Exploring the capacity of aquatic biofilms to act as environmental DNA samplers: Test on macroinvertebrate communities in rivers

Sinziana Rivera, Valentin Vasselon, Nathalie Mary, Olivier Monnier, Fréderic Rimet, Agnes Bouchez

Exploring the capacity of aquatic biofilms to act as environmental DNA samplers: test on macroinvertebrate communities in rivers

Sinziana F. Rivera[1], Valentin Vasselon[2], Nathalie Mary[3], Olivier Monnier[4], Fréderic Rimet[1] & Agnès Bouchez[1]

[1] *Université Savoie Mont-Blanc, INRAE, UMR CARRTEL, 75bis av. de Corzent, FR-74200 Thonon-les-Bains, France*

[2] *OFB, Pôle R&D «ECLA», Site INRAE, UMR CARRTEL, 75bis av. de Corzent, FR-74200 Thonon-les-Bains, France*

[3] *ETHYC'O, B.P. 13821, 98803 Nouméa Cedex, Nouvelle-Calédonie.*

[4] *OFB, Service Mobilisation de la Recherche, 5 square Félix Nadar, FR-94300 Vincennes, France*

Corresponding Author:

Sinziana Rivera

Université Savoie Mont-Blanc, INRAE, UMR CARRTEL,

75bis av. de Corzent

CS 50511, FR-74200 Thonon-les-Bains, France

sinzianaflorina@gmail.com

Abstract

Aquatic biofilms are heterogeneous assemblages of microorganisms surrounded by a matrix of extracellular polymeric substances (EPS). Recent studies suggest that aquatic biofilms can physically act as sorptive sponges of DNA. We took the opportunity from already available samples of stone biofilms and macroinvertebrates specimens collected in parallel at the same sites to test the capacity of biofilms to act as DNA samplers of macroinvertebrate communities in streams. Macroinvertebrate communities are usually studied with metabarcoding using the DNA extracted from their bodies bulk samples, which remains a time-consuming approach and involves the destruction of all individual specimens from the samples. The ability of biofilms to capture DNA was explored on 19 rivers sites of a tropical island (Mayotte Island, France). First, macroinvertebrate specimens were identified based on their morphological characteristics. Second, DNA was extracted from biofilms, and macroinvertebrate communities were targeted using a standard COI barcode. The resulting morphological and molecular inventories were compared. They provided comparable structures and diversities for macroinvertebrate communities when one is working with the unassigned OTU data. After taxonomic assignment of the OTU data, diversity and richness were no longer correlated. The ecological assessment derived from morphological bulk samples was conserved by the biofilms samples. We also showed that the biofilm method allows to detect a higher diversity for some organisms (Cnidaria), that is hardly accessible with the morphological method. The results of this study exploring the DNA signal captured by natural biofilms are encouraging. However, a more detailed study integrating more replicates and comparing the biodiversity signal based on both morphological and molecular bulk macroinvertebrate samples to the one captured by biofilms will be necessary. Better

understanding how the DNA signal captured by natural biofilms represents the biodiversity of a given sampling site is necessary before considering its use for bioassesment applications.

Keywords: Biofilms, Metabarcoding, Environmental DNA, Macroinvertebrates

1. Introduction

Aquatic biofilms can be described as an assemblage of bacteria, algae, fungi, and protozoa that are enclosed within a matrix composed of extracellular polymeric substances (EPS) and develop on wetted surfaces (Romaní et al., 2016). EPS from the biofilm matrix are generally molecules of high molecular weight with both positive and negative binding sites, each playing an important role in molecules' sorption (Wolfaardt et al., 1998). Mora-Gómez et al. (2016) indicated that aquatic biofilms can trap particulate material from the water column and, thus, increase its concentration in contrast to the water column. Several studies have shown that biofilms are able to sorb organic compounds such as pesticides (e.g., Diuron) (Flemming & Wingender, 2010; Chaumet et al., 2019) or polycyclic aromatic hydrocarbons (PAHs) (Froehner et al., 2012). Recently, Shogren et al. (2018) observed that the presence of benthic biofilms in recirculating streams significantly decreased extra-cellular fish DNA concentrations from the water column. Hence, they hypothesized that the biofilm may entrap extra-cellular DNA, resulting in either the temporary or permanent removal of DNA particles from the water column. The quantity of trapped material depends on the biofilm structure and composition, which are greatly influenced by the chemical, physical, and biological conditions (e.g., substrate, light, temperature, nutrients, oxygen, pH, flow, pressure, salinity, and grazing) (Mora-Gómez et al., 2016). The sensitivity of biofilm and, in particular, diatoms (its main component) (Morin et al., 2016) to water chemistry, as well as the geomorphological and physical characteristics of rivers and lakes, led to the use of these organisms for the

monitoring of aquatic ecosystems. Diatoms from biofilms have, indeed, been used as bioindicators of water quality for centuries (Rimet, 2012) and, over the last decade, diatom metabarcoding from biofilms has been applied successfully to the assessment of the ecological status of rivers and lakes (Vasselon et al., 2017; Rivera et al., 2018; Mortágua et al., 2019; Rivera et al., 2020). In the same way as biofilm, environmental biota such as sponges have recently been suggested to have the ability to act as natural eDNA samplers in marine environments (Mariani et al., 2019).

Another group of organisms widely used to monitor the quality of freshwater ecosystems are macroinvertebrates. The conventional method is based on the collection of specimens in the field and their subsequent identification at the lowest practical level (generally genus or species) (Bailey et al., 2001) on the basis of their morphological characteristics. However, this requires taxonomic expertise, especially at immature life stages (larvae and nymphs), where diagnostic morphological characters are not fully developed and identification to the species level is generally not possible (Sweeney et al., 2011). Additionally, during field sampling, macroinvertebrate specimens can be damaged, which hamper their identification. Recent advances in sequencing technologies enabled the development of DNA metabarcoding (Taberlet et al., 2012). This alternative approach enables to identify specimens in multiple samples simultaneously by combining DNA barcoding (Hebert et al., 2003) with High-Throughput Sequencing (HTS). Thus, macroinvertebrate specimens are identified based on a standard fragment of their DNA (DNA barcode) instead of their morphology. This overcomes the taxonomic challenges of the conventional morphological approach. Metabarcoding studies on macroinvertebrate communities are usually performed on DNA extracted from bulk samples. This approach consists of collecting macroinvertebrates in the field, sorting and mixing them together, and then extracting DNA from this mixture. This new method showed

good results in identifying macroinvertebrate specimens for monitoring purposes (Hajibabaei et al., 2011; Carew et al., 2013; Elbrecht & Leese, 2017; Serrana et al., 2019). Nevertheless, some issues remain problematic as is the relationship between biomass or species counts and read abundances (Elbrecht et al., 2017).

Additionally, macroinvertebrate metabarcoding involves the destruction of all individual specimens from the samples, as DNA extraction protocols require the homogenization of the biomass from all organisms. Furthermore, this remains time-consuming, as macroinvertebrate specimens must still be collected and sorted prior to DNA extraction. However, a recent study (Pereira-da-Conceicoa et al., 2020) proposed a much faster approach on unsorted bulk samples, including debris, which reduces sample-processing time and allows identifying taxa at a much finer resolution.

Based on this, the aim of this study is to evaluate the capacity of aquatic biofilms to act as passive samplers of eDNA. This property can be investigated by targeting macroinvertebrate eDNA, which is composed of both intra- and extra-cellular DNA (Taberlet et al., 2018). Therefore, we extracted DNA directly from biofilms and performed DNA metabarcoding of the macroinvertebrate community using a standard DNA barcode (i.e., COI). If macroinvertebrate eDNA is present, this approach could be used to produce a taxonomic inventory of the community. This alternative approach could represent a cheaper, faster, and less complicated method for macroinvertebrate metabarcoding studies. First, we hypothesized that biofilms can catch macroinvertebrate eDNA. Second, we hypothesized that macroinvertebrate community structure and diversity, obtained from biofilm metabarcoding on macroinvertebrate DNA amplicons, could be comparable to that obtained using the conventional approach.

To test these hypotheses, we took the opportunity to use macroinvertebrate specimens and natural biofilms already sampled at same river sites for biomonitoring purposes. Samples came from several river sites from a French tropical island monitoring network (Mayotte). First, macroinvertebrate inventories were obtained through conventional sampling with a Surber net and morphological-based identifications using a binocular. Second, macroinvertebrate inventories were identified from biofilms using DNA metabarcoding. Finally, we compared the macroinvertebrate community structure, diversity and richness obtained with both approaches in order to evaluate the potential of biofilms to be used as environmental DNA sensors. We also explored if the ecological assessment using the taxonomic richness of insect families from the Ephemeroptera, Plecoptera and Trichoptera orders (EPT) provided by the conventional morphological approach was preserved with the biofilm molecular approach.

2. Material and Methods

2.1. Study site

Benthic macroinvertebrates and biofilm samples were collected in August 2017 from 19 sites (12 rivers) of Mayotte, a French tropical island located in the Comoros archipelago in the Mozambique Channel (Figure 1). Sampling sites are part of the EU Water Framework Directive (WFD) regular river monitoring network of Mayotte which corresponds to river sites under poor to good ecological quality conditions (Vasselon et al., 2017, Tapolczai et al., 2017), with physicochemical gradients described in Tapolczai et al. (2019a). Briefly, in these short tropical rivers poor quality conditions are associated with higher nutrient and organic matter concentrations, while good quality conditions are associated with well-oxygenated

waters. An overview of *in situ* conditions measured during the 2017 sampling campaign is available at the supplementary dataS1.

2.2. Field sampling

Macroinvertebrate specimens were collected using a Surber net (0.05 m^2, 250 µm mesh size). At each site, samples of macroinvertebrates were collected from marginal and dominant substrata following the French standard protocol (Afnor, 2016; a summary of the sampling methodology is given in Supporting Information 1). The sampled habitats were diverse: bedrock, clay, silt, sand, gravel, pebbles, cobbles, boulders, leaf litter, tree roots, woody debris, macrophytes, and moss. Immediately after collection, individual samples were fixed with ethanol (90%) to create a final concentration of at least 70%.

In the laboratory, macroinvertebrates were sorted and preserved in 70% ethanol. Morphological identification of the specimens was conducted to the lowest possible taxonomic level, that is to say species level (if possible), under a binocular loupe (Leica MZ 7.5; cold light source Schott KL1500 LED) using both published and unpublished determination keys (Day et al., 2002; De Moor et al., 2003a, 2003b; Keith et al., 2006; Valade et al., 2007; Johanson & Mary, 2009; Tachet et al., 2010; Pešić et al., 2015; Mary, 2017; Pešić et al., 2018; Smit et al., 2010).

At each of the 19 sites, biofilm sampling was performed from five stones collected in flowing parts within the river stretch where the macroinvertebrate specimens were collected. The biofilms were recovered by scraping of the stones' upper surface with a clean toothbrush following the European standard EN 13946 (Afnor, 2014). The 19 samples were stored in a

sterile 50 ml Falcon© tube, fixed in ethanol (70% final concentration) according to European protocols (CEN, 2018), and kept cold (4–7°C) until molecular treatment.

2.3. Extraction of total DNA from biofilms

DNA extractions were performed using 2 mL of the preserved biofilm samples. After centrifugation at 13,000 rpm for 30 min at 4ºC, a supernatant containing ethanol was removed and the pellets were used as a starter for DNA extraction. Total genomic DNA was isolated using the NucleoSpin® Soil Kit (Macherey-Nagel, GmbH, Düren, Germany) following the manufacturer's instructions. Lysis buffer SL1 combined with the chemical additive Enhancer SX was used for the mechanical lysis of the sample material. The quality and quantity of the extracted DNA were checked using a NanoDrop spectrophotometer.

2.4. COI amplification and sequencing

A 461-bp (including primers) fragment of the mitochondrial cytochrome c oxidase subunit I gene (COI) was amplified using a two-step polymerase chain reaction (PCR). To obtain technical replicates, two subsamples of each DNA extract were used for subsequent PCR amplification and HTS. For PCR1, each DNA extract was amplified in triplicate using the equimolar mixes of the BF2 (5'-GCHCCHGAYATRGCHTTYCC-3') and BR2 (5'-TCDGGRTGNCCRAAARAAAYCA-3') primer sets (Elbrecht & Leese, 2017) to which sequencing adapters had been added. Half of the P5 (CTTTCCCTACACGACGCTCTTCCGATCT) and P7 (GGAGTTCAGACGTGTGCTCTTCCGATCT) Illumina adapters were included in the 5′part

of the BF2 and BR2 primers, respectively. PCR1 amplifications were performed using a 24 µL mix containing 1 µL of extracted DNA, 0.625 U of TaKaRa LA Taq polymerase (TaKaRa Bio, Sugatsu, Japan), 2.5 µL of 10X buffer, 1.25 µL of 10µM of each primer, 1.25 µL of 10 g L -1 bovine serum albumin (BSA), 2 µL of 2.5 mM deoxynucleotide (dNTP), and 15.6 µL of H_2O (molecular-biology grade). PCR conditions were an initial denaturation of DNA at 94°C for 3 min followed by 30 cycles of denaturation at 94 °C for 30 s, annealing at 50 °C for 30 s and extension at 65 °C for 2 min 30 s. Two negative PCR controls were performed in order to detect potential contamination during the amplification step. After PCR, amplification of the COI barcode in each sample was confirmed by agarose gel electrophoresis using a 1.5 % agarose gel. All negative controls appeared negative. Because PCR controls did not showed positive amplicons we did not include them in the Illumina library construction and sequencing. However, it would have been interesting to do this as a measure to minimize tag jumps (Schnell et al. 2015). For the PCR2, the three PCR1 replicates prepared for each DNA extract were pooled and sent to the "GenoToul Genomics and Transcriptomics" platform (GeT-PlaGe, Auzeville, France), where subsequent laboratory preparations were performed. The 38 PCR1 amplicons were purified and used as templates in the PCR2, which used Illumina tailed primers targeting the half of the P5 and P7 sequences. Finally, the 38 generated PCR2 amplicons were dual-indexed using homemade 8-base indices and pooled into a single tube. The final pool was sequenced on an Illumina MiSeq flow cell using the V3 paired-end sequencing kit (250 bp × 2) to ensure a high sequencing depth. The quality of the run was checked internally using PhiX, and then each pair-end sequences were assigned to its sample with the help of the previously integrated index.

2.5. Reference library construction

To have a COI macroinvertebrate reference library that is compatible with our bioinformatics pipeline using the Mothur software (Schloss et al., 2009), a local library was built. To that end, COI macroinvertebrate sequences were downloaded from the public data portal of the Barcode of Life Data System v4 (BOLD) with special emphasis on macroinvertebrate taxa identified from Mayotte Island in previous monitoring studies, as well as macroinvertebrate taxa considered as part of the 15 most relevant freshwater invertebrate taxonomical groups for bioassessment (see Elbrecht & Leese [2017] for more details). Sequences were downloaded in FASTA (a text-based file containing the nucleotide sequences) and TSV (a text-based file containing the taxonomic affiliation of each nucleotide sequence) formats. All DNA sequences were filtered by length and quality according to the following criteria: minimum length =400 bp, homopolymer < 7 bp, and ambiguous base = 0. The remaining sequences were combined and analyzed as a whole. Sequences were first aligned in MAFFT and then the 421 bp region was targeted by looking for the BF2 and BR2 primer set. Sequences presenting more than one mismatch in the forward primer and the reverse primer were removed. Finally, sequences were dereplicated and alignment gaps were removed. Taxonomical information about the remaining sequences was recovered from the TSV files. The good identity of the sequences was tested by assigning the library to itself using the naïve Bayesian method (Wang et al., 2007) with a confidence score threshold of 75%. Sequences presenting mismatches in the taxonomical assignment between Mothur and BOLD were removed. The final reference library resulted in 15,399 DNA sequences of the COI barcode from 5 phyla, 11 classes, 19 orders, 526 families, 2974 genera and 6749 species. An overview of the most abundant orders is presented in Figure 2.

2.6. Local reference library adequacy

The adequacy of our local library for the study of the macroinvertebrate DNA diversity found in biofilm samples was assessed through a comparison of its taxonomic list to the list of taxa detected in the morphological inventories. A comparison was performed at the phylum, class, order, and family levels using the interactive tool Venny v.2.1 (Oliveros, 2016). A comparison below family level was not performed because morphological taxa lists contained taxa identified at taxonomic ranks between family and genus levels (e.g., tribe) that were not considered in our local reference library—thus, hampering their comparison.

2.7. HTS data analyses

Sequence data processing was performed starting from demultiplexed Miseq data that were first paired and merged (paired sequences overlap > 35 bp and mismatches <1). Overlapped FASTA files were then filtered by length and quality according to the following criteria: minimum length =451 bp, maximum length =471 bp, homopolymers <8 bp, and ambiguous base = 0. All the resulting FASTA files were combined and de-replicated to keep only unique sequences with read abundances > 3. Chimeric DNA sequences were removed using the *Vsearch* algorithm. The resulting sequences were then assigned a taxonomy using our local macroinvertebrate library and the naïve Bayesian method (Wang et al., 2007) with a confidence score threshold of 75% (i.e. in a bootstrap, the percentage of times that the sequence must match to the same taxonomy in order to be assigned a definitive taxonomic name). This first assignation of all environmental sequences allowed us to detect and subsequently eliminate non-macroinvertebrates sequences. Only DNA sequences belonging to macroinvertebrates were kept for further analysis. Subsequently, a similarity distance matrix was generated using the *dist.seqs* command. Based on this distance matrix, sequences belonging to closely related groups were clustered into OTUs using the furthest neighbor

algorithm at a 97% similarity level. A list of OTUs and their relative abundances was produced for each of the samples based on read abundances per OTU.

Molecular taxa lists were then created by providing a taxonomy to each OTU using the *classify.otu* command with a consensus confidence threshold of 80% (i.e., the consensus taxonomy of the individual sequence units [ISU] within each OTU) (Schloss et al., 2009). Finally, a DNA representative sequence was determined for each OTU using the *get.oturep* command in Mothur. For statistical analysis, the OTU abundances of the two technical replicates were merged and analyzed as a single sample.

As mentioned previously, reads that did not match sequences from our local library remained unclassified and were removed after the taxonomic assignment step. To ensure that these unclassified reads corresponded to non-macroinvertebrate organisms, they were blasted against the NCBI database (Sayers et al., 2019) to determine their taxonomic identity. For that, unclassified reads were clustered into OTUs using the furthest neighbor algorithm at a 97% similarity level. Then a DNA representative sequence was determined for each OTU and blasted against the NCBI database (Standard Nucleotide BLAST) (Sayers et al., 2019). This step allowed us to confirm the choice of our pipeline settings and the good coverage of our database.

2.8. Comparison of community structures obtained with both morphology and biofilms' molecular approaches

The structures of macroinvertebrate communities derived from both morphological and molecular approaches were compared using a Mantel test (Mantel, 1967). This test assesses the correlation between two matrices, allowing the estimation of the relationship between them (Legendre & Fortin, 2010). We wanted to know if the macroinvertebrate community

structure given by the DNA biofilm approach was similar to that given by the classical morphological one. The test was performed between the morphological and the molecular OTU data based on Bray Curtis distances. To avoid biases related to quantification, the comparison of the structure of macroinvertebrate communities derived from both approaches was also compared in terms of presence/absence (based on Jaccard distances).

Similarly, after taxonomic assignment of the OTU data, we tested whether the community structures derived from both morphological and molecular taxonomical inventories were comparable using a Mantel test. The tests were performed in the statistical software R (R Core Team, 2019).

2.9. Comparison of biodiversity assessed with morphology and biofilms' molecular approaches

We compared the diversity and richness of macroinvertebrate communities assessed from morphological bulk samples and biofilm samples. The Shannon index (Shannon & Weaver, 1949) and the observed richness were calculated from the morphological inventories (number of unique taxa occurring in a sample) and the OTU data (number of unique OTU occurring in a sample). Shannon index (H') was calculated as follow: $H' = -\sum_{n=1}^{n} p_i \times \ln p_i$ were p_i is the proportion of total sample represented by species i. Correlations between biodiversity and richness metrics derived from both approaches were tested using the *cor.test* function in the statistical software R (R Core Team, 2019).

Similarly, after the taxonomic assignment of the OTU data, we tested whether diversity and richness metrics derived from both morphological and molecular taxonomical inventories were correlated.

The five most abundant taxa detected with both approaches are compared in Table 1. Additionally, a comparison between both approaches at family level in terms of presence/absence is presented in Table 2.

2.10. Ecological characterization of the sampling sites

The Ephemeroptera (mayflies), Plecoptera (stoneflies) and Trichoptera (caddisflies) (EPT) orders are considered as good indicators of water quality due to their low tolerance to water pollution (Stoyanova et al., 2014). Large percentages of EPT taxa in rivers indicate a high water quality. Water quality of Mayotte rivers is assessed by calculating the EPT index (Lenat, 1987) which is equal to the total number of families represented within the Ephemeroptera and Trichoptera orders (Plecoptera order is absent in Mayotte rivers). Based on the EPT's scores, water quality can be classified in 4 classes: poor (EPT < 2), moderate (2 ≤ EPT < 4), good (5 ≤ EPT < 7) or very good (EPT > 7). This classification is adapted for Mayotte rivers and was based on official's assessment reports (Mary, 2018).

Since many of our sequences were assigned not lower than phylum or class levels, we were not able to determine EPT values directly from the molecular inventories (see Table 2). However, to get an idea of whether the ecological assessment obtained with both approaches is comparable, we used the morphological water quality groups inferred from the morphological EPT values to characterize each sampling site. We performed non-metric multidimensional scaling (nMDS) based on Bray Curtis distances to compare the distribution of the sampling sites (characterised by their water quality group) when assessed through the three approaches: morphology, biofilm molecular with unassigned or assigned OTUs. nMDS were performed in the statistical software PAST 3.25 (Hammer et al., 2001).

3. Results

3.1. Is the reference library adapted to the studied rivers?

The comparison between the morphological inventories and our local reference library showed that the library is quite complete for the analysis of the macroinvertebrate communities in Mayotte rivers. Figure 3 summarizes the comparison between the taxa observed in the morphological inventories and the available taxa in our local library. At the family level, 77% of the observed taxa had references in the library. Only 13 families over 57 that were identified on the basis of morphological criteria were missing from our local reference library. Most of them corresponded to families with abundances lower than 1%.

3.2. Morphological results

A total of 96 taxa belonging to 57 families, 68 genera, and 22 species were identified from all habitats among all the samples. The dominant taxa were *Chironomus* (19,033 individuals, 28.1%), *Simulium* (8,860 individuals, 11.6%), Tanytarsini (7,705 individuals, 11.2%), Naididae (4,152 individuals, 6.1%) and *Thiara scabra* (4,129 individuals, 6.1%) (Supplementary data S2).

3.3. Biofilm molecular results

After contigs, we obtained 14,919,139 reads, among which 3,556,583 (24%) passed the quality filters. From these, 962,220 reads (27%) were assigned to macroinvertebrates using our local reference library, while the remaining 73% could not be assigned to any macroinvertebrate taxa (Figure 4A). The clustering of DNA sequences belonging to macroinvertebrates at the 95% level resulted in 670 OTUs. The number of OTUs per sample

ranged from 41 to 202, with an average of 128 OTUs per sample. The taxonomical assignment of OTUs resulted in 11 genera belonging to five orders and three classes. The dominant molecular taxa were Insecta_unclassified (34.5% of the reads comprising 154 OTUs), Diptera_unclassified (26.3%, 17 OTUs), Hydrozoa_unclassified (10%, 308 OTUs), *Craspedacusta sowerbii* (7.3%, 4 OTUs), and *Hydra sinensis* (6.3%, 6 OTUs) (Supplementary data S2).

Sequences that could not be identified with our local macroinvertebrate reference library represented 73% of the sequence data and could be clustered into 5,256 OTUs. We blasted on NCBI (Sayers et al., 2019) the dominant OTUs (OTUs with abundances higher than 1% and representing 74% of the unidentified sequences). Sixty-six percent of the OTUs matched with algae, 2% matched with bacteria, 4% matched with fungi (4%), and 2% matched with other metazoans (Figure 4B). Sixty-six percent of non-macroinvertebrate sequences corresponds with algae, a major component of biofilms.

3.4. Comparison of community structures obtained with both morphology and biofilms' molecular approaches

The comparison between the molecular and morphological inventories by means of Mantel test revealed a significant relationship between the morphological and the molecular OTU data (r=0.555, p value= 0.0001). This was true also for the morphological and molecular inventories obtained after the taxonomic assignment of the OTU data (r=0.549, p value= 0.0001). When considering presence/absence inventories, the relationship between morphological and molecular OTU macroinvertebrate structure was also significant (r= 0.707, p value= 0.0001). On the contrary, after the taxonomic assignment of the OTU data the

relation between the morphological and molecular inventories in terms of presence/absence was considerable reduced (r= 0.38, p value= 0.0068).

3.5. Comparison of biodiversity assessed with morphology and biofilms' molecular approaches

The Shannon index and the observed richness were calculated for both morphological and molecular OTU data. Correlation tests showed that macroinvertebrate diversity and richness obtained using both approaches are correlated (R=0.48, p value=0.038 and R=0.61, p value=0.006, respectively) (Figure 6). However, the obtained diversity and richness metrics were no longer correlated after the taxonomical assignment of OTUs ($R_{taxa\text{-}Shannon}$=0.34, p value=0.15 and $R_{taxa\text{-}Richness}$=0.14, p value=0.56, respectively).

The relative abundances of taxa at phylum and class levels are compared between morphological and molecular inventories in Figure 7. Comparison of relative abundances was not performed below the class level because the taxonomic level reached with the molecular approach was not the same as that reached with the morphological one, making this comparison difficult. The two dominant Phyla (Arthropoda and Mollusca) were detected with both approaches. Conversely, the molecular approach was unable to detect the Phyla Platyhelminthes and Nemertea but detected a large proportion of Cnidaria, which was seldom observed with the morphological approach (only one individual observed in a single station). Regarding the class level, in both cases, Arthropoda correspond mainly to Insecta class. Mollusca taxa, for their part, correspond mainly to the Gastropoda class. The Cnidaria phylum detected using both approaches corresponded to the Hydrozoa class. In many cases, molecular sequences were not assigned to lower taxonomic levels than phylum and remained "phylum_*unclassified*" (e.g., arthropoda_*unclassified*, annelida_*unclassified*,

mollusca_*unclassified*). Furthermore, when comparing the five most abundant taxa detected with both approaches we observed great differences (Table 1). We also observed great differences between both approaches at the family level (Table 2). Only 11 families could be detected with the biofilm molecular approach. Looking at the EPT taxa, no families from Ephemeroptera order could be detected with the molecular method compared to the morphological one were 3 families where observed (Baetidae, Caenidae; Leptophlebiidae). Concerning the Trichoptera order, the biofilm approach detected only 2 families (Ecnomidae and Philopotamidea) compared to the morphological approach which detected 8 families.

3.6. Ecological characterization of the sampling sites

The ecological characterization of the sampling sites obtained with the morphological approach using the EPT index was conserved with the biofilms approach (Figure 5). Both approaches clearly separate sites with a bad water quality class from the rest of the sites. For the biofilm molecular approach this was true before and after the taxonomic assignment of the OTU data. Morphological EPT index values are available at Supplementary data S3.

4. Discussion

Metabarcoding studies of macroinvertebrate communities are usually performed on DNA extracted from bulk samples of collected and sorted specimens (Carew et al., 2013; Elbrecht & Leese, 2017; Hajibabaei et al., 2011). More rarely, eDNA was extracted directly from preservative ethanol (Hajibabaei et al., 2012), water (Bista et al., 2017; Fernandez et al., 2019;

Hajibabaei et al., 2019; Schneider et al., 2016; Thomsen & Willerslev, 2015) or sediments (Aylagas et al., 2016) to detect freshwater macroinvertebrates.

The retention of eDNA in biofilms has already been studied either as a factor influencing eDNA degradation, or as a factor removing eDNA from the water column (Shogren et al., 2018; Shogren et al., 2016). To our knowledge, this is the first study that looks for macroinvertebrates' eDNA in biofilms. In this exploratory study we demonstrated, first, that environmental river biofilms can hold macroinvertebrate eDNA. Second, we showed similarities between the macroinvertebrate community structures derived from the conventional approach (Surber sampling and morphological identification) and from the approach using biofilm-captured eDNA. Third, we showed that diversity (Shannon index) and richness metrics assessed through eDNA in biofilms, based on OTU data, are correlated to those obtained using the conventional approach. Finally, we showed that the ecological assessment derived from the morphological inventories, based on the EPT index, was conserved when using the macroinvertebrate inventories assessed from eDNA in biofilms.

However, even if these results seem promising, there are a number of discrepancies between both methods. First, even if we observed a good correlation between the diversity and richness metrics obtained using both approaches, this correlation was observed only when using OTUs, prior to taxonomic assignment. After the taxonomic assignment of the OTUs, the correlations were no longer significant. Second, we also observed that the most frequently identified taxa of the molecular inventories were different from those of the morphological inventories (Table 1). Even though our local reference library covers a large majority of the taxa detected in our morphological inventories, in most cases, the taxonomical assignment of the OTUs only reached coarse taxonomic levels (class, order). It was rarely possible to assign OTUs to more precise levels (family, genus or species).

In the following paragraphs, we explain the origin of these differences, which comes from (4.1) technical and (4.2) biological factors. Then, we explain the originality and complementarity of using eDNA in biofilms to assess macroinvertebrates diversity (4.3).

4.1. Technical reasons explain differences between morphological inventories and molecular inventories from biofilms eDNA

4.1.1. Local reference library

Species from tropical zones are known to be badly represented in public reference barcoding libraries. However, even for very well-studied biogeographical regions, like Europe, the coverage of reference libraries is far from complete: for instance for Diptera, Ephemeroptera, or Mollusca, over 40% of species are not yet barcoded (Weigand et al., 2019). Our study area is located in a remote tropical region where macroinvertebrate diversity is still not well-studied. Indeed, several endemic species have been recently described from Mayotte Island (Moubayed-Breil & Mary, 2019; Johanson & Mary, 2009) and all these new species are not barcoded. The macroinvertebrate sequences recovered from BOLD and used to construct our local library did not fit for our tropical freshwater environments. They remained insufficient for purposes of obtaining a precise taxonomical resolution in particular at species level. This lack of endemic tropical taxa in the reference library partly explain why the majority of macroinvertebrate sequences were assigned only to class (Insecta, 34.5%) or order level (Diptera, 26.3%; Hydrozoa, 10%).

On the other hand, the accuracy and reliability of the public database used to construct our reference library may also explain the poor taxonomic assignment. DNA barcode sequences from BOLD may contain some erroneous data (Meiklejohn et al., 2019). If the database contains two DNA sequences belonging to the same taxa but identified using different

taxonomic names, this will cause the failure of the assignment. Thus, these sequences will remain “unclassified” at high taxonomic levels.

4.1.2 Primer bias

The considerable amount of non-macroinvertebrate taxa detected using the biofilm molecular approach (73%) indicates the impact of the primer pair we used (Figure 4). Biofilms host a rich biodiversity covering eukaryotic algae, cyanobacteria, bacteria, fungi, insect larvae, ciliate and flagellate grazers. Diatoms and bacteria are the most abundant taxa (Battin et al., 2016). The dominance of non-macroinvertebrate taxa in biofilms along with the use of the degenerated primers BF2/BR2 can explain why we amplified more untargeted taxa than the macroinvertebrates themselves. The amplification of a large number of untargeted taxa related to the use of BF2/BR2 has already been reported by Zizka et al. (2018). To avoid the amplification of untargeted taxa when one is working with biofilm samples, a solution would be to use more specific primers to increase the amplification accuracy. Recently, Leese et al., (2020) developed a new reverse primer to minimize the amplification of non-target organisms from water samples using primers that amplify short DNA sequences. We could also envisage working on a less ubiquitous gene than COI when working in a such rich diversity matrix as biofilms.

Another problem with BF2/BR2 primers is their difficulty to amplify some Mollusca and Trichoptera taxa (Leese et al., 2020). This may explain why we detected two times more Mollusca in the morphological bulk samples compared to the DNA biofilm samples (Figure 7). We also observed very few trichoptera sequences which may be due to primers bias but also to biological reasons (see below). The use of primers targeting a long barcode (421 bp excluding primers) can also partly explain the high percentage of macroinvertebrate

sequences (70.8%) that could not be assigned beyond the class or order level. Longer barcodes can accumulate more errors during the amplification and sequencing processes (Aylagas et al., 2016; Schirmer et al., 2015). Furthermore, macroinvertebrate eDNA in biofilms should be essentially extra-cellular DNA (DNA released from cell lysis), which is more degraded compared to intracellular DNA (DNA from living cells or living multicellular organisms) (Taberlet et al., 2018). The BF2/BR2 primers were developed to target the COI gene in intracellular DNA extracted from whole organisms. They are probably not the best adapted to amplify degraded extracellular DNA, as we probably had in environmental biofilms. Primers for extracellular DNA must target a short-enough DNA region to be able to amplify more degraded eDNA (Ruppert et al., 2019). Therefore, we would recommend, in the future, the use of primers targeting shorter COI fragments as those developed by Vamos et al. (2017). The preference for shorter COI primers has been also mentioned by Bista et al. (2017), who showed a better taxonomic resolution with shorter (235 bp) amplicons than do larger (658 bp) when working with degraded DNA.

4.2. Biological reasons explain differences between morphological inventories and biofilm-captured molecular inventories

The amount of eDNA that the organisms release into their environment depends mainly on biological factors, e.g., species, size, biomass, age classes within species, biological activity, feeding and reproduction behavior, migration patterns, etc. (Barnes & Turner, 2016). Additionally, some organisms can present morphological features —such as the presence of an exoskeleton— that may reduce the release of DNA in the environment. Indeed, organisms with sclerotized bodies (e.g., arthropods) release less DNA compared to fish and amphibians (Taberlet et al., 2018). We observed the same kind of phenomenon in our study: some insects,

such as several Trichoptera families, are surrounded by a case that may limit DNA release (Zizka et al., 2018). In Mayotte, of the eight Trichoptera families detected using the morphological approach, three (Pisulidae, Leptoridae, and Calamoceritidae) exhibit larvae construction during the whole larval phase while one (Hydroptilidae) does so only at the last larval stage (Tachet et al., 2010). This may partly explain why, using the molecular approach, we were unable to detect these families despite the presence of their sequences in our reference database. Conversely, we detected Trichoptera families which do not built cases (Ecnomidae and Philopotamidae) in our biofilm samples.

The habitat and life forms of the studied organisms are also important factors to take into account when one is working with environmental DNA. Understanding where and how organisms live allows for a better interpretation of the results. As mentioned above, biofilms may contain two different types of macroinvertebrate DNA signals: intracellular eDNA from organisms living in the biofilm (e.g., Diptera larvae) and extracellular eDNA, which corresponds to DNA released from the cells of organisms living in the area (inside the biofilm, in the surrounding river habitats, in upstream habitats, etc.). It is probable that the intracellular signal, with a large quantity of good-quality DNA, may predominate extracellular eDNA, with a lower quantity of degraded DNA. A majority of the *Insecta_unclassified* taxa molecularly detected in our study may correspond mainly to intracellular eDNA from very small larvae of Diptera such as chironomids living in the biofilms rather than to extracellular eDNA from individuals living in the area where the biofilms were collected. However, a better taxonomical resolution, at least to the family level, must be reached to elucidate the identity of these numerous *Insecta_unclassified* and *Diptera_unclassified* sequences.

The preference of particular taxa for some specific habitats and the drift of free DNA from upstream can explain the differences observed between the morphological and the biofilm molecular approach. Indeed, biofilms captured eDNA of the *Odonata*, *Lepidoptera*, and *Hemiptera* orders in low abundances (< 1%). The detection of these taxa, which are known to be absent from stones' biofilms and to live in other habitats, suggests the capacity of biofilms to entrap extracellular DNA.

4.3. Interest and originality of capturing eDNA from biofilms to study macroinvertebrate communities

For ecological studies, the ability of molecular methods to detect very small organisms (sizes smaller than 1 mm like very small *Chironomus* or insect eggs) living inside biofilms can be a bonus compared to the conventional morphological approaches. These small organisms are either impossible to identify or are not directly considered because they are not retained in the sampling nets due to their small size. It is also important to consider that females of many amphibiotic taxa (e.g., many Ephemeroptera) lay their eggs on the water's surface. The eggs are then passively spread by the water flow and settle in many areas of the river that are not systematically representative of favorable habitats for larvae or nymphs (often leading to high mortality at embryonic stage). As a result, biofilms' eDNA can detect insects' eggs.

Cnidaria were much more detected in the biofilm eDNA compared to the conventional morphological approach (Figure 7). This phylum was represented by the genera *Hydra* and *Craspedacusta*. Specimens from the *Hydra* genus were detected in 15 sampling sites using the molecular approach (representing 23.5% of the total detected taxa) while only one individual was observed in one site using the morphological approach. The *Hydra* genus encompasses mainly freshwater polyps living attached to stones and their transparent body do not exceed 1-

2 mm. They are hardly observable in binocular. *Craspedacusta* genus (representing 7.3% of the total detected taxa) is frequently observed in freshwater at the jellyfish stage (planktonic) rather than at the polyp stage (benthic) (Tachet et al., 2010). Hydrozoa reproduction may be asexual or sexual, depending on the environmental conditions. Asexual reproduction is carried out by budding, while during sexual reproduction, free-swimming gametes are released into the environment. The release of genetic material into the freshwater environment during reproduction may also explain the high abundance of eDNA of this taxon in biofilms compared to bulk samples. The sensibility of the biofilm molecular approach to detect this kind of taxa compared to the morphological approach (322 OTUs for Hydrozoa class vs 1 observation) is a significant advantage for diversity studies.

5. Conclusion and perspectives for the future

We showed that macroinvertebrate eDNA is present in biofilms and that biofilm molecular OTU inventories provided structures and diversities that are comparable to those based on morphological inventories. We also showed that the ecological assessment, based on the EPT index, could also be assessed through the biofilm molecular approach. However, further studies are now necessary prior considering biofilms as good surrogates for bulk samples for river bioassessment. Furthermore, some optimizations will be necessary. Among these optimizations, one should be the use of more specific primers targeting shorter barcodes to sequence more easily degraded eDNA that may be captured in the natural biofilms. Another optimization should be the development of a complete reference barcode library adapted to the selected primers and to the study area. These optimizations will allow a more precise detection of macroinvertebrate taxa (at least to the family level), especially EPTs, which are key indicators of water quality assessment (Haidekker & Hering, 2008; Lorenz et al., 2004). On the other hand, the relationship between sequence abundances and biomass or specimens'

abundances has always been a critical problem for metabarcoding studies (e.g. Elbrecht et al., 2017). Recent studies (Beentjes et al., 2018; Buchner et al.,2019; Fernandez et al., 2019) showed that, for monitoring studies, biotic indices based on presence/absence data could be applied with good results for water quality assessment. The use of presence/absence data should bypass the need of abundance data and opens the path for the application of eDNA metabarcoding for monitoring studies.

This exploratory study was dedicated to test the eDNA captation potential of biofilms for macroinvertebrate river communities. Therefore, we did not test the link between this eDNA signal in biofilms with environmental conditions. Yet, taxonomy-free approaches such as the one developed by Tapolczai et al. (2019b) on diatom communities, which is connecting directly environmental data to eDNA sequence data, may be a way to overcome some limitations (e.g. reference database) and to predict ecological status directly from the eDNA pattern of macroinvertebrate communities assessed from biofilms.

Finally, experiments under controlled conditions, including replicates and comparing the molecular and the morphological bulk sample approaches to the biofilm molecular approach are needed to better evaluate biofilms' capacity to catch extracellular DNA. Biofilm architecture should also be taken into account in further studies as it may modulate the retention capacity of biofilms and thereby the captured eDNA signal. Even if many questions still have to be addressed, the use of biofilms to capture eDNA of environmental communities is promising and may not be limited to macroinvertebrates.

Acknowledgements

We thank the OFB (Office Français de la Biodiversité) for its financial support. We also thank Vasco Elbrecht for his constructive advice during DNA extractions and bioinformatics

treatment. We thank Sonia Lacroix for DNA extraction, PCR, and library preparation. We thank Julie Guéguen and Teofana Chonova for their assistance with statistical analyses in R and Philippe Usseglio-Polatera for his careful consideration of this manuscript and his meaningful comments. Finally, we thank COST Action DNAqua-Net (CA15219), which is supported by the COST (European Cooperation in Science and Technology) program for support and fruitful scientific discussions.

References

Afnor. (2014). EN 13946. Water quality—Guidance standard for the routine sampling and pretreatment of benthic diatoms from rivers. 1-17.

Afnor. (2016). NF T90-333. Qualité de l'eau—Prélèvement des macro-invertébrés aquatiques en rivières peu profondes.

Aylagas, E., Borja, Á., Irigoien, X., & Rodríguez-Ezpeleta, N. (2016). Benchmarking DNA metabarcoding for biodiversity-based monitoring and assessment. Frontiers in Marine Science, 3, 96.

Bailey, R. C., Norris, R. H., & Reynoldson, T. B. (2001). Taxonomic resolution of benthic macroinvertebrate communities in bioassessments. *Journal of the North American Benthological Society*, *20*(2), 280–286.

Barnes, M. A., & Turner, C. R. (2016). The ecology of environmental DNA and implications for conservation genetics. Conservation genetics, 17(1), 1–17.

Battin, T. J., Besemer, K., Bengtsson, M. M., Romani, A. M., & Packmann, A. I. (2016). The ecology and biogeochemistry of stream biofilms. Nature Reviews Microbiology, 14(4), 251.

Beentjes, K. K., Speksnijder, A. G., Schilthuizen, M., Schaub, B. E., & van der Hoorn, B. B. (2018). The influence of macroinvertebrate abundance on the assessment of freshwater quality in The Netherlands. *Metabarcoding and Metagenomics*, *2*, e26744.

Beermann, A. J., Zizka, V. M., Elbrecht, V., Baranov, V., & Leese, F. (2018). DNA metabarcoding reveals the complex and hidden responses of chironomids to multiple stressors. Environmental Sciences Europe, 30(1), 26.

Buchner, D., Beermann, A. J., Laini, A., Rolauffs, P., Vitecek, S., Hering, D., & Leese, F. (2019). Analysis of 13,312 benthic invertebrate samples from German streams reveals minor deviations in ecological status class between abundance and presence/absence data. *PloS one*, 14(12), e0226547.

Bista, I., Carvalho, G. R., Walsh, K., Seymour, M., Hajibabaei, M., Lallias, D., Christmas, M., & Creer, S. (2017). Annual time-series analysis of aqueous eDNA reveals ecologically relevant dynamics of lake ecosystem biodiversity. Nature communications, 8, 14087.

Carew, M. E., Pettigrove, V. J., Metzeling, L., & Hoffmann, A. A. (2013). Environmental monitoring using next generation sequencing: Rapid identification of macroinvertebrate bioindicator species. Frontiers in zoology, 10(1), 45.

CEN. (2018). Water quality—CEN/TR 17245—Technical report for the routine sampling of benthic diatoms from rivers and lakes adapted for metabarcoding analyses. European Strandardisation Committee. 1-8.

Chaumet, B., Morin, S., Hourtané, O., Artigas, J., Delest, B., Eon, M., & Mazzella, N. (2019). Flow conditions influence diuron toxicokinetics and toxicodynamics in freshwater biofilms. Science of The Total Environment, 652, 1242–1251.

Day, J. A., Harrison, A. D., & De Moor, I. J. (2002). Guides to the freshwater invertebrates of southern Africa. Volume 9: Diptera. WRC Report No. TT 201/02. Water Research Commission, Pretoria, South Africa.

De Moor, I. J., Day, J. A., & De Moor, F. C. (2003a). Guides to the freshwater invertebrates of southern Africa. Volume 7: Insecta I: Ephemeroptera, Odonata and Plecoptera. WRC Report No. TT 207/03. Water Research Commission, Pretoria, South Africa.

De Moor, I. J., Day, J. A., & De Moor, F. C. (2003b). Guides to the freshwater invertebrates of southern Africa. Volume 8: Insecta II: Hemiptera, Megaloptera, Neuroptera, Trichoptera and Lepidoptera. WRC Report No. TT 214/03. Water Research Commission, Pretoria, South Africa.

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*(7), e0130324.

Elbrecht, V., & Leese, F. (2017). Validation and development of COI metabarcoding primers for freshwater macroinvertebrate bioassessment. Frontiers in Environmental Science, 5, 11.

Elbrecht, V., Vamos, E. E., Meissner, K., Aroviita, J., & Leese, F. (2017). Assessing strengths and weaknesses of DNA metabarcoding-based macroinvertebrate identification for routine stream monitoring. Methods in Ecology and Evolution, 8(10), 1265–1275.

Flemming, H.-C., & Wingender, J. (2010). The biofilm matrix. Nature reviews microbiology, 8(9), 623.

Froehner, S., Machado, K. S., Dombroski, L. F., Nunes, A. C., Kishi, R. T., Bleninger, T., & Sanez, J. (2012). Natural biofilms in freshwater ecosystem : Indicators of the presence

of polycyclic aromatic hydrocarbons. Water, Air, & Soil Pollution, 223(7), 3965–3973.

Haidekker, A., & Hering, D. (2008). Relationship between benthic insects (Ephemeroptera, Plecoptera, Coleoptera, Trichoptera) and temperature in small and medium-sized streams in Germany : A multivariate study. *Aquatic Ecology*, *42*(3), 463–481.

Hajibabaei, M., Porter, T., Robinson, C., Baird, D., Shokralla, S., & Wright, M. (2019). Watered-down biodiversity? A comparison of metabarcoding results from DNA extracted from matched water and bulk tissue biomonitoring samples. BioRxiv, 575928.

Hajibabaei, M., Shokralla, S., Zhou, X., Singer, G. A., & Baird, D. J. (2011). Environmental barcoding : A next-generation sequencing approach for biomonitoring applications using river benthos. PLoS one, 6(4), e17497.

Hajibabaei, M., Spall, J. L., 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(1), 28.

Hammer, Ø., Harper, D. A., & Ryan, P. D. (2001). PAST : Paleontological statistics software package for education and data analysis. *Palaeontologia electronica*, *4*(1), 9.

Hebert, P. D., Cywinska, A., Ball, S. L., & Dewaard, J. R. (2003). Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *270*(1512), 313–321.

Johanson, K. A., & Mary, N. (2009). Description of three new caddisfly species from Mayotte Island, Comoros Archipelago (Insecta : Trichoptera). Zootaxa, 2089(1), 1–9.

Keith, P., Marquet, G., Valade, P., Bosc, P., & Vigneux, E. (2006). Atlas des poissons et des crustacés d'eau douce des Comores, Mascareignes et Seychelles. Collection Patrimoines naturels.

Leese, F., Sander, M., Buchner, D., Elbrecht, V., Haase, P., & Zizka, V. M. (2020). Improved freshwater macroinvertebrate detection from eDNA through minimized non-target amplification. *bioRxiv*.

Legendre, P., & Fortin, M.-J. (2010). Comparison of the Mantel test and alternative approaches for detecting complex multivariate relationships in the spatial analysis of genetic data. *Molecular ecology resources*, *10*(5), 831–844.

Lenat, D. R. (1987). Water quality assessment using a new qualitative collection method for freshwater benthic macroinvertebrates. *North Carolina DEM Tech. Report*, *12*.

Lorenz, A., Hering, D., Feld, C. K., & Rolauffs, P. (2004). A new method for assessing the impact of hydromorphological degradation on the macroinvertebrate fauna of five German stream types. *Hydrobiologia*, *516*(1-3), 107–127.

Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer research*, *27*(2 Part 1), 209–220.

Mariani, S., Baillie, C., Colosimo, G., & Riesgo, A. (2019). Sponges as natural environmental DNA samplers. Current Biology, 29, R395-R402.

Mary, N. (2017). Les macroinvertébrés benthiques des cours d'eau de la Nouvelle-Calédonie. Guide d'identification. Version révisée 2017, DAVAR Nouvelle-Calédonie. Ed. OEIL, CNRT.

Mary, N. (2018). Réseau de contrôle de surveillance des cours d'eau de Mayotte, Suivi 2017. Volet Macroinvertebres benthiques. ETHYC'O.BRGM. 57 pages + annexes.

Meiklejohn, K. A., Damaso, N., & Robertson, J. M. (2019). Assessment of BOLD and GenBank–Their accuracy and reliability for the identification of biological materials. PloS one, 14(6).

Mora-Gómez, J., Freixa, A., Perujo, N., & Barral-Fraga, L. (2016). Limits of the biofilm concept and types of aquatic biofilms. Aquatic biofilms: Ecology, water quality and wastewater treatment, 3–28.

Morin, S., Gómez, N., Tornés, E., Licursi, M., & Rosebery, J. (2016). Benthic diatom monitoring and assessment of freshwater environments : Standard methods and future challenges. In Aquatic Biofilms : Ecology, Water Quality and Water Treatment (p. 111–124). Caister Academic Press UK.

Mortágua, A., Vasselon, V., Oliveira, R., Elias, C., Chardon, C., Bouchez, A., Rimet, F., Feio, M. J., & Almeida, S. F. (2019). Applicability of DNA metabarcoding approach in the bioassessment of Portuguese rivers using diatoms. Ecological Indicators, 106, 105470.

Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., O'hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., & Wagner, H. (2010). Vegan : Community ecology package. R package version 1.17-4. http://cran. r-project. org>. Acesso em, 23, 2010.

Oliveros, J. C. (2016). Venny. An interactive tool for comparing lists with Venn's diagrams. 2007–2015.

Pedersen, M. L., & Friberg, N. (2009). Influence of disturbance on habitats and biological communities in lowland streams. Fundamental and Applied Limnology/Archiv für Hydrobiologie, 174(1), 27–41.

Pereira-da-Conceicoa, L. P., Elbrecht, V., Hall, A. C., Briscoe, A., Barber-James, H., & Price, B. (2020). Metabarcoding unsorted kick-samples facilitates macroinvertebrate-based

biomonitoring with increased taxonomic resolution, while outperforming environmental DNA. Environmental DNA, 00, 1-19.

Peres-Neto, P. R., & Jackson, D. A. (2001). How well do multivariate data sets match? The advantages of a Procrustean superimposition approach over the Mantel test. Oecologia, 129(2), 169–178.

Pešić, V. M., Smit, H., & Mary, N. (2018). Fifth contribution to the knowledge of water mites (Acari : Hydrachnidia) from the Comoros: a checklist and description of one new genus and four new species. Zootaxa, 4483(2), 331–348.

Pešić, V., Smit, H., & Mary, N. (2015). Third contribution to the knowledge of water mites from the Comoros, with the description of two new species (Acari : Hydrachnidia). Zootaxa, 3964(4), 445–459.

R Core Team. (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

Rimet, F. (2012). Recent views on river pollution and diatoms. Hydrobiologia, 683(1), 1–24.

Rivera, S. F., Vasselon, V., Jacquet, S., Bouchez, A., Ariztegui, D., & Rimet, F. (2018). Metabarcoding of lake benthic diatoms : From structure assemblages to ecological assessment. Hydrobiologia, 807(1), 37–51.

Rivera, Sinziana F., Vasselon, V., Bouchez, A., & Rimet, F. (2020). Diatom metabarcoding applied to large scale monitoring networks : Optimization of bioinformatics strategies using Mothur software. Ecological Indicators, 109, 105775.

Romaní, A. M., Guasch, H., & Balaguer, M. D. (2016). Aquatic biofilms. Ecology, Water Quality and Wastewater Treatment Caister. Norfolk, UK: Academic Press.

Ruppert, K. M., Kline, R. J., & Rahman, M. S. (2019). Past, present, and future perspectives of environmental DNA (eDNA) metabarcoding: A systematic review in methods,

monitoring, and applications of global eDNA. Global Ecology and Conservation, e00547.

Sayers, E. W., Agarwala, R., Bolton, E. E., Brister, J. R., Canese, K., Clark, K., Connor, R., Fiorini, N., Funk, K., & Hefferon, T. (2019). Database resources of the national center for biotechnology information. *Nucleic acids research*, *47*(Database issue), D23.

Serrana, J. M., Miyake, Y., Gamboa, M., & Watanabe, K. (2019). Comparison of DNA metabarcoding and morphological identification for stream macroinvertebrate biodiversity assessment and monitoring. *Ecological Indicators*, *101*, 963–972.

Schirmer, M., Ijaz, U. Z., D'Amore, R., Hall, N., Sloan, W. T., & Quince, C. (2015). Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform. Nucleic acids research, 43(6), e37–e37.

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., Lesniewski, R. A., Oakley, B. B., Parks, D. H., & Robinson, C. J. (2009). Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl. Environ. Microbiol., 75(23), 7537–7541.

Schneider, J., Valentini, A., Dejean, T., Montarsi, F., Taberlet, P., Glaizot, O., & Fumagalli, L. (2016). Detection of invasive mosquito vectors using environmental DNA (eDNA) from water samples. PloS one, 11(9), e0162493.

Schnell, I. B., Bohmann, K., & Gilbert, M. T. P. (2015). Tag jumps illuminated – Reducing sequence-to-sample misidentifications in metabarcoding studies. Molecular Ecology Resources, 15(6), 1289–1303.

Shannon, C. E., & Weaver, W. (1949). *The Mathematical Theory of Communication, by CE Shannon (and Recent Contributions to the Mathematical Theory of Communication), W. Weaver*. University of illinois Press.

Shogren, A. J., Tank, J. L., Andruszkiewicz, E. A., Olds, B., Jerde, C., & Bolster, D. (2016). Modelling the transport of environmental DNA through a porous substrate using continuous flow-through column experiments. Journal of the Royal Society Interface, 13(119), 20160290.

Shogren, A. J., Tank, J. L., Egan, S. P., August, O., Rosi, E. J., Hanrahan, B. R., Renshaw, M. A., Gantz, C. A., & Bolster, D. (2018). Water Flow and Biofilm Cover Influence Environmental DNA Detection in Recirculating Streams. Environmental science & technology, 52(15), 8530–8537.

Smit, H., Pešić, V., & Mary-Sasal, N. (2010). Second contribution to the knowledge of water mites from the Comoros, with the description of one new species (Acari: Hydrachnidia). Zootaxa, 2413(1), 51–60.

Stoyanova, T., Vidinova, Y., Yaneva, I., Tyufekchieva, V., Parvanov, D., Traykov, I., & Bogoev, V. (2014). Ephemeroptera, Plecoptera and Trichoptera as indicators for ecological quality of the Luda Reka River, Southwest Bulgaria. *Acta zoologica bulgarica*, *66*(2), 255–260.

Sweeney, B. W., Battle, J. M., Jackson, J. K., & Dapkey, T. (2011). Can DNA barcodes of stream macroinvertebrates improve descriptions of community structure and water quality? Journal of the North American Benthological Society, 30(1), 195–216.

Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C., & Willerslev, E. (2012). Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular ecology*, *21*(8), 2045–2050.

Taberlet, P., Bonin, A., Zinger, L., & Coissac, E. (2018). Environmental DNA: For biodiversity research and monitoring. Oxford University Press.

Tachet, H., Richoux, P., Bournaud, M., & Usseglio-Polatera, P. (2010). Invertébrés d'eau douce : Systématique, biologie, écologie (Vol. 15). CNRS éditions Paris.

Tapolczai, K., Bouchez, A., Stenger-Kovács, C., Padisák, J., & Rimet, F. (2017). Taxonomy- or trait-based ecological assessment for tropical rivers? Case study on benthic diatoms in Mayotte island (France, Indian Ocean). Science of the Total Environment, 607, 1293–1303.

Tapolczai, K., Vasselon, V., Bouchez, A., Stenger-Kovács, C., Padisák, J., & Rimet, F. (2019a). The impact of OTU sequence similarity threshold on diatom-based bioassessment: A case study of the rivers of Mayotte (France, Indian Ocean). Ecology and evolution, 9(1), 166–179.

Tapolczai, K., Keck, F., Bouchez, A., Rimet, F., & Vasselon, V. (2019b). Diatom DNA metabarcoding for biomonitoring : Strategies to avoid major taxonomical and bioinformatical biases limiting molecular indices capacities. Frontiers in Ecology and Evolution, 7, 409.

Thomsen, P. F., & Willerslev, E. (2015). Environmental DNA–An emerging tool in conservation for monitoring past and present biodiversity. Biological conservation, 183, 4–18.

Valade, P., Mary, N., Abdou, A., & Keith, P. (2007). Inventaire des espèces de poissons et d'invertébrés des eaux douces de Mayotte. Rapport final ARDA, MNHN, DAF Mayotte.

Vamos, E. E., Elbrecht, V., & Leese, F. (2017). Short COI markers for freshwater macroinvertebrate metabarcoding. PeerJ Preprints.

Vasselon, V., Rimet, F., Tapolczai, K., & Bouchez, A. (2017). Assessing ecological status with diatoms DNA metabarcoding : Scaling-up on a WFD monitoring network (Mayotte island, France). Ecological Indicators, 82, 1–12.

Wang, Q., Garrity, G. M., Tiedje, J. M., & Cole, J. R. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl. Environ. Microbiol., 73(16), 5261–5267.

Weigand, H., Beermann, A. J., Čiampor, F., Costa, F. O., Csabai, Z., Duarte, S., Geiger, M. F., Grabowski, M., Rimet, F., & Rulik, B. (2019). DNA barcode reference libraries for the monitoring of aquatic biota in Europe : Gap-analysis and recommendations for future work. BioRxiv, 576553.

Wolfaardt, G. M., Lawrence, J. R., Robarts, R. D., & Caldwell, D. E. (1998). In situ characterization of biofilm exopolymers involved in the accumulation of chlorinated organics. Microbial ecology, 35(3-4), 213–223.

Zizka, V. M., Leese, F., Peinert, B., & Geiger, M. F. (2018). DNA metabarcoding from sample fixative as a quick and voucher-preserving biodiversity assessment method. Genome, 62(3), 122–136.

Data accessibility

The Illumina Miseq raw data can be accessed at doi: http://doi.org/ 10.5281/zenodo.3934412

The COI macroinvertebrate reference library is available at doi: https://doi.org/ 10.15454/UNUFWY

Authors Contributions

FR, AB and OM conceived the study. VV conceived laboratory and bioinformatics treatment. SFR performed bioinformatics, data analysis and wrote the paper with significant contributions from all authors. NM performed the macroinvertebrate sampling and morphological identifications of macroinvertebrate specimens. All authors gave final approval for publication

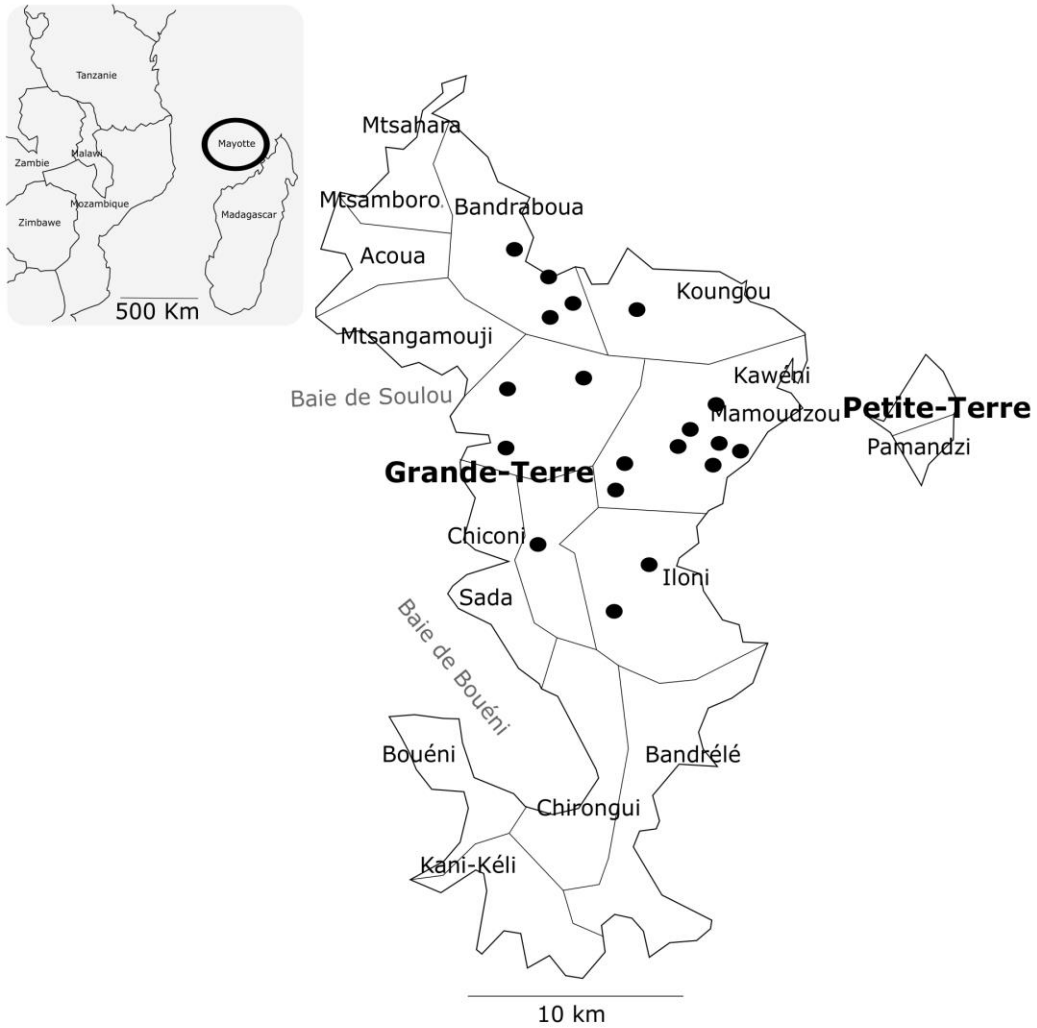


Fig. 1 Location of Mayotte Island and of the river sampling sites.

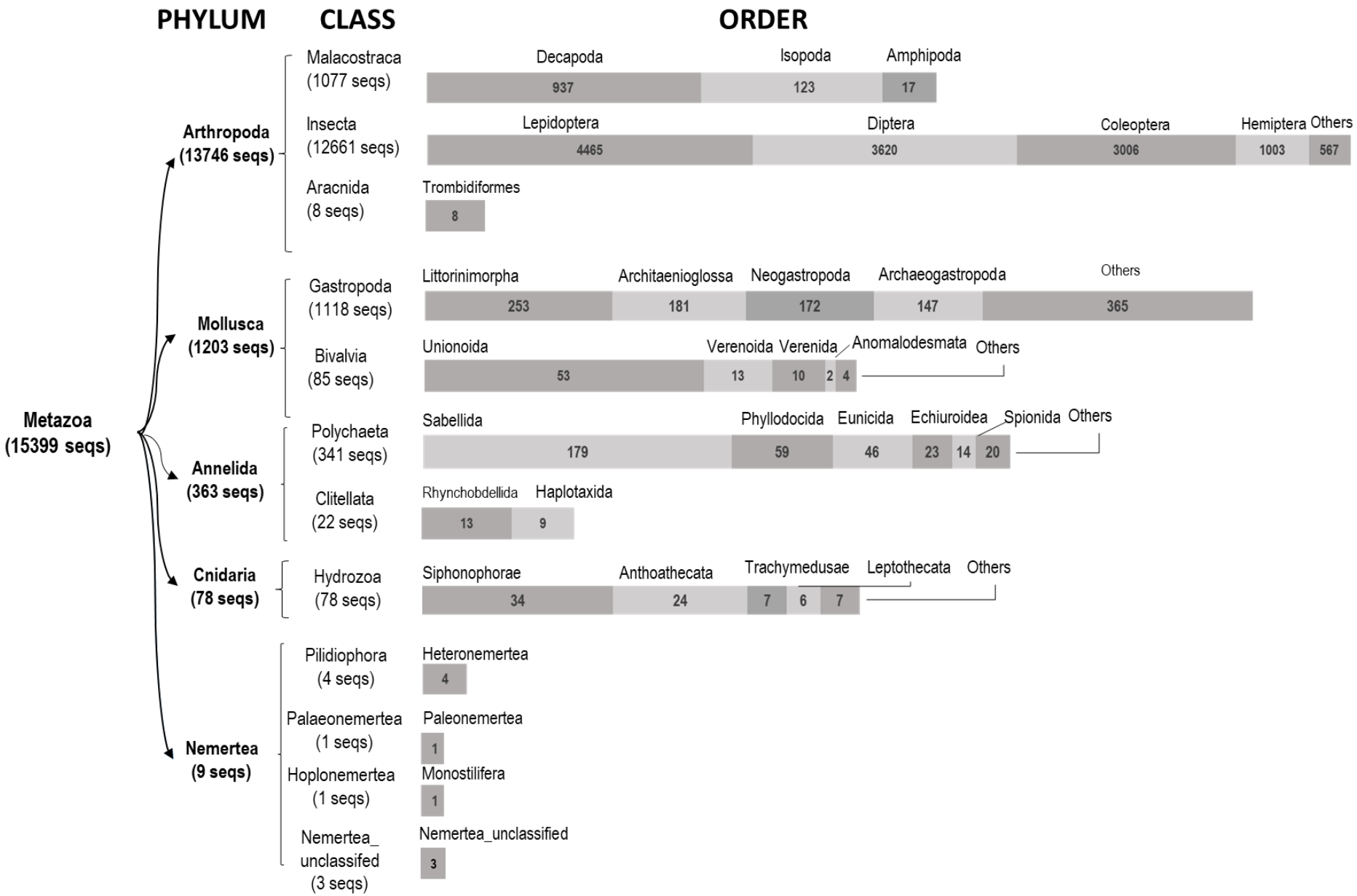


Fig. 2 Overview of the reference library at the phylum, class, and order levels. Only dominant orders are represented. Bar length and numbers give the number of sequences per order.

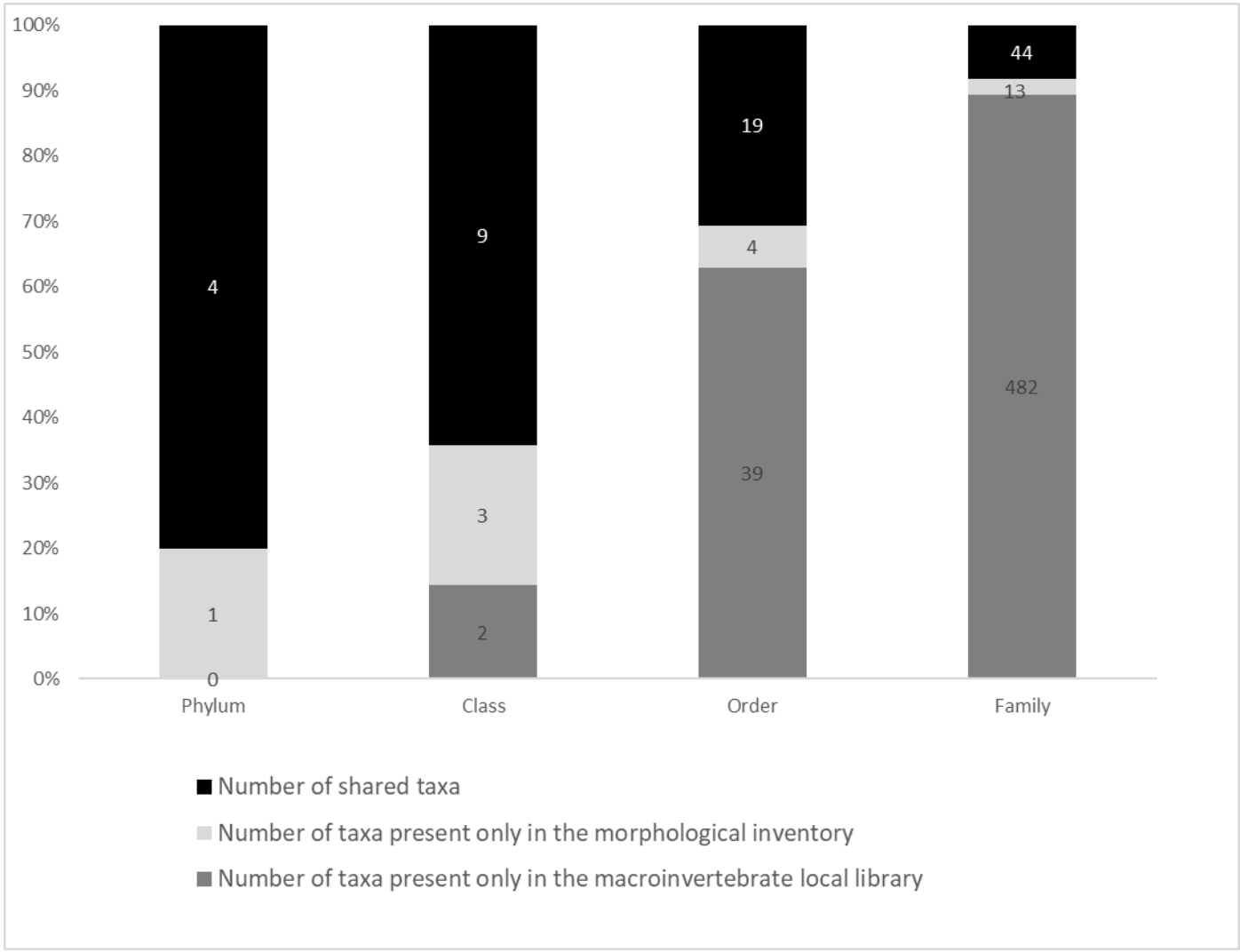


Fig. 3 Comparison between the morphological inventories and our local macroinvertebrate reference library at the phylum, class, order, and family levels.

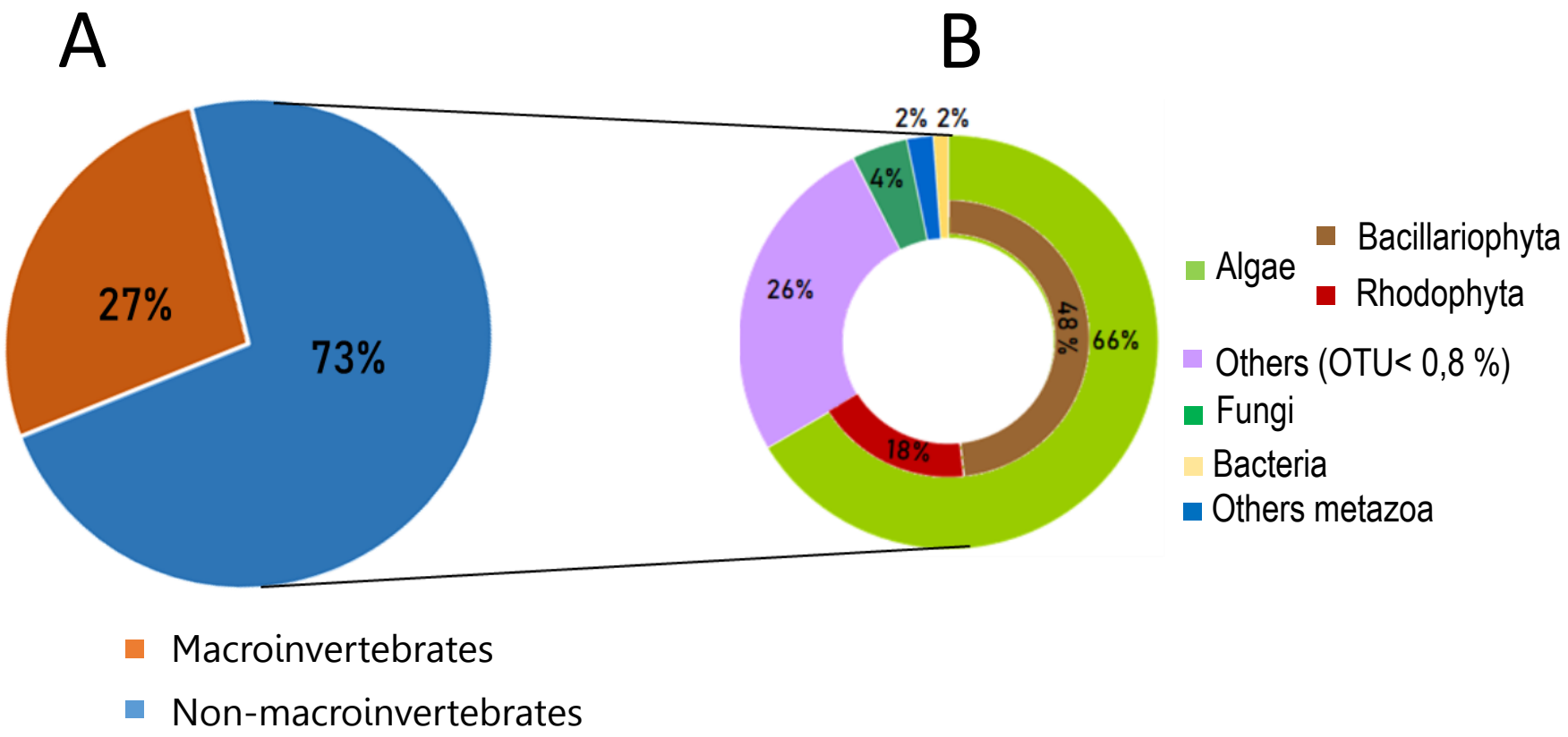


Fig. 4 A. Percentage of macroinvertebrate and non-macroinvertebrate sequences detected in the biofilms **B.** Identity of non-macroinvertebrate OTUs with proportions higher than 1% when blasted to NCBI.

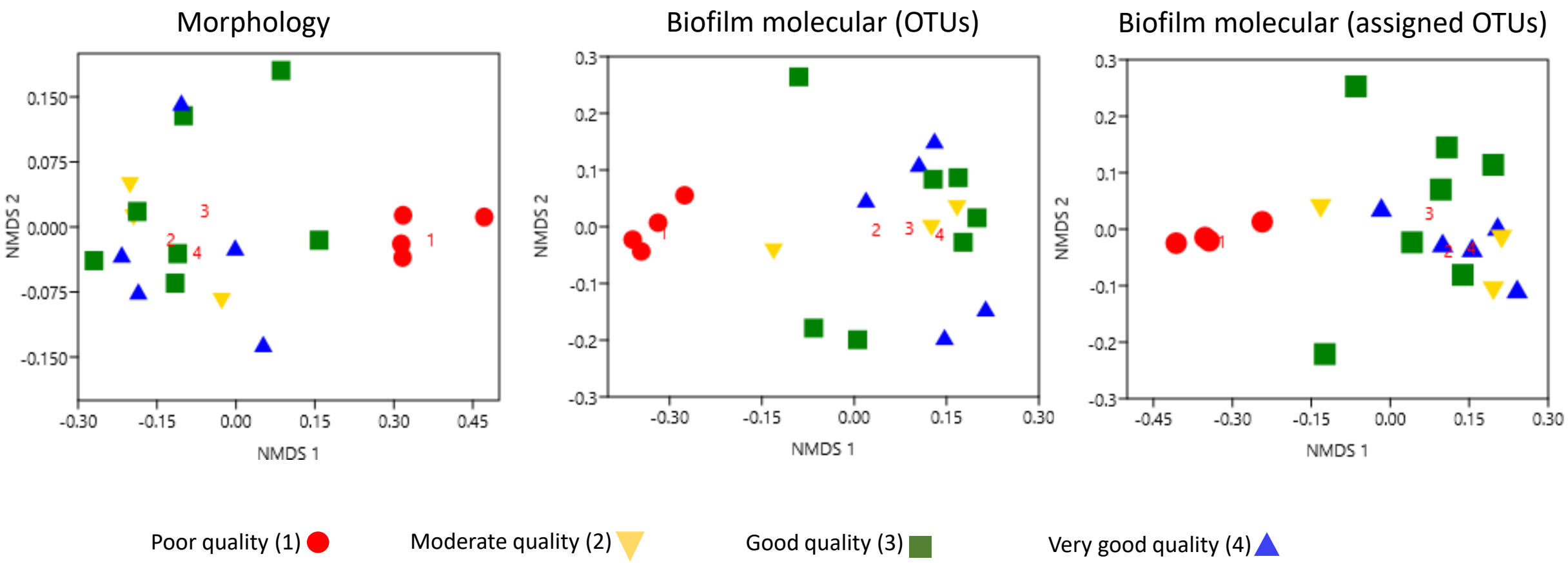


Fig. 5 nMDS ordination plots comparing morphological and biofilm molecular quality assessment. Numbers represent the centroids of each of the 4 water quality classes.

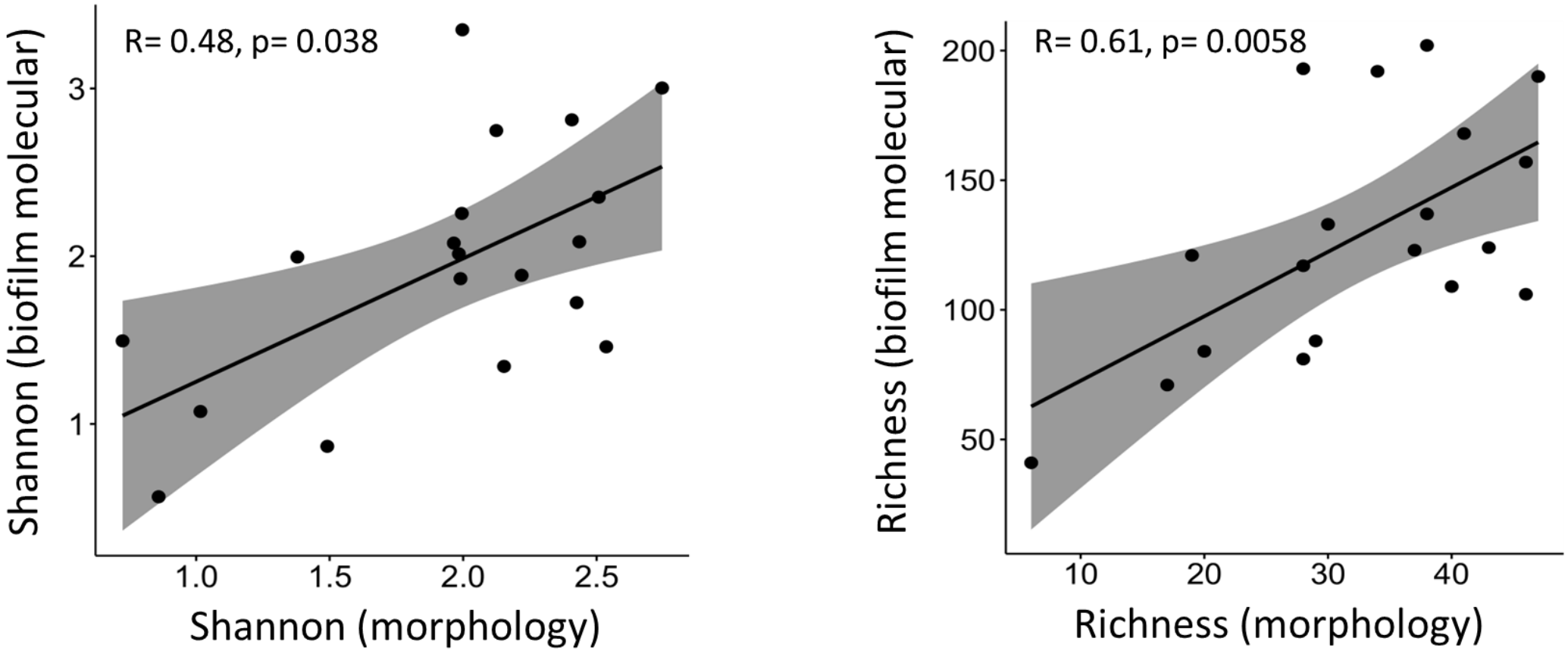


Fig. 6 Correlation between the diversity (Shannon index) and taxa richness obtained with the morphological and the biofilm molecular data. Molecular data corresponds to taxonomically unassigned OTUs.

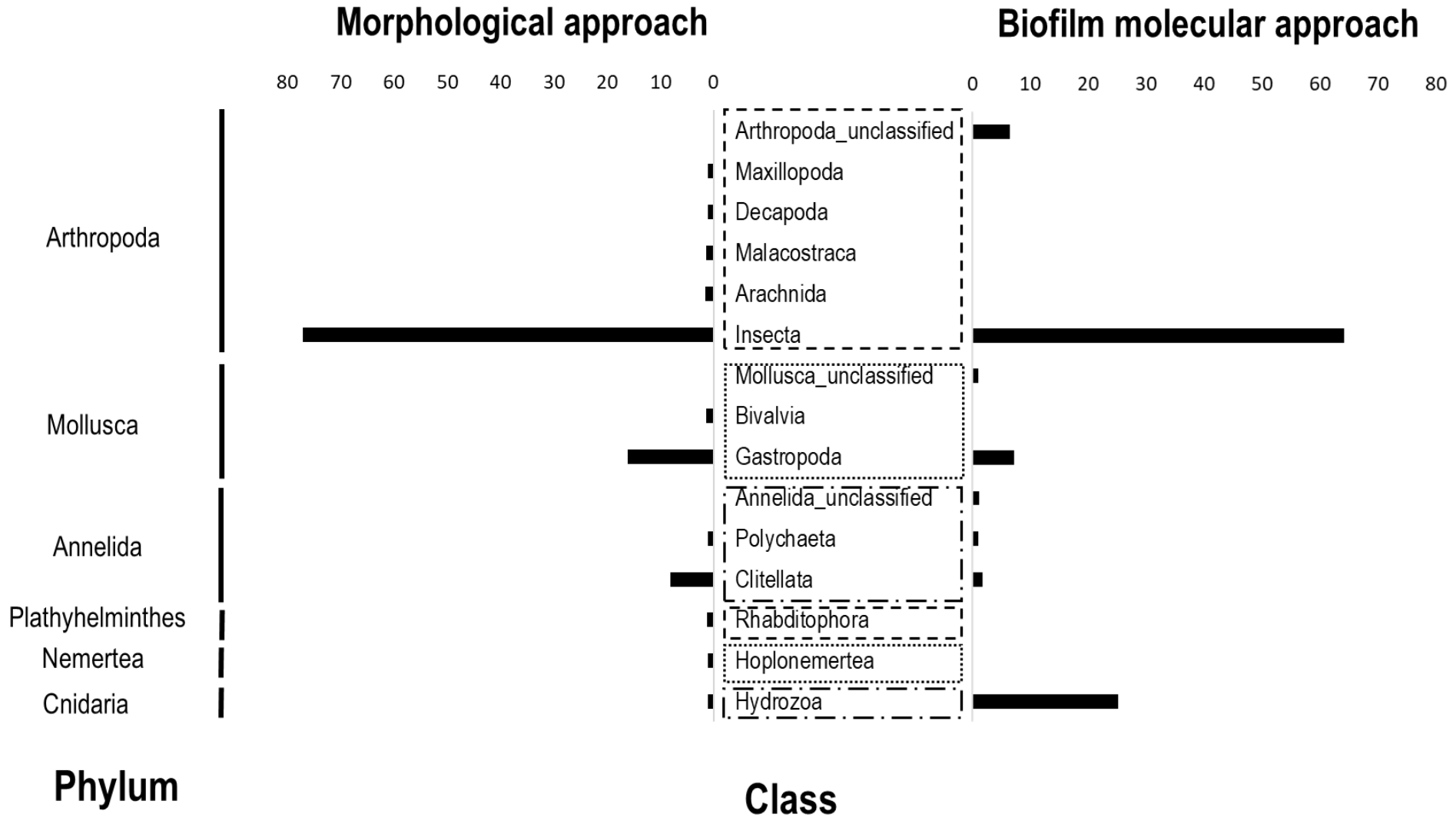


Fig. 7 Relative abundances (%) of the phyla and the class taxa detected using the morphological approach (left) and the biofilm molecular approach (right).

Tables

Table 1 Comparison of the five dominant taxa detected with the morphological (left) and the biofilm molecular (right) approaches. For the biofilm molecular approach, the number of OTUs for each taxon is presented in brackets

Morphological approach		**Biofilm molecular approach**	
Taxon	Abundance (%)	Taxon (metabarcoding)	Abundance (%)
Chironomus spp.	28.1	*Insecta unclassified*	34.5 [154]
Simulium spp.	11.6	*Diptera unclassified*	26.3 [17]
Tanytarsini spp.	12.2	*Hydrozoa unclassified*	10.1 [308]
Naididae_unclassified	6.1	*Craspedacusta sowerbii*	7.3 [4]
Thiara scabra	6.1	*Hydra sinensis*	6.2 [6]

Table 2 Comparison of the morphological (left) and the biofilm molecular (right) inventories at family level in terms of presence/absence. Shaded taxa corresponds to families from the insect class.

ORDER	FAMILY	MORPHOLOGICAL APROACH	BIOFILM APPROACH
Amphipoda	undetermined	X	
Anthoathecata	Hydridae	X	X
Basommatophora	Planorbidae	X	
Coleoptera	Curculionidae	X	X
	Dytiscidae	X	
	Elmidae	X	
	Gyrinidae	X	
	Helodidae	X	
	Hydraenidae	X	
	Hydrophilidae	X	
	Staphylinidae	X	
Collembola	undetermined	X	
Cycloneritimorpha	Neritidae	X	X
Decapoda	Atyidae	X	
	undetermined	X	
Diptera	Cecidomyiidae	X	
	Ceratopogonidae	X	
	Chironomidae	X	X
	Culicidae	X	
	Dixidae	X	
	Empididae	X	
	Ephydridae	X	
	Limoniidae	X	

	Psychodidae	X	
	Simuliidae	X	
	Stratiomyidae	X	
	Tipulidae	X	
Ephemeroptera	Baetidae	X	
	Caenidae	X	
	Leptophlebiidae	X	
Haplotaxida	Naididae	X	X
Hemiptera	Anthocoridae		X
	Gerridae	X	
	Hebridae	X	
	Hydrometridae	X	
	Mesovellidae	X	
	Naucoridae	X	
	Vellidae	X	
Hygrophila	Lymnaeidae	X	
	Physidae	X	
Lepidoptera	Crambidae	X	
	unclassified		X
Limnomedusae	Olindiidae		X
Monostilifera	Tetrastemmatidae	X	
Neritoida	Neritiliidae	X	X
Odonata	Coenagrionidae	X	
	Gomphidae	X	
	Libellulidae	X	
	unclasiffied		X

Oribatida	Oribatidae	X	
Phyllodocida	Nereididae	X	
Rhynchobdellida	Glossiphoniidae	X	
Siphonophorae	*unclassified*		X
Sorbeoconcha	Thiaridae	X	
Trichoptera	Calamoceratidae	X	
	Ecnomidae	X	X
	Hydroptilidae	X	
	Leptoceridae	X	
	Philopotamidae	X	X
	Pisuliidae	X	
	Polycentropodidae	X	
	Psychomyidae	X	
Tricladida	Dugesiidae	X	
Polychaeta_insertae sedis	Aeolosomatidae		X
Veneroida	Sphariidae	X	

Sampling sites

Macroinvertebrate samples

Morphological based-identifications

Species	%
Tiupla sp.	30
T. amarula	8
Oplodonta sp.	3

Morphological inventories

Comparison

Biofilm samples

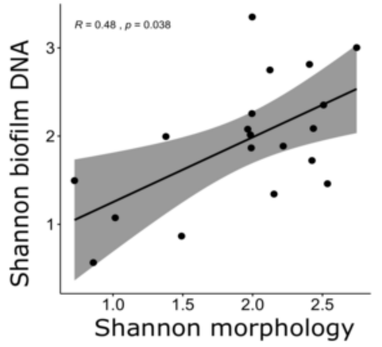
Macroinvertebrates DNA amplification and sequencing

OTUs	%
Otu001	8
Otu002	7
Otu003	5

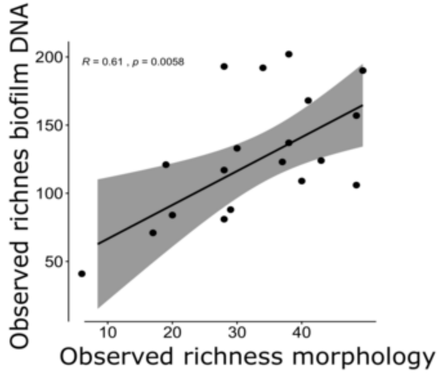
Species	%
Tiupla sp.	15
T. amarula	8
Simulium sp.	3

Molecular inventories (OTU and taxa lists)

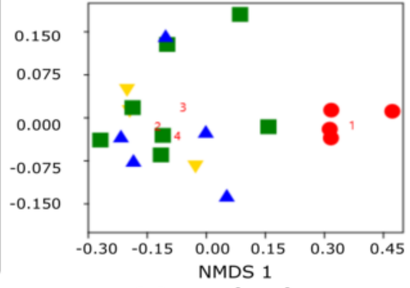
Diversity



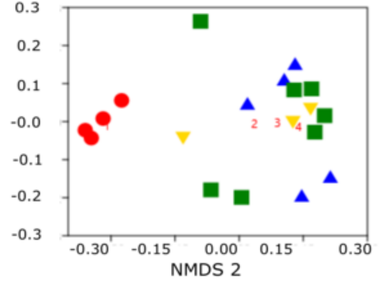
Richness



Ecological assessment



Morphology



Molecular