right) where each point shows the relationship between the relative abundance found in morphological (*x*-axis) or molecular (*y*-axis) counts. The dotted lines represent the results of model II regression with a least-squares fitting for the relative abundances of all samples. The R^2 and *p*-value are indicated for each regression axis.

Diatom Index. The NGS DI-CH index was calculated with the 23 taxa, for which the D and G values were available. When those values were different for a variety or subspecies of the same species, the values of the most abundant and frequent taxa were retained. All the DI-CH values for morphology, DNA, and RNA per site are presented in SI Table S6.

The variations in diatom indices inferred from morphological and molecular (DNA/RNA) data for 27 sites are illustrated in Figure 3. For the majority of sites (25 out of 27), the deviation between the morphological and at least one of the molecular indices (DNA or RNA) was less than 1 unit, and the biological quality status inferred from the two types of data was identical. For 17 sites (63%), the morphological index indicated the same level of water quality as at least one type of molecular data. Both DNA and RNA data were congruent with the morphological index in 7 out of 27 sites. When considered separately, the same level was indicated in 10 and 12 sites for DNA and RNA, respectively. The values of the morphological index exceeded those inferred from DNA and RNA in 16 sites (20 in the case of RNA). As we can see, the correlation between morphological and molecular indices is significant for DNA (Figure 4A) with $R^2 = 0.59$ and *p*-value = 0.0013 and becomes strongly supported in the case of RNA (Figure 4B) with $R^2 = 0.85$ and *p*-value <0.0001.

DISCUSSION

By exhibiting the strong similarity between the DI-CH values inferred from microscopic and NGS analyses of diatom communities, our proof-of-concept study clearly demonstrates the usefulness of NGS diatom data to evaluate water conditions. Our results confirm the previously reported similarity between values of the Specific Pollution Sensitivity biotic index obtained by microscopy and by NGS (pyrosequencing) analysis of SSU and rbcL barcodes.[28] Both studies

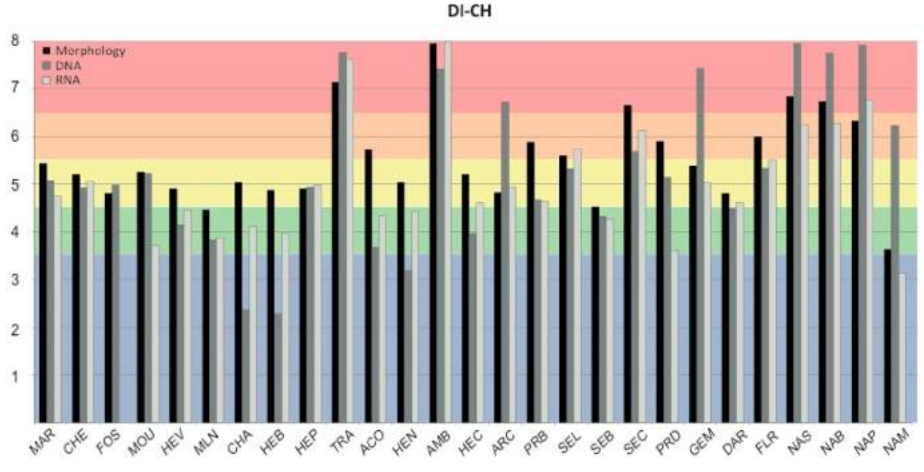


Figure 3. DI-CH values for morphologic analysis (black), DNA (dark gray), and RNA (light gray) per sites. Colors represent the threshold for water quality given by the DI-CH index.

DOI: 10.1021/es506158m
Environ. Sci. Technol. 2015, 49, 7597−7605

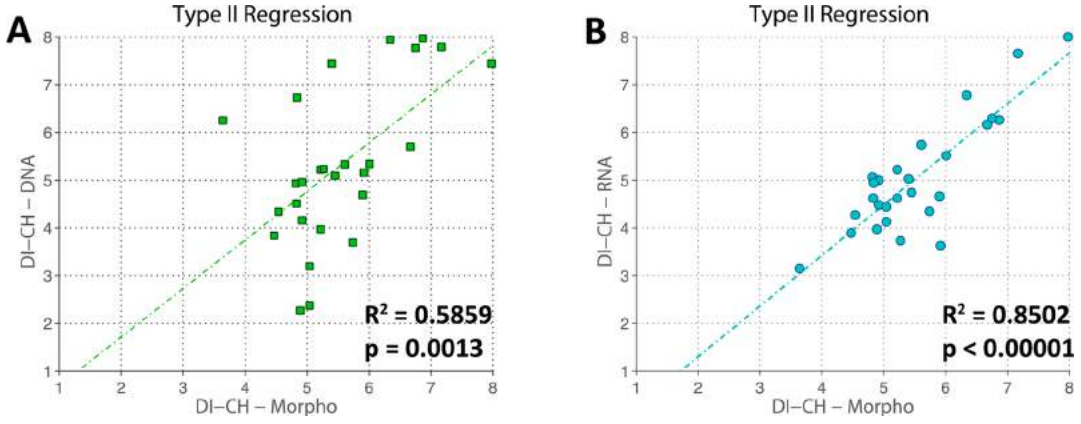


Figure 4. Relationships between the DI-CH inferred from morphological and DNA (A) or RNA (B) abundances per sites. Each point shows the relationship between the DI-CH found in morphological (*x*-axis) or molecular (*y*-axis) counts over all sites. The dotted lines represent the results of model II regression with a least-squares fitting for the relative abundances of all samples. The R^2 and *p*-value are indicated for each regression axis.

fully support the growing evidence that NGS environmental studies have the potential to become new tools for the assessment of aquatic ecosystems health, based on analysis of benthic macroinvertebrates,[24,25] diatoms,[27,29] and other protists.[38]

The congruence between diatom indices inferred either from morphological or NGS data is remarkable, given the poor database coverage and various technical biases. The correlation is especially strong for RNA (Figure 4B), likely because it provides a better depiction of the living diatom community composition. The DNA, however, can be preserved in water for a certain period of time and even carried over long distances.[40] Interestingly, the correlation between NGS and morphology in species relative abundances seems to have limited impact on the correlation between indices. This could be due to the fact that the index is calculated as the sum of a set of species with their respective weighting factors, which tends to reduce the effect of variations for individual species. In fact, a large number of species is assigned to the same set of weights, which means that the abundance of any given species can be replaced by the abundance of a set of several other species. Noticeably, the index correlates better in the sites with lower species richness, which might be related to the reduction of technical or biological biases in low complexity samples.

Although the results of our study are promising, there is still a wide potential to reduce the divergences between molecular and morphological results by addressing the current limitations of NGS data analysis. Some technical biases related to the DNA extraction, PCR conditions, primer specificity, library preparation, and sequence analysis have been extensively discussed in previous studies.[27,41,42] We discuss here the limitations that concern specifically the present study: (1) database incompleteness and inaccuracy, (2) inconsistencies between molecular and morphological taxonomy, and (3) biases in the quantitative analysis of NGS data.

Incompleteness and Inaccuracy of Databases. Gaps and misidentifications in reference databases are commonly believed to be the main hindrance to assigning taxonomy to environmental sequences. In fact, the diatom database is probably more exhaustive than that of any other groups of protists, especially those that cannot be cultivated.[43] The proportion of genetically characterized species in our study (46%) is slightly lower than in other studies targeting well-studied temperate regions (53−78%) but remains higher than those conducted in tropical regions (30−38%).[28] The development of comprehensive databases, like that of Zimmermann et al.,[30] which provided molecular (V4, rbcL) and morphological (LM, SEM) data for 70 cultured diatom strains, is an important step toward filling the gaps in diatom inventories. However, establishing cultures of diatom species for every eco-region could be extremely time-consuming and might not always be successful. An alternative approach could be based on single-cell PCR followed or preceded by LM or SEM study.[44] The success rate of these methods is still very low, but further developments in the field of single-cell genomics might rapidly improve their efficiency.

It should be noted that, although completing the database is important, it does not imply that the sequencing of all morphospecies is necessary. In our study, we assigned species according to very stringent criteria by removing all uncertain cases. Once the reference database is completed for common species such as *Achnanthes lanceolata*, and the identification of *Navicula* species is improved by using more rapidly evolving marker, the correlation between NGS and morphological indices might become even stronger. In fact, the vast majority of species currently missing from the database are rare, with less than 100 specimens per species counted in all samples. Their relative importance in the computation of diatom indices depends on the autoecological value associated with each species. However, it might be sufficient to correctly assign all common species and those rare species with high autoecological value to obtain a perfect match.

Molecular vs Morphological Taxonomy. Another potential source of conflict lies in the divergence between the morphological and molecular (phylogenetic) determination of diatom species. On the one hand, almost all morphospecies are represented by several genetically distinctive types. On the other hand, some morphospecies are subdivided into subspecies or morphological varieties, each with their own specific autoecological values. In the first case, the cryptic diversity may constitute a considerable advantage for biomonitoring, particularly if the cryptic species are associated with some specific ecological conditions. The second case is more problematic because the subspecific taxa are generally uncharacterized genetically.

In this study, we combined all subspecies and morphotypes belonging to the same species because it was impossible to distinguish them genetically. We also combined two species of

DOI: 10.1021/es506158m
Environ. Sci. Technol. 2015, 49, 7597−7605

Cocconeis, to avoid a possible misidentification of numerous phylotypes forming the clade of *C. placentula*, among which *C. pediculus* branches. In our approach, we followed the principle that the species can be grouped if they share the same ecologies and morphologies[45] and if they form a clade in phylogenetic analysis. Grouping at the generic level[46] may be useful, as in the case of *Encyonema*, but it is not necessary and may even be inappropriate in the case of polyphyletic genera.

Taxonomic resolution largely depends on the choice of the DNA barcode. Until now, only the chloroplastic rbcL and nuclear ribosomal 18S V4 region have been used in NGS diatom studies. Here, we chose the V4 region because its amplification from eDNA samples is easier and its size better fits the sequencing length of Illumina Miseq. It has been shown that the taxonomic resolution of V4 (and 18S in general) is lower than rbcL.[27] However, the interspecies variation of a given barcode may change between genera, and its efficiency will depend on the taxonomic composition of diatom community.[29] For example, in our study, the resolution of V4 was too low to unambiguously assign *Navicula* species, but it was sufficient to distinguish most of the species of *Nitzschia* and *Gomphonema*. Ideally, as both V4 and rbcL barcodes are complementary they should be used together in NGS analyses.

Relative abundance. Undoubtedly, the quantitative analysis of NGS data presents the greatest challenge in efforts to alleviate biases in the calculation of diatom indices. Indeed, numerous NGS environmental surveys exhibited discrepancies between the number of sequences assigned to a given species and the number of specimens of the same species in microscopic preparations[47,48] or even mock communities.[49] This lack of correlation between the abundance of reads and individuals could be explained either by technical biases introduced during DNA extraction, PCR amplification or sequencing,[50] or by biological factors such as the variations of rRNA gene copies,[51] which may depend on genome size,[52] number of nuclei,[53] or differences in cell size.[54]

Our study shows that molecular and morphological counts are well correlated in some species, but differ significantly in others (Figure 2). These variations seem taxon-specific and could be explained by variation in the numbers of rRNA gene copies in different diatom species. However, the ground-truth biological data necessary to test such a hypothesis are not available for diatoms. In fact, the correlation between molecular and morphological abundance data was previously observed in the NGS study of changes in foraminiferal[38] and metazoan (unpublished data) communities associated with the environmental impact of fish-farming, as well as in the study of the seasonal abundance in some species of ciliates and chrysophytes.[55] As the match between microscopic and molecular abundances concerns mainly the abundant species, this could explain why the impact of abundance variations on the final computation of the diatom index is relatively moderate.

Future Perspectives. The results presented in this pilot study will require validation by further NGS-based surveys of diatom diversity. In particular, substantial efforts will need to be done by diatom taxonomists and biologists to complete the DNA barcoding reference database and to determine the range of genetic and morphological variation in diatom species. Better knowledge of diatom genomes, especially the quantification of nuclear and chloroplast gene copies, will help in improving the estimation of species abundance from molecular data. Additional NGS studies of diatom communities in different ecological settings are also needed in order to optimize the molecular protocols and improve the accuracy of NGS data analysis, in particular to use the correction factors that would help overcoming the biases in relative abundance estimations.

All these efforts are worthwhile considering the tremendous benefits that the routine application of NGS approaches would bring to diatom-based monitoring. First, the use of DNA barcodes will allow standardization of species identification, which will help in overcoming the recurrent problems of misidentification and will facilitate the comparison of species inventories. Second, the molecular approach will provide more accurate real-time assessment of living communities, especially if RNA is analyzed rather than DNA. Third, the use of NGS technology coupled with the automation of molecular protocols will considerably reduce the time for sample processing, which will, in turn, allow an increase in the number of monitored sites. Finally, given the rapidly diminishing costs of NGS technologies, the application of these new tools will allow important savings.

■ ASSOCIATED CONTENT

Supporting Information

Table S1, Site locations, geographic references, and sampling dates performed along the Geneva basin (Switzerland) in collaboration with SECOE-DETA and used for the study. Table S2, Showing the filtering process on libraries DIATOM 2013 and DIATOM 2014. Table S3, Relative abundance and DI-CH values of morphological data per site location. Table S4, List and counting of species found during the morphological analysis of the two campaigns and their presence in the database (DN) and in the molecular assignation (NGS). Table S5, Relative abundance of morphologic, DNA and RNA data per sites. Table S6, DI-CH values for morphologic, DNA, and RNA data per sites. Figure S1, Relative abundance of 23 assigned taxa inferred for morphology (red), DNA (light green), and RNA (blue). The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/es506158m.

■ AUTHOR INFORMATION

Corresponding Author

*Phone: +41 22 3793069; fax: +41 22 379 33 40; e-mail: Jan.Pawlowski@unige.ch (J.P.).

Author Contributions

J.A.V. and L.A.P.G. contributed equally to this work. The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Funding

Swiss National Science Foundation, Swiss Federal Office for the Environment, G&L Claraz Donation

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Frédérique Rimet and Agnez Bouchez for the cultures of diatoms and helpful discussion, and Andrew Gooday for comments on the manuscript. We also thank François Pasquini from Water Ecology Service of the canton of Geneva for providing the infrastructure and equipment. Financial support was provided by the Swiss National Science Foundation (Grants 316030_150817 and 31003A-140766)

DOI: 10.1021/es506158m
Environ. Sci. Technol. 2015, 49, 7597−7605

and G & L Claraz Donation. This study is a part of the SwissBOL program supported by the Swiss Federal Office for the Environment.

■ ABBREVIATIONS

NGS next generation sequencing
eDNA environmental DNA
DI-CH Swiss Diatom Index

■ REFERENCES

(1) Stevenson, R. J.; Pan, Y.; van Dam, H. Assessing environmental conditions in rivers and streams with diatoms. In *The Diatoms: Applications of the Environmental and Earth Sciences*; Stoermer, E. F., Smol, J. P., Eds.; Cambridge University Press: Cambridge UK, 2010; 57 p.

(2) Rimet, F. Recent views on river pollution and diatoms. *Hydrobiologia* **2012**, *683*, 1−24.

(3) Belore, L. M.; Winter, J. G.; Duthie, H. C. Use of diatoms and macroinvertebrates as bioindicators of water quality in southern Ontario rivers. *Can. Water Resour. J.* **2002**, *27*, 457−484.

(4) Lobo, E. A.; Callegaro, V. L. M.; Hermany, G.; Bes, D.; Wetzel, C. A.; Oliveira, M. A. Use of epilithic diatoms as bioindicators from lotic systems in southern Brazil, with special emphasis in eutrophication. *Acta Limnol. Bras.* **2004**, *16*, 25−40.

(5) Poulickova, A.; Duchoslav, M.; Dokulil, M. Littoral diatom assemblages as bioindicators of lake trophic status: a case study from perialpine lakes in Austria. *Eur. J. Phycol.* **2004**, *39*, 143−152.

(6) Martin, G.; Fernandez, M. R. Diatoms and indicators of water quality and ecological status: sampling, analysis and some ecological remarks. In *Ecological Water Quality—Water Treatment and Reuse*; Voudouris, K., Ed.; InTech: North Canton, OH, 2012; pp 182−203.

(7) Kelly, M. G.; Bennett, C.; Coste, M.; Delgado, C.; Delmas, F.; et al. A comparison of national approaches to setting ecological status boundaries in phytobenthos assessment for the European Water Framework Directive: results of an intercalibration exercise. *Hydrobiologia* **2009**, *621*, 169−182.

(8) Directive 2000/60/EC of the European Parliament and of the Council of 23 October, 2000, establishing a framework for Community action in the field of water policy. *Official Journal L 327*, **22/12/2000** pp 0001−0073.

(9) Hürlimann, J.; Niederhauser, P. *Méthodes d'Analyse et d'Appréciation des Cours d'Eau. Diatomées Niveau R (region)*; Etat de l'environnement n° 0740. Office Fédéral de l'Environnement: Berne, 2007, 132p.

(10) *WPO—Water Protection Ordinance 814.201; 1998*. The Swiss Federal Council, based on Articles 9, 14 paragraph 7, 16, 19 paragraph 1, 27 paragraph 2, 36a paragraph 2, 46 paragraph 2, 47 paragraph 1, and 57 paragraph 4 of the Waters Protection Act of 24 January 1991 (WPA).

(11) Mann, D. G.; Droop, S. J. M.; Kristiansen, J. Biodiversity, biogeography and conservation of diatoms. Biogeography of freshwater algae. *Hydrobiologia.* **1996**, *336*, 19−32.

(12) Mann, D. G.; Sato, S.; Trobajo, R.; Vanormelingen, P.; Souffreau, C. DNA barcoding for species identification and discovery in diatoms. *Cryptogam.: Algol.* **2010**, *31*, 557−577.

(13) Evans, K. M.; Wortley, A. H.; Mann, D. G. An assessment of potential diatom "barcode" genes (cox1, rbcL, 18S and ITS rDNA) and their effectiveness in determining relationships in *Sellaphora* (Bacillariophyta). *Protist* **2007**, *158*, 349−364.

(14) Evans, K. M.; Mann, D. G. A proposed protocol for nomenclaturally effective DNA barcoding of microalgae. *Phycologia* **2009**, *48* (1), 70−74.

(15) Hamsher, S. E.; Evans, K. M.; Mann, D. G.; Poulickova, A.; Saunders, G. W. Barcoding diatoms: exploring alternatives to COI-5P. *Protist* **2011**, *162*, 405−422.

(16) Macgillivary, M. L.; Kaczmarska, I. Survey of the efficacy of a short fragment of the rbcL gene as a supplemental DNA barcode for diatoms. *J. Euk. Microbiol.* **2011**, *58*, 529−536.

(17) Moniz, M. B. J.; Kaczmarska, I. Barcoding micro- and meso-fauna. Barcoding diatoms: is there a good marker? *Mol. Ecol. Res.* **2009**, *9*, 65−74.

(18) Moniz, M. B. J.; Kaczmarska, I. Barcoding of diatoms: nuclear encoded ITS revisited. *Protist* **2010**, *161*, 7−34.

(19) Zimmermann, J.; Jahn, R.; Gemeinholzer, B. Barcoding diatoms: evaluation of the V4 subregion on the 18S rRNA gene, including new primers and protocols. *Org. Divers. Evol.* **2011**, *11*, 173−192.

(20) Luddington, I. A.; Kaczmarska, I.; Lovejoy, C. Distance and character-based evaluation of the V4 region of the 18S rRNA gene for the identification of diatoms (Bacillariophyceae). *PLoS One* **2012**, *7*, 1−11.

(21) Baird, D. J.; Hajibabaei, M. Biomonitoring 2.0: a new paradigm in ecosystem assessment made possible by next-generation DNA sequencing. *Mol. Ecol.* **2012**, *21*, 2039−2044.

(22) Bohmann, K.; Evans, A.; Gilbert, T. M. P.; Carvalho, G. R.; Creer, S.; Knapp, M.; Yu, D. W.; de Bruyn, M. Environmental DNA for wildlife biology and biodiversity monitoring. *Trends Ecol. Evol.* **2014**, *29* (6), 358−367.

(23) Taberlet, P.; Coissac, E.; Pompanon, F.; Brochmann, C.; Willerslev, E. Towards next-generation biodiversity assessment using DNA metabarcoding. *Mol. Ecol.* **2012**, *21*, 2045−2050.

(24) Hajibabaei, M.; Shokralla, S.; Zhou, X.; Singer, G. A. C.; Baird, D. J. Environmental barcoding: a next-generation sequencing approach for biomonitoring applications using river benthos. *PLoS One* **2011**, *6*, e17497.

(25) Hajibabaei, M.; Spall, J. F.; Shokralla, S.; van Konynenburg, S. Assessing biodiversity of a freshwater benthic macroinvertebrate community through non-destructive environmental barcoding of DNA from preservative ethanol. *BMC Ecol.* **2012**, *12*, 28.

(26) Carew, M. E.; Pettigrove, V. J.; Metzeling, L.; Hoffmann, A. A. Environmental monitoring using next generation sequencing: rapid identification of macroinvertebrate bioindicator species. *Front. Zool.* **2013**, *10*, 45.

(27) Kermarrec, L.; Franc, A.; Rimet, F.; Chaumeil, P.; Humbert, J. F.; Bouchez, A. Next-generation sequencing to inventory taxonomic diversity in eukaryotic communities: a test for freshwater diatoms. *Mol. Ecol. Res.* **2013**, *13*, 607−619.

(28) Kermarrec, L.; Franc, A.; Rimet, F.; Chaumeil, P.; Frigerio, J. M.; Jean-François Humbert, J. F.; Bouchez, A. A next-generation sequencing approach to river biomonitoring using benthic diatoms. *Freshwater Sci.* **2014**, *33*, 349−363.

(29) Zimmermann, J.; Glöckner, G.; Jahn, R.; Enke, N.; Gemeinholzer, B. Metabarcoding vs. morphological identification to assess diatom diversity in environmental studies. *Mol. Ecol. Res.* **2014**, *15*, 526−542.

(30) Zimmermann, J.; Abarca, N.; Enk, N.; Skibbe, O.; Kusber, W.-H.; Jahn, R. Taxonomic reference libraries for environmental barcoding: a best practice example from diatom research. *PLoS One* **2014**, *9*, e108793.

(31) Cordonier, A.; Gallina, N.; Nirel, P. M. Essay on the characterization of environmental factors structuring communities of epilithic diatoms in the major rivers of the canton of Geneva, Switzerland. *Vie Milieu/Life Environ.* **2010**, *60*, 223−231.

(32) Krammer, K.; Lange-Bertalot, H. Bacillariophyceae. In *Süsswasserflora von Mitteleuropa*; Gustav Fischer Verlag: Stuttgart, Germany. 1986−1991a,b. Teil 1−4.

(33) Lange-Bertalot, H., Ed. *Diatoms of the European Inland Waters and Comparable Habitats*; ARG Gantner Verlag KG: Ruggell, Lichtenstein, 2001−2003, Vols. *2*−*4*.

(34) Lange-Bertalot, H.; Metzeltin, D. *Indicators of Oligotrophy*; Koeltz Scientific Books: Königstein im Taunus, Germany, 1996, 390 p.

(35) Reichardt, E. *Zur Revision der Gattung Gomphonema*; Koeltz Scientific Books: Königstein im Taunus, Germany, 1999, 203 p.

(36) Hofmann, G.; Werum, M.; Lange-Bertalot, H. *Diatomeen im Süßwasser—Benthos von Mitteleuropa*; Koeltz Scientific Books: Königstein im Taunus, Germany, 2011.

DOI: 10.1021/es506158m
Environ. Sci. Technol. 2015, 49, 7597−7605

(37) Gouy, M.; Guindon, S.; Gascuel, O. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.* **2010**, *27*, 221−224.
(38) Pawlowski, J.; Esling, P.; Lejzerowicz, F.; Cedhagen, T.; Wilding, T. A. Environmental monitoring through protist next-generation sequencing metabarcoding: assessing the impact of fish farming on benthic foraminifera communities. *Mol. Ecol. Resour.* **2014**, *14*, 1129−40.
(39) Stamatakis, A.; Aberer, A. J.; Goll, C.; Smith, S. A.; Berger, S. A.; Izquierdo-Carrasco, F. RAxML-Light: a tool for computing terabyte phylogenies. *Bioinformatics.* **2012**, *28*, 2064−6.
(40) Deiner, K.; Altermatt, F. Transport distance of invertebrate environmental DNA in a natural river. *PLoS One* **2014**, *9*, e88786.
(41) Lee, C. K.; Herbold, C. W.; Polson, S. W.; Wommack, K. E.; Williamson, S. J.; McDonald, I. R.; Cary, S. C. Groundtruthing next-gen sequencing for microbial ecology−biases and errors in community structure estimates from PCR amplicon pyrosequencing. *PLoS One* **2012**, *7*, e44224.
(42) Esling, P.; Lejzerowicz, F.; Pawlowski, J. Accurate multiplexing and filtering for high-throughput amplicon-sequencing. *Nucleic Acids Res.* **2015**, *43*, 2513−2524.
(43) Pawlowski, J.; Audic, S.; Adl, S.; Bass, D.; Belbahri, L.; et al. CBOL protist working group: barcoding eukaryotic richness beyond the animal, plant, and fungal kingdoms. *PLOS Biol.* **2012**, *10*, 1−5.
(44) Lang, I.; Kaczmarska, I. A protocol for a single-cell PCR of diatoms from fixed samples: method validation using *Ditylium brightwellii* (T. West) Grunow. *Diatom Res.* **2013**, *26*, 43−49.
(45) DeNicola, D. M. A review of diatoms found in highly acidic environments. *Hydrobiologia* **2000**, *433*, 111−122.
(46) Rimet, R.; Bouchez, A. Biomonitoring river diatoms: Implications of taxonomic resolution. *Ecol. Indic.* **2012**, *15*, 92.
(47) Nolte, V.; Pandey, R. V.; Jost, S.; Medinger, R.; Ottenwälder, B.; Boenigk, J.; Schlötterer, C. Contrasting seasonal niche separation between rare and abundant taxa conceals the extent of protist diversity. *Mol. Ecol.* **2010**, *19*, 2908−2915.
(48) Stoeck, T.; Breiner, H.-W.; Filker, S.; Ostermaier, V.; Kammerlander, B.; Sonntag, B. A morphogenetic survey on ciliate plankton from a mountain lake pinpoints the necessity of lineage-specific barcode markers in microbial ecology. *Environ. Microbiol.* **2014**, *16*, 430−444.
(49) Amend, A.; Seifert, K. A.; Bruns, T. D. Quantifying microbial communities with 454 pyrosequencing: does read abundance count? *Mol. Ecol.* **2010**, *19*, 5555−5565.
(50) Pawlowski, J.; F. rowicz, F.; Esling, P. Next-generation environmental diversity surveys of Foraminifera: preparing the future. *Biol. Bull.* **2014**, *227*, 93−106.
(51) Weber, A.; Pawlowski, J. Can abundance of protists be inferred from sequence data? A case study of cultured Foraminifera. *PLoS One* **2013**, *8*, e56739.
(52) Prokopowich, C. D.; Gregory, T. R.; Crease, T. J. The correlation between rDNA copy number and genome size in eukaryotes. *Genome* **2003**, *46*, 48−50.
(53) Heyse, G.; Jönsson, F.; Chang, W.-J.; Lipps, H. J. RNA-dependent control of gene amplification. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 22134−22139.
(54) Godhe, A.; Asplund, M. E.; Höm, K.; Saravanan, V.; Tyagi, A.; Karunasagar, I. Quantification of diatom and dinoflagellate biomasses in coastal marine seawater samples by real-time PCR. *Appl. Environ. Microbiol.* **2008**, *74*, 7174−7182.
(55) Medinger, R.; Nolte, V.; Pandey, R. V.; Jost, S.; Ottenwälder, B.; Schlötterer, C.; Boenigk, J. Diversity in a hidden world: potential and limitation of next-generation sequencing for surveys of molecular diversity of eukaryotic microorganisms. *Mol. Ecol.* **2010**, *19*, 32−40.

DOI: 10.1021/es506158m
Environ. Sci. Technol. 2015, 49, 7597−7605

MOLECULAR ECOLOGY
RESOURCES
Molecular Ecology Resources (2017) doi: 10.1111/1755-0998.12668

Taxonomy-free molecular diatom index for high-throughput eDNA biomonitoring

LAURE APOTHÉLOZ-PERRET-GENTIL,* ARIELLE CORDONIER,† FRANÇOIS STRAUB,‡ JENNIFER ISELI,‡ PHILIPPE ESLING§ and JAN PAWLOWSKI*

*Department of Genetics and Evolution, University of Geneva, boulevard d'Yvoy 4, 1205, Geneva, Switzerland, †Water Ecology Service, Department of Territorial Management, Canton of Geneva, avenue de Sainte-Clotilde 23, 1211, Geneva, Switzerland, ‡PhycoEco, Rue des XXII-Cantons 39, 2300, La Chaux-de-Fonds, Switzerland, §IRCAM, UMR 9912, Université Pierre et Marie Curie, place Igor Stravinsky 1, 75004, Paris, France

Abstract

Current biodiversity assessment and biomonitoring are largely based on the morphological identification of selected bioindicator taxa. Recently, several attempts have been made to use eDNA metabarcoding as an alternative tool. However, until now, most applied metabarcoding studies have been based on the taxonomic assignment of sequences that provides reference to morphospecies ecology. Usually, only a small portion of metabarcoding data can be used due to a limited reference database and a lack of phylogenetic resolution. Here, we investigate the possibility to overcome these limitations using a taxonomy-free approach that allows the computing of a molecular index directly from eDNA data without any reference to morphotaxonomy. As a case study, we use the benthic diatoms index, commonly used for monitoring the biological quality of rivers and streams. We analysed 87 epilithic samples from Swiss rivers, the ecological status of which was established based on the microscopic identification of diatom species. We compared the diatom index derived from eDNA data obtained with or without taxonomic assignment. Our taxonomy-free approach yields promising results by providing a correct assessment for 77% of examined sites. The main advantage of this method is that almost 95% of OTUs could be used for index calculation, compared to 35% in the case of the taxonomic assignment approach. Its main limitations are under-sampling and the need to calibrate the index based on the microscopic assessment of diatoms communities. However, once calibrated, the taxonomy-free molecular index can be easily standardized and applied in routine biomonitoring, as a complementary tool allowing fast and cost-effective assessment of the biological quality of watercourses.

Keywords: bioindication, environmental DNA, metabarcoding, water quality

Received 28 July 2016; revision received 6 March 2017; accepted 7 March 2017

Introduction

Various biotic indices are widely used for the assessment of water quality. Traditionally, the indices are calculated based on the diversity of selected bioindicator taxa identified morphologically (Borja & Dauer 2008; Poikane *et al.* 2011). Recently, several attempts have been made to use eDNA data to infer the community structure of bio-indicator species (Baird & Hajibabaei 2012; Chariton *et al.* 2015). Several factors have been identified that may potentially impede the correct assignment of sequences to morphospecies and therefore the calculation of accurate indices. In particular, the incompleteness of the genetic database, the lack of resolution of phylogenetic markers and cryptic diversity (Yu *et al.* 2012; Carew *et al.* 2013; Eiler *et al.* 2013) have been highlighted as major issues. To overcome these limitations, we examine here whether it is possible to infer a molecular index directly from eDNA data without referring to the morphotaxonomy.

As a case study, we chose benthic diatoms, which are widely used as bioindicators of rivers and streams because of their high sensitivity to environmental changes and well-established taxon-specific ecological tolerances and preferences (Stevenson *et al.* 2010). In 2000, the European Union published a directive, the Water Framework Directive (Directive 2000/60/EC), that commits all member states to evaluate the status of their water bodies and to achieve a good status for them by a set deadline, recommending diatoms as one of the ideal bioindicators for river assessment. Different biotic indices are used across the different countries (Kelly

Correspondence: Pawlowski Jan, E-mail: jan.pawlowski@unige.ch

© 2017 John Wiley & Sons Ltd

et al. 2008). In Switzerland, two biological indices are used to comply with the concomitant ecological objectives specified by the Swiss decree on water protection (Swiss Federal Council 1998), the IB-CH using macrozoobenthos and the DI-CH, using diatoms. The Swiss Diatom Index (DI-CH) is based on chemical parameters indicating anthropogenic pollution and classifies the water quality into five different ecological classes on a scale from 1 to 8 (1–3.5: very good; 3.5–4.5: good; 4.5–5.5: average; 5.5–6.5: bad; 6.5–8: very bad). The calculation follows the weighted average equation of Zelinka & Marvan (1961) and is defined as

$$\mathrm{DI-CH} = \frac{\sum_{i=1}^{n} D_i G_i H_i}{\sum_{i=1}^{n} G_i H_i}$$

This equation involves an autecological value D and a weighting factor G, which are specific to each species. It also uses an additional parameter H, which corresponds to the relative frequency of a particular taxon in the sample.

Like other diatom indices (Kelly *et al.* 2001; Coste *et al.* 2009), the DI-CH requires a morphologic determination to the species level. This requirement is a major weakness of the currently used system. Indeed, diatoms are a highly diverse group of protists and the identification of their tiny frustules requires special sample preparation, high-quality microscopes and in-depth taxonomic expertise. Inter-calibration exercises among specialists are organized to validate the robustness of the indices. These time-consuming limiting factors contrast with the need for the fast routine assessment of water quality required by Water Framework Directive and the Swiss Federal Office for the Environment.

The development of high-throughput sequencing (HTS) technologies applied to diversity surveys of microbial eukaryotes communities provided a possibility to overcome some of these limitations (Pawlowski *et al.* 2016). Several attempts have been made to use HTS eDNA metabarcoding as a tool for identifying diatom species either in mock communities (Kermarrec *et al.* 2013, 2014) or in environmental samples (Kermarrec *et al.* 2014; Zimmermann *et al.* 2014, 2015; Visco *et al.* 2015). Some authors attempted to infer diatom indices from metabarcoding data (Kermarrec *et al.* 2014; Visco *et al.* 2015; Keck *et al.* 2016). However, the results of these studies were not entirely satisfactory due to uncertainties concerning the correct assignment of sequences to morphospecies and various biases involved in qualitative and quantitative analyses of molecular data.

Here, we propose a taxonomy-free approach to calculate the Swiss Diatom Index values directly from sequence data. To test this new approach, we analyse 87 epilithic samples from Swiss rivers, mostly located in the Geneva basin, using the hypervariable region V4 of 18S rDNA as the diatom DNA barcode and the Illumina Miseq platform for sequencing. As illustrated in Fig. 1, we calculate the DI-CH values inferred from molecular data with two methods. First, by phylogenetic assignment of OTUs to morphospecies (DI-MOLTAXASSIGN – pathway 2), as previously described in Visco *et al.* 2015. Second, by assigning OTUs directly to ecological classes (DI-MOLTAXFREE – pathway 3). Finally, we compare those values with the ones derived from traditional microscopic studies (DI-CH – pathway 1).

Material and methods

Sampling

In total, 87 samples were collected during the 2013–2015 period in the Geneva and Neuchâtel cantons in Switzerland (Table S1, Fig. S1, Supporting information). This number includes 27 samples already published in (Visco *et al.* 2015). All the samples were collected as part of the monitoring program for water quality performed by the Service of Water Ecology (SECOE) of the Department of Environment, Transport and Agriculture of the Geneva canton and the Service of Energy and Environment of the Neuchâtel canton. The biofilm containing epilithic diatoms was collected following the directives established by the Swiss Federal Office for the Environment (Hürlimann & Niederhauser 2007). Each sample was divided into two subsamples for morphological and molecular analyses. Morphological samples were preserved with a final concentration of at least 4% of formaldehyde, while molecular samples were kept cold (ca. 0 °C) during sampling. In the laboratory, about 1 mL of each sample suspension was centrifuged and pellets were stored at −80 °C until further investigations.

Morphological analysis

The preparation of diatoms slides for microscopic observation was performed as recommended by the Swiss Federal Office for the Environment (Hürlimann & Niederhauser 2007). About 500 valves per sample were counted and identified mainly with the bibliographic support of The Flora of Diatoms (Krammer & Lange-Bertalot 1986-1992), Diatoms of Europe (Lange-Bertalot 2001) and Iconographia Diatomologica (Lange-Bertalot & Metzeltin 1996; Reichardt 1999), and Diatomeen im Süsswasser-Benthos von Mitteleuropa (Hofmann *et al.* 2011). In the case of the samples from Neuchâtel, after the 500 valves had been counted, the preparations were scanned for 20 min to

© 2017 John Wiley & Sons Ltd

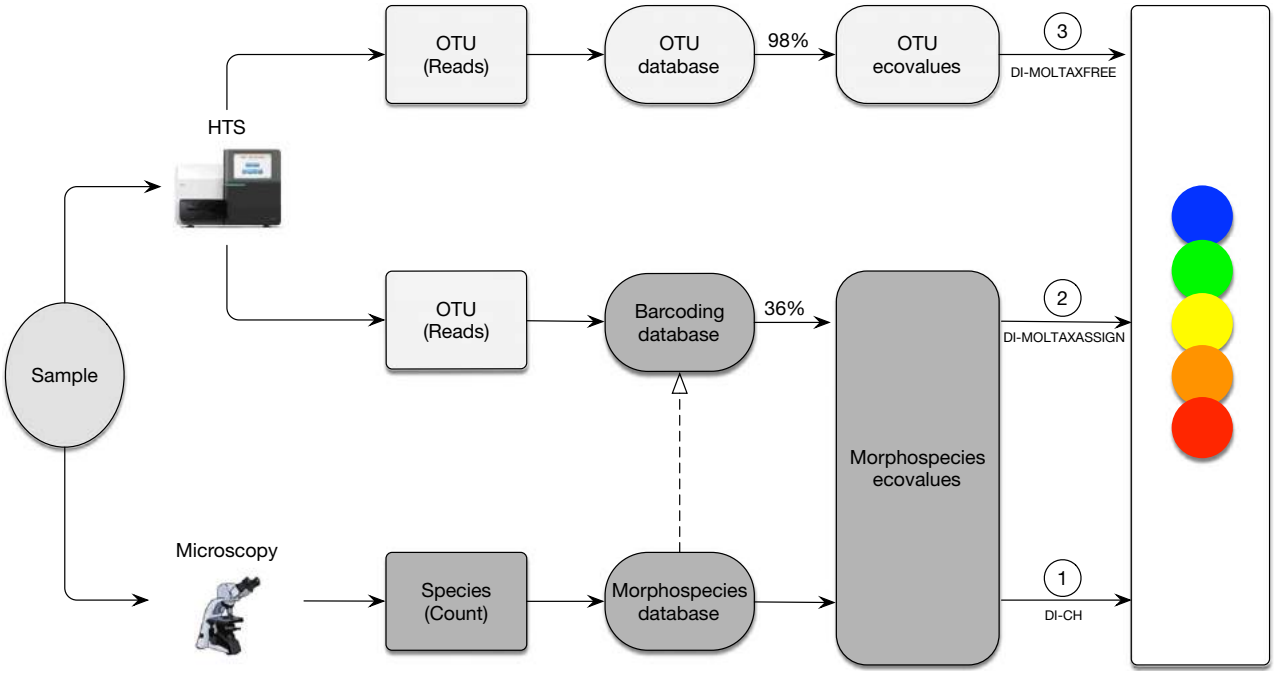


Fig. 1 Workflow illustrating the different methods used in this paper.

find rare species. Finally, the DI-CH values for each site were calculated following the equation described above.

Reference database

We chose the V4 region following the work of (Zimmermann *et al.* 2011) and our previous study (Visco *et al.* 2015). Although alternative diatom barcodes, such as rbcL, seem to offer better taxonomic resolution, we favour the V4 region because its amplification from eDNA samples is easier and its size better fits the sequencing length of the Illumina Miseq platform.

We built a reference database of the 18S V4 region of diatoms using online databases GenBank Release 212 and R-syst::diatom v5 (Rimet *et al.* 2016) and Sanger sequences from previous environmental studies in the Geneva basin (Visco *et al.* 2015). The region of interest was cut from downloaded sequences and aligned using the Seaview program (Gouy *et al.* 2010). The alignment was checked manually. Environmental sequences were screened using Uchime for chimeras (Edgar *et al.* 2011), which were then removed. The remaining sequences were analysed by Maximum Likelihood (ML) phylogenetic inference and those that did not branch in the clade corresponding to their morphological identification were discarded. After filtering, 1297 unique diatom sequences were kept, including 155 environmental sequences coming from the same geographic area as the study (Table S2, Supporting information).

Molecular analysis

DNA was extracted with the PowerBiofilm® DNA Isolation kit (MO BIO Laboratories Inc.) according to the manufacturer instructions. Three extraction replicates were performed for each sample. The hypervariable region V4 of the 18S rRNA gene of diatoms was then enriched by PCR amplification using specific diatom primers modified after (Zimmermann *et al.* 2011). Following previous studies, PCRs were performed as described in Visco *et al.* (2015), using unique combinations of forward and reverse primers tagged with individual tags composed of eight nucleotides attached at each primers 5′-extremities (Esling *et al.* 2015). A total of 20 different forward and reverse tagged primers were designed to enable multiplexing of all PCR products in a unique sequencing library. The sequences of tags and primers are provided in Table S3 (Supporting information).

Two PCR replicates were performed for each extraction and were then pooled for purification with High Pure PCR Cleanup Micro kit (Roche Diagnostics). In total, six PCR replicates were pooled for each sample. Purified PCR products were quantified with QuBit HS ds DNA kit (Invitrogen) and pooled in equimolar quantities. Two libraries were prepared (DIATOM03 for 2014 samples and DIATOM05 for 2015 samples, containing 24 and 36 samples, respectively) using Illumina TruSeq® DNA PCR-Free Library Preparation Kit following the manufacturer's instructions. The libraries were

© 2017 John Wiley & Sons Ltd

then quantified with qPCR using KAPA Library Quantification Kit and sequenced on a MiSeq instrument using paired-end sequencing for 500 cycles with NANO KIT v2.

HTS data analysis

Quality filtering and assembly were performed according to the method described in Visco *et al.* 2015;. The two runs from our previous study and the two from this study were combined, and this complete data set was de-replicated; that is, the identical sequences were grouped together to obtain unique sequences, called Independent Sequence Units (ISUs). An abundance threshold of 10 was used for the minimum number of reads required for each ISU (Bokulich *et al.* 2013). We removed the ISUs that did not match any diatom sequences in the NCBI database with at least 99% coverage and 97% identity. ISUs were then grouped at 99% using complete-linkage clustering method. Finally, we removed chimeric sequences found with manual inspection of Uchime (Edgar *et al.* 2011) candidates.

Phylogenetic analyses

Taxonomic assignment of the operational taxonomic units (OTUs) was checked by phylogenetic analyses. The most abundant ISUs were used as the representative sequence for each OTU and were aligned to the reference database. The Maximum Likelihood (ML) phylogeny was constructed using RAxML v.7.2.8 (Stamatakis 2014) with GTR + G as model of evolution and 1000 replicates for the bootstrap analysis. The OTUs were then assigned to a morphospecies if they formed a clade supported by bootstrap values >60, following our previous study (Visco *et al.* 2015) and that of (Zimmermann *et al.* 2015). After the OTUs were assigned, DI-CHMOLTAXASSIGN scores were calculated based on the molecular data, using the D and G values given by the assigned species and the relative frequency of reads for the H factor.

Calculation of ecological values

To calculate the autecological value D and the weighting factor G for each OTU, we rely on an approach similar to that used to create the DI-CH index itself (Hürlimann & Niederhauser 2007). For the calibration, the reference status for each site was given by the DI-CH values. For the calculation, only the OTUs with a relative frequency >1% in at least one sample were kept. To find the autecological value D, the samples were grouped into 15 classes from 1 to 8 with a step of 0.5 according to their ecological status. For each OTU, the class with the highest 80th percentile of relative frequencies was then kept as the D value. For the weighting factor G, the samples were grouped into eight ecological classes. For each OTU, the distribution of 80% of its total abundance across the eight classes was used to determine the weighting factor, using the following thresholds. 8: OTUs present in classes 1–3 and 7–8, corresponding to extreme ecological status. 4: OTUs present in 1 class only. 2: OTUs present in 2 classes. 1: OTUs present in 3 classes. 0.5: abundant OTUs present in a minimum of 4 classes or representing at least 3% in 3 classes. The workflow for this computation is summarized in Fig. S2 (Supporting information). This calculation was first done with the complete data set to compare the values given by the species assigned with the ones inferred from the DI-MOLTAXFREE approach.

Inference of the molecular index and cross-validation

The molecular index was inferred from HTS data based either on those OTUs that could be assigned to morphospecies (DI-MOLTAXASSIGN) or all OTUs having a relative abundance of more than 1% in at least one sample of the data set (DI-MOLTAXFREE). In the second case, the ecological values D and G were calculated as described above, while the H values were equal to the relative number of sequences (reads) for each OTU.

To evaluate the status of the taxonomy-free index (DI-MOLTAXFREE), two cross-validation tests were performed. In each case, the D and G values were recalculated without the tested samples. First, we used a leave-one-out cross-validation. To do so, one sample was removed from the data set for the calculation of the value D and the factor G. Then, these D and G values were used to calculate the DI-MOLTAXFREE index of the removed sample. This process was repeated for each sample. Second, we performed a 25/75 cross-validation in which the D and G values were calculated for 65 sites and the evaluation of the index on the 22 remaining sample. The sites were randomly chosen, and the validation was repeated for 1000 trials. The formula used to calculate the DI-MOLTAXFREE was the same as for the calculation of the morphological DI-CH presented in the introduction.

Results

HTS data

The samples were sequenced in four independent Illumina runs. A total number of 2 206 456 good reads distributed across the 87 samples remained after filtering. The details for each run are described in Table S4 (Supporting information). The reads from all runs were dereplicated, resulting in 3079 ISUs. The ISUs were

© 2017 John Wiley & Sons Ltd

clustered into 663 OTUs. After chimera removal, a final number of 440 OTUs was used for further analyses. The distribution of these OTUs and the number of reads per site are detailed in Table S5 (Supporting information). The number of OTUs per site varied from 1 (FOS) to 77 (VXB) with a median value of 27 (Table S6, Supporting information).

Morphological analysis

Morphospecies were counted, and the relative abundance of each taxon was calculated for each site (Table S7, Supporting information). A total of 269 morphospecies was identified across the 87 sites. The number of taxa per site varied from 5 (AMB) to 72 (PTH) with a median value of 24 (Table S6, Supporting information). The ecological status values ranged between 1.61 (VXD) and 7.98 (AMB). The different ecological classes (very good, good, average, bad and very bad) were represented by 15, 26, 25, 12 and 9 sites, respectively (Table S8, Supporting information). These DI-CH values were used as references for the molecular analysis.

Taxonomic assignment

We built a ML tree with our reference database and all OTUs (Fig. S3, Supporting information). After analysis, 152 OTUs (35%) were assigned to 43 morphospecies, of which 28 were found in the morphological analyses, while 15 matched to morphospecies not found microscopically in our samples. Figure S4 (Supporting information) shows the number of morphospecies recognized through morphological analysis, and in the genetic database and our HTS data set after phylogenetic assignment. Almost 70% of the morphospecies (185/269) found in the morphological counts were not represented in the database, leaving 84 morphospecies that were represented in the database. However, among these only 28 species were assigned in the molecular data set.

Ecological values comparison

In this section, we compare the D and G values provided by the morphological database with those inferred from molecular data (DI-MOLTAXFREE). To do so, we selected 78 of 152 taxonomically assigned OTUs that

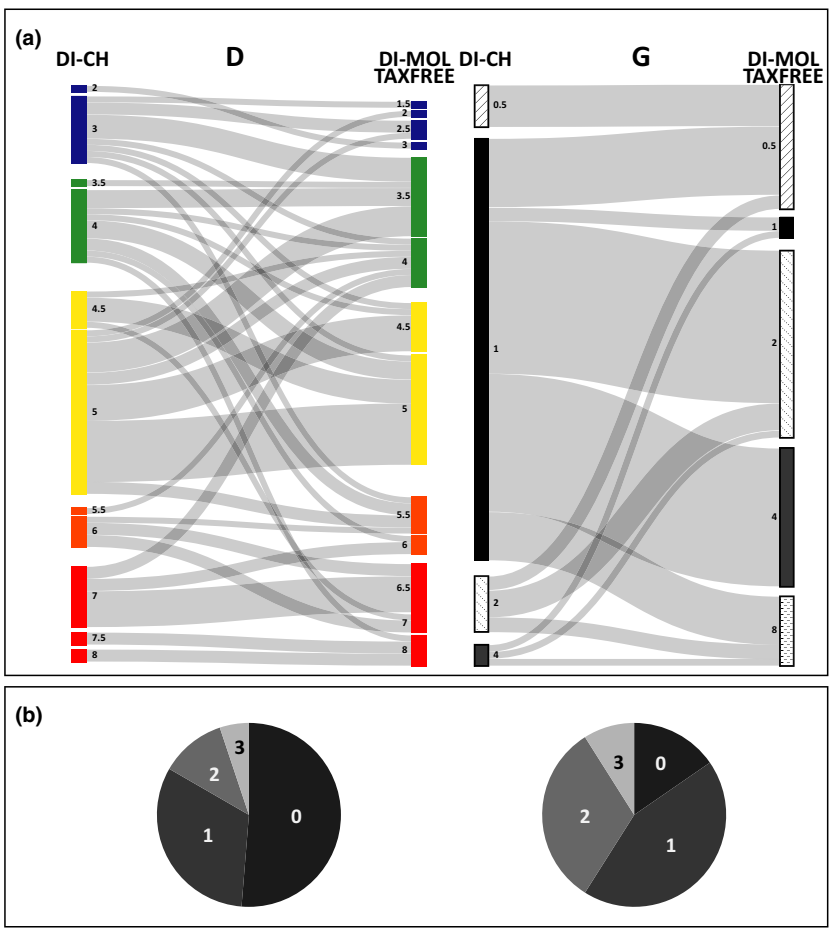


Fig. 2 Comparison of DG values for 78 assigned OTUs. The figure is separated into two parts: D values on the left and G values on the right. For each value, the sankey diagram (a) represents the relationship between the values inferred from morphology (DI_CH), and those inferred by the molecular index (DI-MOLTAXFREE). The links represent the assigned OTUs. Pie charts (b) represent the proportion of assigned OTUs as a function of the number of classes that change between their two values. No class changes are indicated in black, one class changes in dark grey, two classes change in medium grey and three classes change in light grey. For the D value, the class are separated as follows: 1–3.5: very good; 3.5–4.5: good; 4.5–5.5: average; 5.5–6.5: bad; 6.5–8: very bad and the scale of the G value is 0.5, 1, 2, 4 and 8.

© 2017 John Wiley & Sons Ltd

could be given the D and G values of the related morphospecies and represented more than 1% of the total number of sequences in at least one sample of the data set. The selected OTUs were assigned to 23 different morphospecies. Their D and G values obtained from the morphotaxonomic database were compared to the values obtained by the taxonomy-free approach (Fig. 2).

More than half of the 78 OTUs show a morphological and a molecular D value indicating the same ecological status and 15% of the OTUs show exactly the same G values. These numbers increase to 83% and 59% with a maximum of one change for the D value and G value, respectively. For both values, <10% show a drastic change of three categories difference. The D and G values are given for each assigned OTUs in Table S9 (Supporting information).

Relative abundance

Besides the ecological values D and G, we also compared the relative abundance of each species based on microscopic counts of specimens found at a particular site to the relative abundance of the corresponding OTU represented by the number of HTS reads (sequences). In Fig. S5 (Supporting information), we provide the results of this comparison for the 23 assigned morphospecies. In the majority of cases, we observed that the relative abundance of sequences is higher compared to the abundance of specimens (circles are located above the triangles). However, in few cases (e.g. *Sellaphora seminulum*), the opposite is observed. We calculated the correlation between the morphological and the molecular abundance for the four most abundant species. As shown in Fig. S6 (Supporting information), three species (*Cocconeis placentula*, *Eolimna minima*, *Planothidium lanceolatum*) showed a strong correlation (R^2 = 0.79, 0.76 and 0.90, respectively, with *P*-values < 0.0001), whereas *Achnanthidium minutissimum* did not (R^2 = 0.41).

Diatom index

The molecular scores inferred using the taxonomic assignment (DI-MOLTAXASSIGN) and the taxonomy-free method (DI-MOLTAXFREE) were compared to examine the coverage of the HTS data set by each of those two approaches. The range of the values calculated by the DI-MOLTAXASSIGN was 3.00–7.98 compared with 2.7–6.93 for the DI-MOLTAXFREE method. As illustrated in Fig. 3, the taxonomic assignment method utilized 36% of the reads, whereas the taxonomy-free approach utilized 98% of the data set. Similar proportions were found in the number of OTUs, with 38% and 85% of OTUs included in the taxonomic assignment and taxonomy-free approaches, respectively. For only one site (HEP), the number of OTUs used in the taxonomic assignment method was greater than in the taxonomy-free approach. This particular site shows a huge genetic diversity in *Cocconeis placentula* (17 different OTUs), although six of them were very rare and therefore were removed from the taxonomy-free analysis.

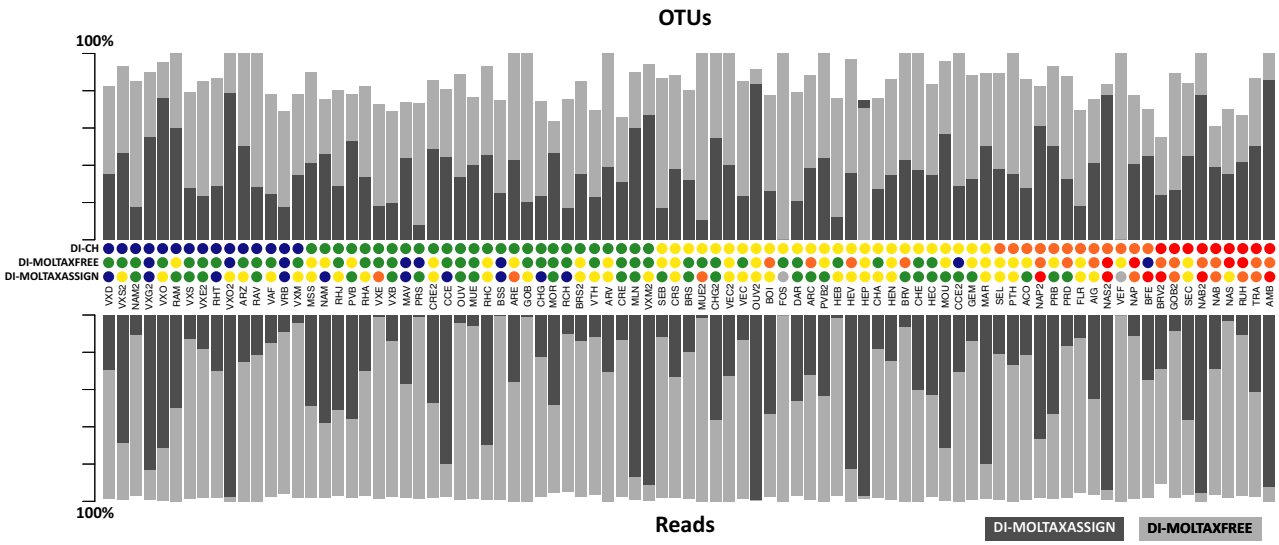


Fig. 3 Percentage of the HTS data set used by the taxonomic assignment (dark grey) and the molecular index (light grey) methods for each site. The OTUs are illustrated at the top and the reads at the bottom. In the middle, the coloured dots represent the ecological status given by the calculation of DI-CH values with Morphology (DI-CH), Molecular index (DI-MOLTAXFREE) or Taxonomic assignment (DI_MOLASSIGN). For the molecular index, the results of the leave-one-out cross-validation are used. The very good, good, average, bad and very bad statuses are represented with blue, green, yellow, orange and red colour, respectively.

© 2017 John Wiley & Sons Ltd

The central part of Fig. 3 indicates the ecological status inferred by each approach. The two molecular methods (taxonomic assignment and taxonomy-free) give the same ecological status for 45% (38/85) of the samples; 14 of them are congruent with the morphological evaluation. For 38% (33/87) of the samples, the DI-MOLTAXFREE gave the same class as the DI-CH compared with 30% (26/85) for the DI-MOLTAXASSIGN. For two sites (FOS and VEF), no sequences could be assigned and, therefore, no taxonomic assignment evaluation was possible.

The taxonomic assignment and taxonomy-free molecular indices are compared further in Fig. 4, which shows the correlations of each index with the values of the morphological index (DI-CH) and indicates the difference compared to the values of DI-CH. The correlation between DI-MOLTAXASSIGN and the DI-CH (R^2 = 0.57 and P-value < 0.0001) is lower than the correlation between DI-MOLTAXFREE and the DI-CH (R^2 = 0.67 and P-value < 0.0001). The values of the indexes differ by <1 in 77% of the samples for the DI-MOLTAXFREE, compared to 52% for the DI-MOLTAXASSIGN. The proportion of sites correctly assessed with the DI-MOLTAXASSIGN increases to 88% for the most sampled sites belonging to the good and average classes, which are the best represented in our data set. The under-sampled classes show less good results, with 75%, 67% and 46% of correctly assessed sites for the bad, very bad and very good classes, respectively (Fig. S7, Supporting information).

In the case of DI-MOLTAXFREE, the leave-one-out cross-validation test was used to better illustrate the comparison with DI-MOLTAXASSIGN (Fig. 4). However, similar results were obtained using the 25/75 cross-validation tests (shown as stars in Fig. 4 and illustrated in Figs S8 and S9, Supporting information). The seven most problematic sites remain the same in the two cross-validation tests. In those cases, the difference compared to the DI-CH is >1.5 for the leave-one-out analysis and for the 25/75 cross-validation, <6% of the trials show a difference below 1. Four of these seven sites belong to the very good quality class.

Discussion

Overcoming the taxonomic assignment issue

The main objective of this study was to test whether the step of taxonomic assignment is necessary to calculate a molecular diatom index with eDNA data. Previous studies highlighted various biases introduced by this step but still kept it as an integral part of their analyses (Kermarrec *et al.* 2014; Visco *et al.* 2015; Zimmermann *et al.* 2015). The present study shows that the molecular index computed with (DI-MOLTAXASSIGN) or without (DI-

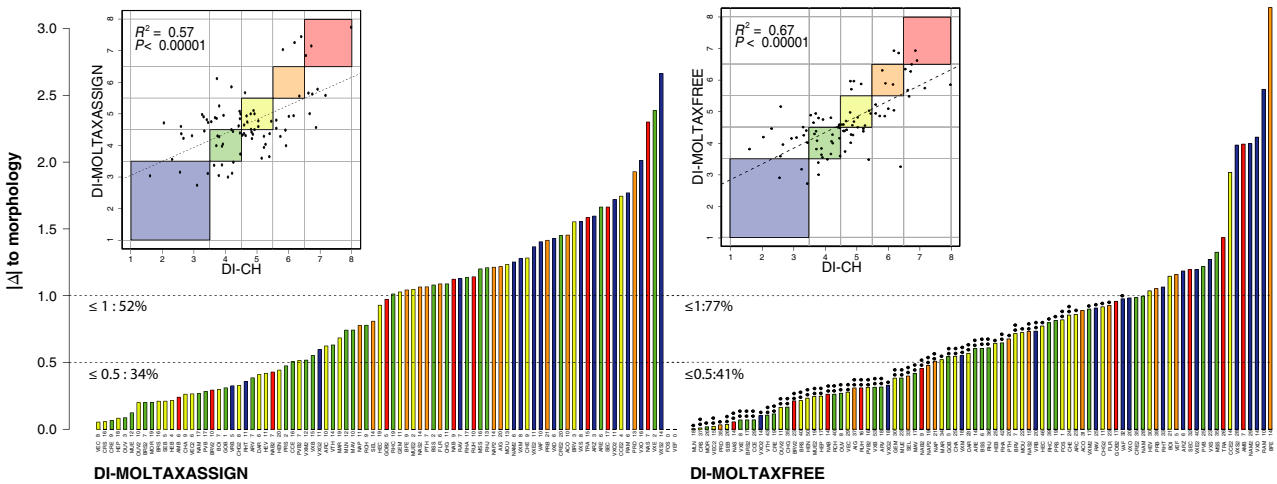


Fig. 4 Comparison between the DI-CH values given by morphology and molecular methods (taxonomic assignment on the left and leave-one-out cross-validation on the right). For each method, two types of graphics are represented. The scatter plots show the relationships between the DI-CH inferred from morphological (x-axis) and the molecular methods (y-axis). Coloured boxes represent the ecological status given by the DI-CH (blue: very good, green: good, yellow: average, orange: poor, red: bad). The regression line for all samples is represented by dashed line, and the R^2 and P-value are indicated for each graph. The bar plots show for each site the absolute difference between the DI-CH values given by the morphology and the molecular methods. Sites are coloured in function of their DI-CH value (blue: very good, green: good, yellow: average, orange: poor, red: bad). Above each site name, the number of OTUs used to calculate the index is indicated. Dashed lines are drawn at the 0.5 and 1 difference thresholds. Percentages of sites below these thresholds are indicated in the graphs. For each site, the black dots show the results of the 25/75 cross-validation. One dot is used if at least 70% of the replicates gave an absolute difference with the morphological DI-CH below 1 and two dots are used if this percentage is above 90%.

© 2017 John Wiley & Sons Ltd

MOLTAXFREE) taxonomic assignment is not significantly different. Moreover, we observe a higher correlation between morphological and molecular indices in the case of the taxonomy-free approach (Fig. 4), suggesting that taxonomic assignment may not be essential for eDNA-based diatom monitoring.

Our results suggest that the main benefit of taxonomy-free approach lies in its much higher data coverage compared to the use of taxonomic assignment. The latter step considerably reduces the amount of available data due to the incompleteness of genetic reference databases, which comprise only 31% of the morphospecies identified in this study. This small number is reduced further to 10%, as 56 morphospecies present in genetic database (many belonging to the genus *Navicula*) could not be correctly assigned because of the lack of resolution of the 18S V4 marker. The selection of another marker (e.g. *rbc*L proposed by Kermarrec *et al.* 2013 and MacGillivary & Kaczmarska 2011) could probably improve the phylogenetic assignment for some species. However, it is uncertain whether the global data coverage would be much better.

Even if all morphospecies were sequenced with a more highly resolving marker, the taxonomic assignment will still be compromised by the issue of cryptic genetic diversity. It is well known that, in common with many other protists, the majority of diatom morphospecies are represented by large numbers of OTUs that are not always monophyletic (Amato *et al.* 2007; Beszteri *et al.* 2007; Rimet *et al.* 2014; Rovira *et al.* 2015; Van den Wyngaert *et al.* 2015). For example, *Cocconeis placentula* is represented in our data by 17 OTUs. Although this species complex has been split morphologically into several subspecies, their correspondence to numerous OTUs branching within the *C. placentula* clade is not well established. As a result, it is not possible to use different ecological values assigned to these subspecies and, conversely, to take advantage of ecological values assigned to *C. placentula* OTUs by the taxonomy-free approach. Regarding the practical application of the diatom index, the main problem with the taxonomic assignment approach is not so much the lack of correspondence between OTUs and morphospecies, but the difficulty of avoiding the errors introduced by the direct translation of ecological values associated with morphospecies to corresponding OTUs.

Accuracy of ecological values

By overcoming the step of taxonomic assignment, our method provides an independent assessment of ecological values. These values have been estimated directly from the HTS data, using morphological analyses as a reference to establish the ecological status of each site. As such estimations have never been attempted before, we examine the difference between these newly calculated values and those given by morphological observations. Although this comparison could only be performed on a few assigned OTUs and a limited number of sites, the results shown in Fig. 2 and Fig. S6 (Supporting information) are promising.

In the case of the autecological D values, the same ecological status was obtained for most of the OTUs. On the contrary, the variations of the weighting factor G are wider, with most of the OTUs having G values more or less equally distributed between 0.5 and 8, while most morphospecies are characterized by a G value of 1 (Fig. 2). As the G value reflects the occurrence of species/OTU across the sites, it is possible that these wider variations are related to the presence of extracellular DNA that can be dispersed over large distances (Deiner & Altermatt 2014). Alternatively, it is possible that the G values are affected by low amplification efficiency, which artificially reduces the range of occurrence, making the ecological tolerance of a given OTU appear narrower than in morphological surveys.

The accuracy of the DI-MOLTAXFREE also depends on the stability of D and G values during cross-validation. As illustrated in Fig. S10 (Supporting information), the values of the weighting factor G are relatively stable, with 83% of 228 analysed OTUs changing less than one category. In the case of D values, the variations are greater, although they rarely exceed two points. These large variations can be an effect of under-sampling, limiting the number of sites where an OTU occurs. This probably applies in the case of OTU 427, which is responsible for the highest difference between the DI-MOLTAXFREE and the DI-CH index found at the site BFE. Another possibility is that morphological misidentification leads to an erroneous assessment of some sites where an OTU is present. Such misidentifications can occur when the samples are processed routinely without a detailed scanning electron microscope examination of each specimen. To avoid such errors, it is necessary to stabilize the D and G values by increasing the number of sites and adapting the D and G values to the specificities of molecular data.

The issue of relative abundance

The third factor that influences the molecular index is relative abundance. This is also examined here. It is widely accepted that different technical and biological biases impact the relative abundance of specimens and sequences, making impossible the use of quantitative data in HTS surveys (Elbrecht & Leese 2015). However, this was not confirmed by the present study, at least as far as the most abundant species are concerned (Fig. S6, Supporting information). The same tendency was observed in other protists, such as foraminifera, where

© 2017 John Wiley & Sons Ltd

the same species dominated morphological and molecular assemblages (Pawlowski *et al.* 2014). We could speculate that this relatively good match between the numbers of specimens and sequences of abundant species is reinforced by the exponential character of PCR amplification. As shown in the case of *C. placentula* and *E. minima* (Fig. S6, Supporting information), when a species is very abundant in microscopic counts, it is often even more abundant in HTS reads. However, this is not always true. For example, at some sites, the relative abundance of specimens of *Sellaphora seminulum* exceeds the abundance of reads (Fig. S7, Supporting information), suggesting that the PCR amplification may not be very efficient in this species.

In general, the importance of quantitative biases seems to be reduced in the case of small, single-cell organisms such as diatoms or foraminifera. However, the biomass of protistan cells can also vary considerably and the variability of rRNA copy numbers has been demonstrated in some diatoms (Alverson & Kolnick 2005; Godhe *et al.* 2008) and other protists (Gong *et al.* 2013; Weber & Pawlowski 2013). The taxonomy-free approach avoids this problem, because it does not involve the direct comparison of the relative abundance of specimens and sequences. Assuming that the PCR and other technical biases are the same across the samples for a given OTU (as long as the experimental conditions remain unchanged), the impact of these biases on the accuracy of taxonomy-free molecular index will be less important and easier to control than in the case of taxonomic assignment approach. Nevertheless, the formulae on which current indices are based are not adapted specifically for quantitative HTS data; a special effort will be required to address this issue in future studies.

Limitations of taxonomy-free approach

Although, as mentioned above, the taxonomy-free index has many advantages, it also has some important limitations that have to be overcome before the index can be used routinely. In view of our results, the most important factor causing incongruence between molecular and morphological indices is the lack of comprehensive sampling. As illustrated in Fig. 4, the DI-MOLTAXFREE approach considerably reduced the number of incorrectly assigned sites compared to the DI_MOLTAXASSIGN method. Yet, there are still sites that differ significantly from their status according to the morphological DI-CH method, and remarkably, most of them belonging to the under-sampled classes of very bad, bad and very good water quality.

The effect of under-sampling is particularly dramatic in the case of very good (blue) sites, half of which lie outside the 1-point limit (Fig. S5, Supporting information). This can be explained by the fact that these very good water quality sites are not characterized by specific indicator species but rather by different species-rich communities (Whitton *et al.* 1991; Hürlimann & Niederhauser 2007), which might be difficult to reconstruct without an extensive sampling. Conversely, the lack of congruence observed in the case of the bad and very bad quality sites can be explained the fact that these sites are usually characterized by high abundances of a few indicator species (Hill *et al.* 2001; Stevenso *et al.* 2010). When these sites appear rarely in the data set because of under-sampling, the absence of indicator species/OTUs in cross-validation studies may lead to the totally wrong assignment of a given site, as possibly happened in the case of sites AMB and BFE in our analyses (Fig. 4).

These few examples highlight the importance of sampling effort to ensure the accuracy of ecological values associated with OTUs in the taxonomy-free approach. However, even the most extensive eDNA sampling will not be able to alleviate all limitations of using OTUs rather than morphospecies to evaluate the quality of the environment. In particular, the metabarcoding data are unable to provide the kind of ecological information that is available through microscopic observations. For example, the list of OTUs and their relative frequencies says nothing about the physiological state of species, which can be measured by the proportion of teratological morphotypes in microscopic analyses (reviewed in Falasco *et al.* 2009). In general, the extensive knowledge of the taxonomy, biology and ecology of diatoms that can be derived from microscopic observations cannot be easily applied to the interpretation of molecular data. Therefore, the taxonomy-free index should be considered as a complementary tool rather than as a replacement for morphology-based studies.

Future challenges and perspectives

Our study raises several questions concerning the applicability of taxonomy-free approach in routine biomonitoring. Some of these questions, concerning the geographic range of OTUs and their ecological preferences, can hardly be answered without extensive sampling. Therefore, to further test the taxonomy-free index, the most important challenge is to obtain data from a much broader geographic area and from more diverse habitats. As shown by our results, the assessment of water quality is relatively good in the case of sites of average and good ecological status that dominate in our sampling. On the contrary, the diatom communities of the very good and very bad quality sites are not yet sufficiently represented in our data sets and, therefore, the inferred ecological values are not accurate enough. This highlights the importance of having not only numerous

© 2017 John Wiley & Sons Ltd

sites but also sufficiently varied sampling habitats to cover the widest diversity possible.

Another important challenge is the calibration of the taxonomy-free index. In the present study, we relied on a well-established diatom index that is routinely used to characterize water quality in Swiss rivers and streams. The Swiss index and other diatom indices currently available are based on decades of microscopic data collection that has provided comprehensive information about diatom species ecology and distribution. These morphological data are essential to calibrate the taxonomy-free index and ensure its accuracy and robustness. However, where morpho-taxonomic data are not available due to a lack of taxonomic expertise, other types of data, such as chemical parameters or macro-invertebrate surveys, could serve as alternative calibration options. The most readily available data are chemical parameters. Yet, to be useful for diatom index calibration, the chemical analyses have to be conducted over longer periods of time. Depending on the diversity and geographic ranges of diatom OTUs, calibration of the taxonomy-free index would be necessary for different habitats and geographic localities. However, once the index is properly calibrated, the ecological values for each OTU will be more stable and the values of diatom index will be more reliable.

To conclude, our study demonstrates the great potential of the taxonomy-free molecular index for environmental biomonitoring. Although our work focuses on diatoms and the specific case of the Swiss diatom index, the taxonomy-free approach could easily be applied to other groups of single-cell bioindicators, such as ciliates (Lee *et al.* 2004; Chen *et al.* 2008; Jiang *et al.* 2011), and foraminifera (Schönfeld *et al.* 2012; Vidovic *et al.* 2014; Alve *et al.* 2016). New molecular indices could also be tested for microbial and meiofaunal taxa that are not currently used as bioindicators. The implementation of these new indices would help to extend the range of monitored sites and increase the frequency of monitoring. Once established, molecular indices could provide a fast, easily standardized and highly sensitive tool that complements the current morphology-based methods available for the water quality assessment.

Acknowledgements

We thank Francois Pasquini from Water Ecology Service of the Canton of Geneva and Isabelle Butty from the division of water and soil of the Canton of Neuchâtel for their permission to use the samples and the morphological data. We also than Andrew Gooday from National Oceanographi Centre, Southampton for very careful editing of the manuscript and discussion. Financial support was provided by the Swiss National Science Foundation (grants 316030_150817 and 31003A-140766) and G & L Claraz Donation. This study is a part of the SwissBOL program supported by the Swiss Federal Office for the Environment.

References

Alve E, Korsun S, Schönfeld J *et al.* (2016) Foram-AMBI: a sensitivity index based on benthic foraminiferal faunas from North-East Atlantic and Arctic fjords, continental shelves and slopes. *Marine Micropaleontology*, **122**, 1–12.

Alverson AJ, Kolnick L (2005) Intragenomic nucleotide polymorphism among small subunit (18s) Rdna paralogs in the diatom genus skeletonema (bacillariophyta)1. *Journal of Phycology*, **41**, 1248–1257.

Amato A, Kooistra WHCF, Ghiron JHL *et al.* (2007) Reproductive isolation among sympatric cryptic species in marine diatoms. *Protist*, **158**, 193–207.

Baird DJ, Hajibabaei M (2012) Biomonitoring 2.0: a new paradigm in ecosystem assessment made possible by next-generation DNA sequencing. *Molecular Ecology*, **21**, 2039–2044.

Beszteri B, John U, Medlin LK (2007) An assessment of cryptic genetic diversity within the *Cyclotella meneghiniana* species complex (Bacillariophyta) based on nuclear and plastid genes, and amplified fragment length polymorphisms. *European Journal of Phycology*, **42**, 47–60.

Bokulich NA, Subramanian S, Faith J *et al.* (2013) Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nature methods*, **10**, 57–59.

Borja A, Dauer DM (2008) Assessing the environmental quality status in estuarine and coastal systems: Comparing methodologies and indices. *Ecological Indicators*, **8**, 331–337.

Carew ME, Pettigrove VJ, Metzeling L, Hoffmann AA (2013) Environmental monitoring using next generation sequencing: rapid identification of macroinvertebrate bioindicator species. *Frontiers in Zoology*, **10**, 45.

Chariton AA, Stephenson S, Morgan MJ *et al.* (2015) Metabarcoding of benthic eukaryote communities predicts the ecological condition of estuaries. *Environmental Pollution (Barking, Essex: 1987)*, **203**, 165–174.

Chen Q-H, Xu R-L, Tam NFY, Cheung SG, Shin PKS (2008) Use of ciliates (Protozoa: Ciliophora) as bioindicator to assess sediment quality of two constructed mangrove sewage treatment belts in Southern China. *Marine Pollution Bulletin*, **57**, 689–694.

Coste M, Boutry S, Tison-Rosebery J, Delmas F (2009) Improvements of the Biological Diatom Index (BDI): description and efficiency of the new version (BDI-2006). *Ecological Indicators*, **9**, 621–650.

Deiner K, Altermatt F (2014) Transport distance of invertebrate environmental DNA in a natural river. *PLoS ONE*, **9**, e88786.

Directive 2000/60/EC of the European Parliament and of the Council of 23 October (2000) Establishing a framework for Community action in the field of water policy. *Official Journal L*, **327**, 1–73.

Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R (2011) UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*, **27**, 2194–2200.

Eiler A, Drakare S, Bertilsson S *et al.* (2013) Unveiling distribution patterns of freshwater phytoplankton by a next generation sequencing based approach. *PLoS ONE*, **8**, e53516.

Elbrecht V, Leese F (2015) Can DNA-based ecosystem assessments quantify species abundance? Testing primer bias and biomass—sequence relationships with an innovative metabarcoding protocol. *PLoS ONE*, **10**, e0130324.

Esling P, Lejzerowicz F, Pawlowski J (2015) Accurate multiplexing and filtering for high-throughput amplicon-sequencing. *Nucleic Acids Research*, **43**, 2513–2524.

Falasco E, Bona F, Badino G, Hoffmann L, Ector L (2009) Diatom teratological forms and environmental alterations: a review. *Hydrobiologia*, **623**, 1–35.

Godhe A, Asplund ME, Härnström K *et al.* (2008) Quantification of diatom and dinoflagellate biomasses in coastal marine seawater samples

© 2017 John Wiley & Sons Ltd

by real-time PCR. *Applied and Environmental Microbiology*, **74**, 7174–7182.
Gong J, Dong J, Liu X, Massana R (2013) Extremely high copy numbers and polymorphisms of the rDNA operon estimated from single cell analysis of oligotrich and peritrich ciliates. *Protist*, **164**, 369–379.
Gouy M, Guindon S, Gascuel O (2010) SeaView Version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Molecular Biology and Evolution*, **27**, 221–224.
Hill BH, Stevenson RJ, Pan Y *et al.* (2001) Comparison of correlations between environmental characteristics and stream diatom assemblages characterized at genus and species levels. *Journal of the North American Benthological Society*, **20**, 299–310.
Hofmann G, Werum M, Lange-Bertalot H (2011) k ; über 700 der häufigsten Arten und ihre Ökologie. Gantner.
Hürlimann J, Niederhauser P (2007) Méthodes d'analyse et d'appréciation des cours d'eau. Diatomées Niveau R (region). Etat de l'environnement n° 0740. Office fédéral de l'environnement, Berne.
Jiang Y, Xu H, Hu X *et al.* (2011) An approach to analyzing spatial patterns of planktonic ciliate communities for monitoring water quality in Jiaozhou Bay, northern China. *Marine Pollution Bulletin*, **62**, 227–235.
Keck F, Bouchez A, Franc A, Rimet F (2016) Linking phylogenetic similarity and pollution sensitivity to develop ecological assessment methods: a test with river diatoms. *Journal of Applied Ecology*, **53**, 856–864.
Kelly MG, Adams C, Graves AC (2001) *The Trophic Diatom Index: a User's Manual*; Revised Edition. Environment Agency, Rotherham.
Kelly M, Bennett C, Coste M *et al.* (2008) A comparison of national approaches to setting ecological status boundaries in phytobenthos assessment for the European Water Framework Directive: results of an intercalibration exercise. *Hydrobiologia*, **621**, 169–182.
Kermarrec L, Franc A, Rimet F *et al.* (2013) Next-generation sequencing to inventory taxonomic diversity in eukaryotic communities: a test for freshwater diatoms. *Molecular Ecology Resources*, **13**, 607–619.
Kermarrec L, Franc A, Rimet F *et al.* (2014) A next-generation sequencing approach to river biomonitoring using benthic diatoms. *Freshwater Science*, **33**, 349–363.
Krammer K, Lange-Bertalot H (1986-1992) *Bacillariophyceae. Süsswasserflora von Mitteleuropa*, Stuttgart Germany.
Lange-Bertalot H (2001) *Diatoms of the European Inland Waters and Comparable Habitats*. Ruggell, Lichtenstein.
Lange-Bertalot H, Metzeltin D (1996) Indicators of oligotrophy: 800 taxa representative of three ecologically distinct lake types : carbonate buffered, oligodystrophic, weakly buffered soft water. *Iconographia Diatomologica*, **2**, 390.
Lee S, Basu S, Tyler CW, Wei IW (2004) Ciliate populations as bio-indicators at Deer Island Treatment Plant. *Advances in Environmental Research*, **8**, 371–378.
MacGillivary ML, Kaczmarska I (2011) Survey of the efficacy of a short fragment of the rbcL gene as a supplemental DNA barcode for diatoms. *The Journal of Eukaryotic Microbiology*, **58**, 529–536.
Pawlowski J, Esling P, Lejzerowicz F, Cedhagen T, Wilding TA (2014) Environmental monitoring through protist next-generation sequencing metabarcoding: assessing the impact of fish farming on benthic foraminifera communities. *Molecular Ecology Resources*, **14**, 1129–1140.
Pawlowski J, Lejzerowicz F, Apotheloz-Perret-Gentil L, Visco J, Esling P (2016) Protist metabarcoding and environmental biomonitoring: time for change. *European Journal of Protistology*, **55**, 12–25.
Poikane S, van den Berg M, Hellsten S *et al.* (2011) Lake ecological assessment systems and intercalibration for the European Water Framework Directive: aims, achievements and further challenges. *Procedia Environmental Sciences*, **9**, 153–168.
Reichardt E (1999) Zur Revision der Gattung Gomphonema: Die Arten um G. affine/insigne, G. angustatum/micropus, G. acuminatum sowie gomphonemoide Diatomeen aus dem Oberoligozän in Böhmen. *Iconographia Diatomologica*, **8**, 250.
Rimet F, Trobajo R, Mann DG *et al.* (2014) When is sampling complete? the effects of geographical range and marker choice on perceived diversity in *Nitzschia palea* (Bacillariophyta). *Protist*, **165**, 245–259.
Rimet F, Chaumeil P, Keck F *et al.* (2016) R-Syst::diatom: an open-access and curated barcode database for diatoms and freshwater monitoring. *Database: The Journal of Biological Databases and Curation*, **2016**, baw016.
Rovira L, Trobajo R, Sato S, Ibáñez C, Mann DG (2015) Genetic and physiological diversity in the diatom *Nitzschia inconspicua*. *The Journal of Eukaryotic Microbiology*, **62**, 815–832.
Schönfeld J, Alve E, Geslin E *et al.* (2012) The FOBIMO (FOraminiferal BIo-MOnitoring) initiative—Towards a standardised protocol for soft-bottom benthic foraminiferal monitoring studies. *Marine Micropaleontology*, **94–95**, 1–13.
Stamatakis A (2014) RAxML Version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, **30**, 1312–1313.
Stevenson RJ, Pan YD, van Dam H (2010) Assessing environmental conditions in rivers and streams with diatoms, In: *The Diatoms: Applications for the Environmental and Earth Sciences*, 2nd edn, pp. 57–85. Cambridge University Press, Cambridge, UK.
Swiss Federal Council (1998) *Waters Protection Ordinance*, https://www.admin.ch/opc/en/classified-compilation/19983281/index.html.
Van den Wyngaert S, Möst M, Freimann R, Ibelings BW, Spaak P (2015) Hidden diversity in the freshwater planktonic diatom *Asterionella formosa*. *Molecular Ecology*, **24**, 2955–2972.
Vidovic J, Dolenec M, Dolenec T, Karamarko V, Žvab Rožič P (2014) Benthic foraminifera assemblages as elemental pollution bioindicator in marine sediments around fish farm (Vrgada Island, Central Adriatic, Croatia). *Marine Pollution Bulletin*, **83**, 198–213.
Visco J, Apothéloz-Perret-Gentil L, Cordonier A *et al.* (2015) Environmental monitoring: inferring the diatom index from next-generation sequencing data. *Environmental Science & Technology*, **49**, 7597–7605.
Weber AA-T, Pawlowski J (2013) Can abundance of protists be inferred from sequence data: a case study of foraminifera (P López-García, Ed,). *PLoS ONE*, **8**, e56739.
Whitton BA, Rott E, Friedrich G (1991) Use of algae for monitoring rivers. *Journal of Applied Phycology*, **3**, 287–288.
Yu DW, Ji Y, Emerson BC *et al.* (2012) Biodiversity soup: metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. *Methods in Ecology and Evolution*, **3**, 613–623.
Zelinka M, Marvan P (1961) Zur Präzisierung der biologischen Klassifikation der Reinheit fliessender Gewässer. *Archiv fur Hydrobiologie*, **57**, 389–407.
Zimmermann J, Jahn R, Gemeinholzer B (2011) Barcoding diatoms: evaluation of the V4 subregion on the 18S rRNA gene, including new primers and protocols. *Organisms Diversity & Evolution*, **11**, 173–192.
Zimmermann J, Abarca N, Enke N *et al.* (2014) Taxonomic reference libraries for environmental barcoding: a best practice example from diatom research. *PLoS ONE*, **9**, e108793.
Zimmermann J, Glöckner G, Jahn R *et al.* (2015) Metabarcoding vs. morphological identification to assess diatom diversity in environmental studies. *Molecular Ecology Resources*, **15**, 526–542.

J.P., L.A.P.G. and P.E. conceived and designed the experiments. L.A.P.G. performed the experiments and analysed the data. A.C., F.S. and J.I. performed all the morphological work. J.P. and L.A.P.G. wrote the manuscript.

Data accessibility

Illumina raw data are deposited in Dryad (doi:10.5061/dryad.8m3kv).

© 2017 John Wiley & Sons Ltd

Supporting Information

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

Fig. S1 Map of sampling sites.

Fig. S2 A. Schematic representation of the calculation of the D and G value for the molecular method.

Fig. S3 RAxML tree with sequences from the database and the OTUs from the HTS analysis.

Fig. S4 Venn diagram of morphospecies represented in the database (yellow), morphological analysis (blue) and found in HTS dataset by taxonomic assignment method (red).

Fig. S5 Scatter plot of the relative frequency for all the assigned species.

Fig. S6 Scatter plot of the relative frequency for the 4 most represented morphospecies in the HTS dataset.

Fig. S7 Graphical table representing the DI-MOLTAXFREE cross-validation results.

Fig. S8 Box plot of the DI-MOLTAXFREE Cross-Validation 25:75 test.

Fig. S9 Graphical representation of the DI-MOLTAXFREE Cross-Validation 25:75 test.

Fig. S10 Bar plots representing the proportion of D (blue) and G (green) values in function of their change during the cross-validation test.

Table S1 Illumina run code, station code, location, sampling date and geographic references for each site used in this study.

Table S2 List of database entries description with their NCBI or Rsyst accession number. Environmental sequences (ENV) are marked.

Table S3 List of primers and tags used in this study.

Table S4 Filtering process of the four Illumina runs used in this study.

Table S5 List of OTUs with their number of reads per sample.

Table S6 Number of OTUs from HTS analysis and species from morphological analysis for each site.

Table S7 List of species found during the morphological analysis with their relative abundance per site.

Table S8 DI-CH values given by morphology (DI-CH), taxonomic assignment (DI-MOLTAXASSIGN) and Leave-one-out cross-validation (DIMOLTAXFREE) for each site.

Table S9 Comparison of DG values given by morphology (D and G) and molecular (MOL-D and MOL-G) indices for each assigned OTUs.

© 2017 John Wiley & Sons Ltd