Morphological vs. DNA metabarcoding approaches for the evaluation of stream ecological status with benthic invertebrates: testing different combinations of markers and strategies of data filtering

Albin MEYER (1), Frédéric BOYER (2), Alice VALENTINI (3), Aurélie BONIN (2,3,4), G. Francesco FICETOLA (2,4), Jean-Nicolas BEISEL (5), Jonathan BOUQUEREL (1), Philippe WAGNER (1), Coline GABORIAUD (3), Florian LEESE (6), Tony DEJEAN (3), Pierre TABERLET (2,7), Philippe USSEGLIO-POLATERA* (1)

* Corresponding author's email address: philippe.usseglio-polatera@univ-lorraine.fr

(1) Université de Lorraine, CNRS, LIEC, F-57000 Metz, France

(2) Univ. Grenoble Alpes, CNRS, LECA, Laboratoire d'Ecologie Alpine, F-38000 Grenoble, France

(3) SPYGEN, F-73370 Le Bourget du Lac, France

(4) Department of Environmental Science and Policy, Università degli Studi di Milano, Via Celoria 26, Milano 20133, Italy

(5) Laboratoire Image Ville et Environnement (LIVE, UMR 7362), Strasbourg, France

(6) University of Duisburg-Essen, Aquatic Ecosystem Research, Universitaetsstr. 5, 45141 Essen

(7) UiT – The Arctic University of Norway, Tromsø Museum, Tromsø, Norway

Keywords: bioassessment, WFD, bulk, taxonomy, biotic index, metrics

Abstract (ca. 250 words)

Macroinvertebrate assemblages are the most common bioindicators used for stream biomonitoring, yet the standard approach exhibits several time-consuming steps, including the sorting and identification of organisms based on morphological criteria. In this study, we examined if DNA metabarcoding could be used as an efficient molecular-based alternative to the morphology-based monitoring of stream using macroinvertebrates. We compared results achieved with the standard morphological identification of organisms sampled in 18 sites located on 15 French wadeable streams to results obtained with the DNA metabarcoding identification of sorted bulks of the same macroinvertebrate samples, using read numbers (expressed as relative frequencies) as a proxy for abundances. We especially evaluated how combining and filtering metabarcoding data obtained from three different markers (COI: BF1-BR2, 18S: Euka02 and 16S: Inse01) could improve the efficiency of bioassessment.

One hundred and forty taxa were identified based on morphological criteria, and 127 were identified based on DNA metabarcoding of the three markers, with an overlap of 99 taxa. The threshold values used for sequence filtering based on the “best identity” criterion and the number of reads had an effect on the assessment efficiency of data obtained with each marker. Compared to single marker results, combining data from different markers allowed improving the match between biotic index values obtained with the bulk-DNA versus morphology-based approaches. Both approaches assigned the same ecological quality class to a majority (86%) of the site sampling events, highlighting both the efficiency of metabarcoding as a biomonitoring tool but also the need for further research to improve this efficiency.

Introduction

Human activities have major negative impacts on freshwater ecosystems with drastic consequences on their biodiversity at the global scale (Dudgeon et al., 2006, Vörösmarty et al., 2010). In this context, important directives have been implemented to assess the ecological status of freshwater systems (*e.g.* the Water Framework Directive, hereafter abbreviated "WFD", in Europe [Directive 2000/60/EC]) or to protect their integrity (the Clean Water Act in the USA [Public Law 92-500]). There is therefore a major need for tools that allow large-scale, efficient monitoring of the ecological status of water bodies, with the ultimate objectives of identifying the underlying causes of the observed deterioration in water quality (or habitat suitability) and taking the appropriate measures to improve the ecological status of the monitored ecosystems. Such monitoring usually involves the survey of specific groups of organisms, *i.e.* Biological Quality Elements (BQEs) in the WFD: fishes, macroinvertebrates, macrophytes, phytoplankton and diatoms. Standard monitoring approaches are generally based on the morphological identification of sampled organisms using harmonized, intercalibrated protocols. Depending on the standards and the studied BQE, these approaches may exhibit several limitations: (i) they can be destructive/invasive, (ii) they can be resource-intensive, *i.e.* time-consuming and financially expensive, and (iii) they require taxonomic expertise for morphology-based identification, whereas such expertise is continually declining in many countries (Hutchings, 2017; Terlizzi, Bevilacqua, Fraschetti, & Boero, 2003).

DNA metabarcoding is an innovative molecular-based alternative for ecosystem monitoring. This approach consists first in extracting DNA from environment samples of sediment, soil, water, faeces, or directly from community bulks (Hering et al., 2018; Taberlet, Coissac, Pompanon, Brochmann, & Willerslev, 2012). DNA is then amplified using versatile molecular markers and sequenced through high-throughput sequencing. These sequences are compared to those found in reference databases, online and/or developed for a given project (*e.g.* Baird, Pascoe, Zhou, & Hajibabaei, 2011; Rimet et al., 2016), in order to obtain a

73 list of taxa potentially present in the monitored ecosystem or in the bulk sample of the surveyed
74 community.

75 Compared to traditional morphology-based methods, the metabarcoding approach is
76 usually considered as non-invasive (for instance, when directly extracting DNA from water
77 samples; Dejean et al., 2012; Valentini, Pompanon, & Taberlet, 2009), and comparably
78 inexpensive and rapid (Baird & Hajibabaei, 2012; Ji et al., 2013; Taberlet et al., 2012). It can
79 also allow for better taxon identification than morphological expertise (Sweeney, Battle,
80 Jackson, & Dapkey, 2011), and an overall better detection of all the species in aquatic
81 ecosystems (Civade et al., 2016; Valentini et al., 2016). Moreover, metabarcoding also allows
82 a sound estimate of the beta diversity (Ji et al., 2013; Serrana, Miyake, Gamboa, & Watanabe,
83 2019; Sweeney et al., 2011; Yu et al., 2012), and is a reliable source of information for policy
84 making (Ji et al., 2013). Therefore, metabarcoding has been considered as a potential and
85 credible alternative to morphology-based monitoring for both terrestrial and aquatic
86 ecosystems (Baird & Hajibabaei, 2012; Elbrecht, Vamos, Meissner, Aroviita, & Leese, 2017;
87 Shaw, Weyrich, & Cooper, 2017). In freshwaters, several studies have highlighted the
88 biomonitoring potential of metabarcoding, which could efficiently discriminate streams
89 according to their ecological quality (Gibson et al., 2015; Hajibabaei et al., 2011; Ji et al., 2013;
90 Kuntke, de Jonge, Hesselsøe, & Lund Nielsen, 2020; Mächler et al., 2019; Serrana et al., 2019;
91 Sweeney et al., 2011; Zizka, Geiger, & Leese, 2020). For instance, eDNA information on
92 eukaryotic communities in bottom sediments has been strongly associated to land-use
93 pressure types (Xie et al., 2017), and macroinvertebrate bulk data have been used to infer key
94 gradients of stream condition, including dissolved oxygen, dissolved organic carbon, total
95 nitrogen and conductivity (Emilson et al., 2017).

96 The main objective of this study was to test the ability of metabarcoding, applied to
97 standardized bulk samples of benthic macroinvertebrates, to assess the ecological status of
98 streams based on a large-scale biomonitoring program performed within the WFD context.

Using metabarcoding on bulk samples could allow to bypass the organism identification step, which is time-consuming and a source of uncertainty due to operator misidentifications and inter-operator identification variability (Metzeling, Chessman, Hardwick, & Wong, 2003). We also investigated how to improve the bioassesment ability of metabarcoding by testing:

(i) combinations of different markers and primer sets, selected for their ability to identify a large range of benthic macroinvertebrate groups. If individual markers can exhibit some bias in the detection of taxa in a given sample (Elbrecht & Leese, 2015, 2017; Piñol, Mir, Gomez-Polo, & Agustí, 2015), using multiple markers which efficiently amplify different taxonomic groups can limit the global bias of detection. However, increasing the number of markers inflates the time and cost of analyses (Clarke, Soubrier, Weyrich, & Cooper, 2014); and,

(ii) varying threshold values for metabarcoding data filtering. Namely, we investigated the minimal threshold values for best identity match (the percentage of similarity between a sequence and a barcode found in a reference database) and the number of reads for the identified Molecular Operational Taxonomic Units (MOTUs). Classically, a single threshold value is chosen for best identity matches, either one value per taxonomic level of interest (*e.g.* species, genus) or a unique value for all the taxonomic levels. For arthropods, a unique threshold value generally close to 97% is frequently found in the literature (*e.g.* 97% in Elbrecht & Leese, 2017; Serrana et al., 2018; 2019; Yu et al., 2012; 97.5% in Carew, Kellar, Pettigrove, & Hoffmann, 2018; 98% in Lobo, Shokralla, Costa, Hajibabaei, & Costa, 2017). Low abundance (in terms of number of reads) MOTUs, *e.g.* exhibiting a relative abundance lower than 0.003% (Elbrecht & Leese, 2017) or 0.005% (Bokulich et al., 2013; Carew et al., 2018) in a given sample, are usually excluded from metabarcoding data, as they are considered unreliable (*e.g.* Elbrecht & Leese, 2017; Elbrecht et al., 2017).

123 The selection of genetic markers (*e.g.* mitochondrial vs. nuclear, single vs. multicopy,
124 fast evolving vs. conserved, protein-coding vs. ribosomal), the respective primer sets as well
125 as the threshold values are important decisions to take in a metabarcoding experiment, but
126 for which no clear guidelines exist (but see Bokulich et al., 2013). Such decisions may have a
127 strong impact on the DNA-based bioassessment, which could be exacerbated by the high
128 phylogenetic diversity of benthic invertebrates in a single sample. For example, taxonomic
129 groups that are not as well amplified as other groups by a given primer set or marker could be
130 filtered out of the dataset due to their low numbers of reads. The amplification rate can indeed
131 vary greatly among the major taxonomic groups in benthic invertebrate assemblages, even
132 when accounting for biomass (Elbrecht et al., 2017).

133 We systematically studied the efficiency of the three selected markers and of all the
134 possible combinations of these markers, by comparing bulk-DNA and standard results through
135 the lens of the French Multimetric Invertebrate Index (I_2M_2; Mondy, Villeneuve, Archaimbault,
136 & Usseglio-Polatera, 2012). The I_2M_2 was designed as a WFD-compliant index for the
137 invertebrate-based ecological assessment of French wadeable streams. The standardized
138 taxonomic levels needed for calculating the I_2M_2 are mainly the genus level, except for Diptera,
139 Hirudinea and Turbellaria (family level) and Nematoda or Oligochaeta identified as such
140 (standard XP T90-388; AFNOR, 2010). We tested the usefulness of metabarcoding for
141 assessing and discriminating the ecological status of streams, based on a set of streams with
142 a wide range of ecological features.

143

Material & Methods

Sampling sites and data acquisition

Sampling was conducted in 18 sites on 15 streams belonging to five different stream types defined according to stream order (Strahler, 1957) and French hydroecoregions (Wasson, Chandesris, Pella, & Blanc, 2002; Wasson, Chandesris, Pella, Sauquet, & Mengin, 2006) (Fig. 1; see also Table S1 in the Supplementary Material). Streams were selected from two national networks, surveying (i) reference sites (Réseau de Référence, RdR, about 400 sites) and (ii) the mean ecological quality of French streams via the long-term survey of a large selection of sites (Réseau de Contrôle de Surveillance, RCS, about 1500 sites). These surveys have allowed gathering a lot of information on chemical and hydro-morphological pressures impairing water quality and habitats since 2007 (Larras et al., 2017; Mondy et al., 2012). The selection of sites was based on three criteria: their geographic origin (two different hydro-eco-regions x nine streams), pressure intensity and category (water quality degradation or hydrological alteration), and stream type (Wasson et al., 2002). Pressure intensity ranged from “very low” impairment corresponding to “Least Impaired River Reaches” (LIRRs; following Dolédec & Statzner, 2008; Mondy et al., 2012; Statzner, Bady, Dolédec, & Schöll, 2005), to “moderate” or “strong” impairment, both corresponding to “Impaired River Reaches” (IRRs). We selected IRRs which had been impaired by only one main pressure category (water quality degradation or hydrological alteration) over the 2007-2012 period. Sites were *a priori* grouped by triplet within a given stream type, including one LIRR and two IRRs, one IRR exhibiting a moderate impairment and the other a stronger impairment level.

Macroinvertebrate community sampling occurred in autumn 2014 and spring 2015. Thirty-six macroinvertebrate field sampling events were done following a standardized protocol (French standard XP T90-333; AFNOR, 2009), commonly used in France in the

context of the WFD. This protocol advocates the sampling of macroinvertebrates in eight dominant habitats (*i.e.* with an individual share of at least 5% coverage at reach scale) and four marginal ones (*i.e.* with an individual share of less than 5% coverage) in three successive phases (Fig. 2A). Each habitat is characterized by its substrate type (among twelve categories) and its superficial current velocity range in front of the substrate (among four categories). Sampled marginal substrates (phase A; samples 1-4) and the first four dominant ones (phase B, samples 5-8) are selected according to a decreasing gradient of "hosting capacity" (*i.e.* their ability to support a rich and diverse invertebrate assemblage; this gradient is defined by the norm XP T90-333). The four other samples performed on dominant substrates (phase C, samples 9-12) are made in proportion to their individual benthic coverages. A Surber sampler (net mesh size = 500 µm, opening area = 1/20 m²) was used to sample macroinvertebrates in each habitat. Samples were preserved in undenatured alcohol (*ca.* 70% final concentration), for up to one year. In laboratory, organisms were sorted, numbered and identified at the standardized taxonomic level (standard XP T90-388; AFNOR, 2010). Some individuals, difficult to identify at the required taxonomic level (*e.g.* early instars or organisms altered during the sampling process), were identified at the best taxonomic level possible (*i.e.* 29 individuals from eight taxa over 89157 individuals catch during the whole study). After taxonomic identification on morphological criteria, all the sorted organisms, *i.e.* the "bulk samples" (N = 432), were stored in 95° undenatured alcohol (for up to one year) and sent for metabarcoding.

I_2M_2 index calculation

Faunal data (*i.e.* abundance per taxon x phase x site) were combined per site, as indicated in Fig. 2A, to calculate the values of the five individual metrics aggregated in the I_2M_2 index (Mondy et al., 2012). These five metrics are the total taxonomic richness, the Shannon-Weaver diversity index (Shannon & Weaver, 1963), the Average Score Per Taxon (ASPT;

194 Armitage, Moss, Wright, & Furse, 1983), and the proportions of ovoviviparous (Ovoviviparity)
195 and polyvoltine (Polyvoltinism) organisms in the invertebrate assemblage. The I_2M_2 index was
196 constructed and calibrated over 10 chemical and 7 hydromorphological pressure categories
197 (see Mondy et al., 2012, for further details). One sub-index per pressure category is first
198 calculated as the weighted mean of the values of the five metrics, with the weights equal to
199 the ability of each metric to discriminate between LIRRs and IRRs for this pressure category
200 (quantified by its "discrimination efficiency", DE; Ofenböck, Moog, Gerritsen, & Barbour, 2004).
201 Then, the final I_2M_2 index value is calculated as the arithmetic mean of the 17 sub-index values
202 (Mondy et al., 2012).

203

204 *Metabarcoding*

205 Bulk samples were homogenized using an IKA ULTRA-TURRAX Tube Drive control
206 system with sterile 20 mL tubes and 10 steel beads (5 mm Ø) by grinding at 4,000 rpm for 15
207 min (IKA, Staufen im Breisgau, Germany). Complete samples were then incubated overnight
208 at 56 °C in 5 mL of lysis buffer (Tris-HCl 0.1 M, EDTA 0.1 M, NaCl 0.01 M and N-lauroyl
209 sarcosine 1%, pH 7.5–8.0). Extractions were then completed using the DNeasy Blood Tissue
210 Kit (Qiagen GmbH, Hilden, Germany), according to the manufacturer's instructions. DNA
211 extracts were recovered in a total volume of 300 µL. Two DNA extractions were performed per
212 bulk sample. Negative extractions without samples were systematically performed to monitor
213 possible contaminations.
214 Three primer pairs respectively corresponding to three different markers were used for
215 each sample, Inse01 for a mitochondrial 16S rDNA region (Elbrecht et al., 2016; Taberlet,
216 Bonin, Zinger, & Coissac, 2018), Euka02 for a nuclear 18S rDNA region (Guardiola et al.,
217 2015) and BF1-BR2 for the Cytochrome c Oxidase I (COI) region (Elbrecht & Leese, 2017).
218 DNA amplifications were performed in a final volume of 20 µL, using 2 µL of extract DNA as

template. The amplification mixture contained 10 µL of Applied Biosystems™ Master Mix AmpliTaq Gold™ 360, 0.2 µg/µL of bovine serum albumin (BSA, Roche Diagnostic, Basel, Switzerland) and 0.5 µM of each primer for COI and 16S primers or 0.2 µM for 18S primers. Two PCR replicates were amplified for each DNA extraction and each primer pair, for a total of four PCR replicates per bulk sample. The primers were 5'-labeled with an eight-nucleotide tag unique to each replicate (with at least five differences between any pair of tags) allowing the assignment of each sequence to the corresponding sample during sequence analysis. The PCR mixture was denatured at 95°C for 10 min, followed by 35 cycles of 30 s at 95°C, 30 s at 52°C for COI and Inse01 or 45°C for Euka02 and 1 min at 72°C (1m 30s for COI), and followed by a final elongation at 72°C for 7 min. Negative PCR controls (ultrapure water, with 12 replicates as well) were analyzed in parallel to the samples to monitor possible contaminations during the PCR step.

After PCR amplification, PCR products from the same marker were combined in equal volumes and purified using the MinElute (Qiagen GmbH, Hilden, Germany) purification kit. Purified amplicons were checked by high resolution capillary electrophoresis (QIAxcel System, Qiagen GmbH, Hilden, Germany) and sent to Fasteris (Geneva, Switzerland) for library preparation and sequencing. Libraries were prepared according to the PCR-free MetaFast protocol (Taberlet et al., 2018; for further details, also see: https://www.fasteris.com/dna/?q=content/metafast-protocol-amplicon-metagenomic-analysis), which limits chimera formation. The Inse01 and Euka02 amplicons (three libraries each) were sequenced on a HiSeq 2500 platform (Illumina, San Diego, CA, USA) with a paired-end approach (2 × 125 bp), while the COI amplicons (two libraries) were sequenced on a MiSeq platform (Illumina, San Diego, CA, USA) producing 2 x 250 bp paired-end reads.

Workflow for metabarcoding data

Sequences were processed using the OBITools software (Boyer et al., 2016). Each pair of raw reads was paired end merged with `illuminapairedend` to recover the full amplicon sequences. Pairs of reads that did merge with an alignment score above 40 (equivalent to align 10bp of maximal quality on both read ends) were discarded. For the Euka02 primer, as the read length did not allow to recover the full amplicon sequences for important taxa such as Gammaridae, Coleoptera and Trichoptera, pairs of reads whose ends could not be aligned (score < 40) were concatenated and kept separately for further processing. Recovered amplicon sequences were then assigned to their respective sample with `ngsfilter` and dereplicated to get MOTUs with `obiuniq`. MOTUs were then aligned against dedicated reference sequence databases (Ficetola et al., 2020) for each primer pair using `ecotag`. The Euka02 barcodes kept as concatenated sequences were processed separately and the alignment score was based on combining the alignments obtained for both ends of the barcode.

After the taxonomic assignment step, metabarcoding data were subjected to subsequent steps of preparation and filtering before the ultimate step of I_2M_2 index calculation (Fig. 2B). Discordant PCR replicates (*i.e.* that did not cluster when compared to other technical replicates of the same bulk sample) were identified using an iterative process. This iterative process was akin to minimizing the intra-sample distances (between PCR replicates of a given bulk sample) while maximizing the inter-samples distances. At each step of this “min-max” process, the algorithm identified which replicates were the most discordant, *i.e.* the replicates that were too distant from the other replicates of the same bulk sample, and it removed them before iterating. The assumption of this process is that PCR replicates from the same sample should be more similar to one another than to PCR replicates from other samples. To be more specific, at each iteration all the PCR replicates were projected on a 2D space using a Correspondence Analysis (implemented in the ade4 R package; Dray & Dufour, 2007) based on their square root transformed counts. Euclidian distances between all the PCR replicates

in this 2D space were then computed. Distances were partitioned depending on whether or not they involved PCR replicates of the same sample or different samples. These two distance distributions were then compared to pinpoint outlier distances for PCR replicates of the same sample. PCR replicates responsible for these distances were removed and the remaining PCR replicates were used again for a new iteration until no PCR replicates had to be removed. Respectively, we removed 43.4, 7.5 and 9.7% of the replicates for the primer sets COI, Euka02 and Inse01. PCR replicates were thus available for 68.8, 100 and 97.9% of the bulk samples amplified with COI, Euka02 and Inse01, respectively (mean numbers of available replicates per bulk sample = 2.3, 3.7 and 3.6, respectively for COI, Euka02 and Inse01). The remaining PCR replicates were then pooled together, by summing the numbers of reads per MOTU found across all the replicates, for a given bulk sample.

Next, MOTUs were filtered based on the best identity percentage (*i.e.* the percentage of similarity between the MOTU sequence and the closest one identified in the reference database; Ficetola et al., 2020), then, after further sample pooling (Fig. 2A), they were filtered based on their total number of reads (Fig. 2B). For the filtering step based on best identity values, we tested several thresholds, between 80 to 100%. After this first filtering step, we standardized the MOTUs data using the reference list of taxa taken into account for the calculation of the I_2M_2 index value (norm XP T90-388; AFNOR, 2010). As a result, the reads were either pooled by genus, sub-family or family level according to the taxa (or even at a higher taxonomic level; *e.g.* for Oligochaeta), for the MOTUs which could be aggregated to the taxonomic level requested by the standard. MOTUs identified at a taxonomic level too high for being used in the bioevaluation process have been removed from the dataset (*e.g.* MOTUs such as Metazoa, Neoptera or Holometabola). Figure S1, available in the Supplementary Material, shows the remaining total number of reads per PCR replicate at this step of the process. We also tested how a uniform sequencing/read depth would influence the performance of the markers, by filtering out MOTUs with read abundance equal to or lower

than 0.003% per PCR replicate at this step of the process. Preliminary analyses had shown that this added filtering step has only a negligible effect on the biomonitoring results compared to the other filtering steps (namely for best identity and minimal number of reads; see the next paragraph). Therefore, results presented in this study do not include this optional filtering step.

According to the standard used for the index calculation (XP T90-333; AFNOR, 2009; Fig. 2A), the numbers of reads per MOTU were further pooled together within each group of four samples corresponding to the three successive phases of the field sampling protocol (A, B and C, respectively; Fig. 2A). A table summarizing the main identified MOTUs by each marker is available in the Supplementary Material (Table S2). Taxa were then filtered according to their total number of reads in each pooled sample for each marker. We went for a naive and global approach, and systematically tested several threshold values of minimal number of reads, ranging from 1 to 100 reads, five reads by five reads (*i.e.* 1, 5, 10, 15, up to 100).

As we tested all the possible combinations of threshold values for both filtering steps, we obtained a total of 441 datasets for each marker, *i.e.* 21 threshold values for the best identity percentage multiplied by 21 threshold values for the minimal number of reads. After this step, read data were transformed either in relative frequencies (RF) or in presence/absence (PA).

Marker combinations

For each mode of data expression (either RF or PA) we tested the individual marker but also all the different combinations of two (3) or three (1) markers. In order to limit the number of tested marker combinations over all the possible combinations of tested threshold values, we first identified the threshold values that could best allow us to maximize the

correlation (*i.e.* exhibiting the highest adjusted-R^2; see next section) between the results obtained with both the standard and the bulk-DNA approaches. Thus, we first selected the three "best" threshold values of best identity for each individual marker. Then, we selected the five best threshold values for the minimal number of reads, based on the datasets already filtered with the three best identity thresholds previously identified. As a result, we obtained 15 datasets (one per combination of best identity threshold x read minimal number threshold) for each individual marker, 225 datasets (15 x 15 combinations) for each pair of markers (e.g. [COI + Euka02]), and 3375 datasets (*i.e.* 15 x 15 x 15 combinations) for the combination of three markers ([COI + Euka02 + Inse01]).

When combining data from different markers, a taxon was considered as "present" if it was present at least in one of the two or three datasets included in the marker combination. For data coded in RF, for a given taxon, RF values were averaged over all the data provided by the two or three combined datasets.

Statistical analyses

The values of the I_2M_2 index and its five individual metrics were calculated for all the available site sampling events, combinations of markers and selected thresholds, using relative frequencies of reads as a proxy for abundances (*i.e.* for all the metrics except total taxonomic richness and ASPT). For data expressed in presence/absence, the abundance was fixed as equal to one for each identified taxon. The values of the bulk-based index (B-I_2M_2) and its metrics were compared to the values provided by the standard approach ("Morphology-based" I_2M_2; M-I_2M_2 for the index). These values were expressed as Ecological Quality Ratios (EQR; range [0;1]). Reference values, known as "BEST" and "WORST" and needed for the expression of all the metric values as EQRs, were based on the revised I_2M_2 construction dataset (see next paragraph) for both the standard and the bulk-DNA approaches. We also

calculated the discrimination efficiency (DE; Ofenböck et al., 2004) of the I_2M_2 index on each dataset. Here, DE is the relative frequency of IRRs exhibiting I_2M_2 values lower than the first quartile of the distribution of the I_2M_2 values in the LIRRs.

The I_2M_2 was updated in 2016 thanks to a new and bigger available dataset (with 10,074 sampling events, versus only 4,132 sampling events in the original dataset used in Mondy et al., 2012). Based on this new dataset, more robust reference and DE values were (re)calculated, and combinations of phases used for individual metric calculation were revised in order to optimize the discrimination efficiency of metrics (see Fig. 2A). In this study we used this revised version of the I_2M_2. (ref?)

Linear regressions were calculated between the two sets of I_2M_2 index values (or individual metric values) obtained with the standard and bulk-DNA based approaches. For each regression we used the adjusted R^2 as a measure of the variance explained by the model (Crawley, 2007). The distributions of index values provided by both methods were statistically compared over all the site sampling events, with Wilcoxon signed rank tests for paired data. Friedman rank sum tests were used to identify whether or not the tested thresholds for best identity and minimal number of reads (after having selected the best three thresholds for best identity) led to significant differences in adjusted R^2 values over the whole range of tested values. All statistical analyses were done in R (version 3.6.3; R Core Team, 2020).

Results

Taxonomic identification

One hundred and forty morphotaxa were identified in the 36 site sampling events (Fig. 3). The bulk data filtered with the lowest threshold for best identity (80%), allowed to recover 75, 66 and 57 of the morphotaxa, for the markers COI, Euka02, and Inse01, respectively. Twenty-eight morphotaxa were independently recovered by all three markers. Forty-one taxa identified on morphological criteria were not recovered by at least one of the markers. Of these forty-one taxa, six were not referenced as such in our marker-specific reference databases: the phylum Nematoda, the family Rhagionidae, and the genera *Cyphon*, *Agriotypus*, *Capnioneura* and *Lasiocephala*. Twenty-eight taxa were obtained by metabarcoding but not based on morphological identification: 12 taxa for COI, 15 taxa for Euka02 and 9 taxa for Inse01.

Relative frequencies vs. presence/absence

We first examined how the mode of data expression (RF or PA) affected the efficiency of metabarcoding-based bioassessment. Whatever the mode of data expression, we generally found significant and positive linear relationships between B-I_2M_2 and M-I_2M_2 values. For the primer sets COI and Euka02, the relationship was significantly higher on RF data than on PA data (Wilcoxon signed rank test for paired data; $p < 0.001$ for both primer sets; Fig. 4A). For Inse01, adjusted R^2 were not significantly different between RF and PA data (Wilcoxon signed rank test for paired data; $p > 0.05$). Nevertheless, the primer set Inse01 exhibited the highest values of adjusted R^2 with PA data, with a maximum of 0.738. The maximum adjusted R^2 for the primer sets COI and Euka02 were respectively 0.524 and 0.671, with RF data. Based on

these preliminary results, we chose to focus on data expressed in relative frequencies (RF) for the remaining analyses.

Threshold selection for markers

We used adjusted R^2 for examining how the selected threshold value for best identity modified the bioassessment efficiency of each marker. The relationship between the adjusted R^2 and the tested thresholds for the best identity greatly varied among markers (Friedman rank sum tests; $p < 0.001$ for each primer set; Fig. 5). For COI, the thresholds equal or lower than 84% of identity provided the best values of adjusted R^2 (medians > 0.50; Fig. 5A). Adjusted R^2 values decreased with increasing identity thresholds from 85 to 100%, albeit with a small increase of R^2 for identity comprised between 96 and 99%. For Euka02, the adjusted R^2 slowly increased with the threshold values until 96% of identity, with all the medians higher than 0.61 (Fig. 5B), while quickly decreased (median values close to 0.49) at very high values of identity (97-100%). For Inse01, the adjusted R^2 median values globally ranged between 0.58 and 0.62 for thresholds between 80 and 97% of identity. The highest R^2 values were observed with identity values of 98-100% (Fig. 5C). Based on these results, we selected the best identity threshold values of 80, 81 and 82% for COI; 94, 95 and 96% for Euka02; and 98, 99, and 100% for Inse01 (Fig. 5).

The relationships between the adjusted R^2 and the tested thresholds for the minimal number of reads exhibited different patterns of change according to the marker and the selected best identity threshold (Fig. 6). The adjusted R^2 varied significantly according to the tested thresholds for the minimal number of reads (Friedman rank sum tests; $p < 0.001$ for each primer set). For COI, adjusted R^2 was higher than 0.50 for thresholds equal to 5 and 10, and for thresholds higher than 50 reads, but was lower if the selected minimal number of reads was equal to 1 or comprised between 15 and 45 (Fig. 6A). For Euka02, adjusted R^2 increased

412 with increasing minimal number of reads until a plateau at about 65 reads (Fig. 6B). For both
413 markers, the relationships between adjusted R^2 and thresholds for the minimum number of
414 reads were similar for all the tested thresholds for best identity. The patterns of response of
415 Inse01 were similar for the 98 and 99% thresholds for best identity (Fig. 6C), but differed from
416 that obtained with a threshold of 100% of identity. The highest adjusted R^2 values were
417 observed for all the identity thresholds using thresholds comprised between 50 and 70 reads.
418 Based on these results, the threshold values 5, 60, 65, 70 and 100 for COI; 80, 85, 90, 95 and
419 100 for Euka02; and, 50, 55, 60, 65 and 70 for Inse01, were respectively selected (Fig. 6).

420

421 *Biomonitoring efficiency of the bulk-DNA approach*

422 After having identified, for each marker, the best mode of data expression (relative
423 frequencies) and the best ranges of identity threshold and minimal number of reads, we
424 analyzed the global bioassessment efficiency of each individual marker and each combination
425 of markers. We compared the values of the B-I_2M_2, its related metrics and the allocated
426 ecological quality classes with those provided by the standard approach. We also compared
427 the discrimination efficiency of the B-I_2M_2 calculated on each tested combination.

428

429 • *Biomonitoring efficiency of individual markers*

430 The marker Euka02 provided the best regression between M-I_2M_2 and B-I_2M_2 (Fig. 7A)
431 with a best adjusted R^2 of 0.671 for the combination of thresholds “96%:85” (for 96% of best
432 identity and a minimum of 85 reads, afterwards abbreviated min.r), among the 15 combinations
433 of selected thresholds for Euka02. The best adjusted R^2 were equal to 0.524 and 0.665 for
434 COI (80%:5 min.r) and Inse01 (100%:70 min.r), respectively and among the 15 combinations
435 of selected thresholds for each primer set. The I_2M_2 values differed significantly between

morphology-based versus bulk-DNA approaches for all three markers (Wilcoxon signed rank test for paired data; p-values < 0.05; Tab. 1). The ecological quality classes allocated to sites based on both approaches differed also in many cases: *i.e.* 5/36, 16/36 and 18/36 times for Inse01, Euka02 and COI respectively (Tab. 1). Both COI and Euka02 often allocated a worse ecological status to sites than did the morphology-based approach, while Inse01 tended to provide a good match between ecological classes assigned with both approaches (Tab. 1). The discrimination efficiency of the I_2M_2 index values provided by the bulk-DNA approach with COI (DE = 0.875; Tab. 1) was higher than that obtained with the I_2M_2 index values provided by the standard approach ($DE_{standard}$ = 0.833), while for the other markers the discrimination efficiency was poorer (DE = 0.75 for both markers).

The values of the ASPT and the relative frequency of polyvoltine organisms in the invertebrate assemblage (Polyvoltinism in Fig. 7) were correctly modelled from bulk-DNA data for all the markers. Adjusted R^2 ranged from 0.522 (Euka02) to 0.779 (Inse01) for ASPT, and between 0.457 (Inse01) and 0.768 (Euka02) for polyvoltinism. The relative frequency of ovoviviparous organisms (Ovoviviparity in Fig. 7) was correctly modelled by Euka02 and COI, with adjusted R^2 equal to 0.398 and 0.358, respectively, but not by Inse01. Richness was correctly modelled by COI (adj-R^2 = 0.520), but not by Euka02 (adj-R^2 < 0) and Inse01. Whatever the marker, the bulk-DNA approach did not correctly assess the Shannon-Weaver diversity, with adjusted R^2 ranging from -0.004 (COI) to 0.046 (Inse01).

- *Biomonitoring efficiency of marker combinations*

Combining bulk-DNA information provided by different markers improved the match between the standard and bulk-DNA based I_2M_2 values (Fig. 4B & 7B). The best adjusted R^2 between M-I_2M_2 and B-I_2M_2 increased from 0.671 (Euka02) to 0.717 when we combined all

markers information (Fig. 4B & 7B). DE values of marker combinations ranged from 0.833 (for [COI + Inse01]) to 0.958 (for [COI + Euka02] and [COI + Inse01 + Euka02]) (Tab. 1).

The increased quality of the model was readily explained by an increase in the adjusted R^2 values for four out of the five individual metrics included in the I_2M_2 index, although some of these increases remained modest (Fig. 7A & 7B). The adjusted R^2 of total richness and the relative frequency of ovoviviparous organisms in the assemblage greatly increased from 0.520 (COI) to 0.634 [COI + Inse01], and from 0.398 (Euka02) to 0.610 [COI + Euka02], respectively. In contrast, the adjusted R^2 of the Shannon-Weaver index and of the relative frequency of polyvoltine organisms increased marginally from 0.046 (Inse01) to 0.064 [Inse01+ Euka02] and from 0.768 (Euka02) to 0.770 [COI + Euka02], respectively. The adjusted R^2 of the ASPT index decreased for combinations of primer sets from 0.779 (Inse01) to 0.710 [COI + Inse01].

Even if adjusted R^2 values increased, ecological quality classes allocated to tested sites still differed for 9/36, 8/36, 5/36 and 9/36 sampling events respectively for [COI + Euka02], [COI + Inse01], [Inse01 + Euka02] and for the combination of the three markers (Tab. 1). Globally, the combinations of markers allocated a rather worse ecological status than the standard approach (Tab. 1). Values differed significantly between M-I_2M_2 and B-I_2M_2 for [COI + Euka02] and [COI + Inse01] (paired Wilcoxon's test; both $p < 0.01$; Tab. 1), but not for [Inse01 + Euka02] and for the combinations of the three markers (both $p > 0.05$; Tab. 1).

Discussion

Bulk-DNA and biomonitoring

In this study, our objective was to evaluate the efficiency of bulk-DNA metabarcoding of benthic macroinvertebrates to assess the ecological status of wadeable streams. We therefore compared results based on morphological identification of organisms using a standardized protocol (the I_2M_2; Mondy et al., 2012) to those provided by different combinations of metabarcoding markers. We have confirmed the usefulness of bulk-DNA metabarcoding for invertebrate-based stream bioassessment and have identified several strategies to maximize the match between metabarcoding and standard approaches. In addition, B-I_2M_2 and M-I_2M_2 values matched better for two out of the three markers when using the relative number of reads (RF) for each MOTU, instead of their presence/absence only. Aylagas, Borja, Muxika, and Rodríguez-Ezpeleta (2018) also observed better bioassessment efficiency when using the number of reads instead of the presence/absence of MOTUs. Nonetheless, results from Buchner et al. (2019) and Zizka, Geiger, and Leese (2020) have suggested that macroinvertebrate presence/absence data can lead to similar bioassesment results when compared to abundance-based data.

The good similarity observed between B-I_2M_2 and M-I_2M_2 could be explained by the good correlations observed between the values of the individual metrics of the I_2M_2 obtained with both bulk-DNA and standard approaches for four out of the five metrics. Several studies confirmed that bulk data would readily allow to efficiently retrieve metric values initially calculated on data achieved with a standard approach (Elbrecht et al., 2017; Emilson et al., 2017; Gibson et al., 2015; Serrana et al., 2019).

Combining data from multiple markers

Our results also confirmed that combining data from different markers allows improving the appraisal of biodiversity based on bulk samples (Clarke et al., 2014; Hajibabaei, Spall, Shokralla, & van Konynenburg, 2012). Overall, the marker COI (primer set BF1-BR2) exhibited the best performance in terms of number of recovered taxa, but stream bioassessment based on COI data alone was the least efficient. Combining metabarcoding data obtained with at least two different markers has allowed us to increase both the number of taxa recovered with metabarcoding and the quality of the bioassessment. Adding data obtained with a third marker increased the number of recovered taxa, but not the quality of the bioassessment. We therefore suggest that a minimum of two different markers should be used for the biomonitoring of stream based on the bulk-DNA metabarcoding of benthic macroinvertebrates.

Moreover, our results have also highlighted that, depending on the type(s) of metrics included in a biotic index, the choice of the markers should be quite logically also governed by their ability to recover important indicator taxa (*e.g.* EPT) in order to minimize the risk of missing such taxa. For instance, evaluating the ecological status of French streams with the I_2M_2 index would need a combination of markers able to recover Gammaridae (and the generally abundant genus *Gammarus*). This taxon is indeed highly contributing to the trait-based metrics “ovoviviparity” and “polyvoltinism”, and therefore its absence (when filtered out) would explain why both the COI and Euka02 markers exhibited drops of their R^2 past certain best identity thresholds (respectively at 85 and 98%). This observation also confirms the importance of the threshold selection step (also addressed in the next section).

On the importance of thresholds

We have highlighted the importance of the selected thresholds for the best identity value and for the minimal (absolute or relative) number of reads when filtering the taxa which will be considered as present in the sample during the bioinformatics step of the metabarcoding approach. Slightly different thresholds can lead to far different results in terms of M-I_2M_2 and B-I_2M_2 correlations, at least when using individual markers, confirming the results observed by Tapolczai et al. (2019) for diatoms. Moreover, our results have shown that best threshold values were highly different among markers. Therefore, in order to maximize the robustness of metabarcoding-based biomonitoring indices, we suggest the preliminary selection of marker-specific thresholds, based – for instance – on correlation tests between metabarcoding data and abiotic information (*e.g.* metabarcoding-based index values calculated along a pressure gradient, Tapolczai et al., 2019) or biotic information (*e.g.* between metabarcoding- and morphology-based index values, this study).

Ecological quality class allocation

The B-I_2M_2 values efficiently discriminate between least impaired and significantly impaired sites, but misclassifications (*i.e.* differences in the ecological quality class allocated by the metabarcoding-based and the morphology-based approaches) were observed. Comparing both methods, Dowle, Pochon, Banks, Shearer, and Wood (2016) indicated that such changes in quality class allocation were often due to changes in index values that were already close to an ecological quality class boundary. In estuarine and coastal sediment, Aylagas et al. (2018) observed changes in the quality class allocation for 14 of 18 stations, based on benthic macroinvertebrate communities. They also identified a lower rate of change for the stations allocated to the ecological quality classes exhibiting the widest range of values. It should be noted that in our study, the majority of the sampling events (22/36) were performed in sites of “high” ecological quality (based on their M-I_2M_2 value), which is also the quality class

exhibiting the largest extent. This “high” ecological status of many stream sites within the database may also have limited the number of misclassifications compared to other studies. It may as well explain why worse evaluations of the ecological status were mainly observed when classifying site sampling events based on the metabarcoding approach.

In addition, misclassifications were mainly due to discrepancies between the values of the I_2M_2 metrics calculated with both approaches. Such discrepancies were due to: (i) differences in the provided taxonomic lists; (ii) differences in the estimated relative abundances of taxa (*i.e.* relative numbers of individuals vs. relative numbers of reads); and, (iii) inadequacy of the “reference” values needed to calculate the EQR (calculations of Ecological Quality Ratios, as advocated by the WFD, need to normalize metrics values in a 0-1 range, using “reference” values). Potentially promising prospects for improvement will be discussed in the following sections.

On individual metrics

Combining data obtained with at least two different markers has allowed to correctly model the values of four out of the five metrics of the I_2M_2. Namely, we observed good results for the ASPT, the total taxonomic richness, and the relative frequencies of the polyvoltine and ovoviviparous organisms within the invertebrate assemblage, but not for the Shannon-Weaver diversity index. Gibson et al. (2015) also observed a lack of significant, positive relationships between data obtained with standard and bulk-DNA approaches, for Pielou’s evenness and Simpson’s dominance indices. In contrast with the Shannon-Weaver index, the ASPT calculation only needs information on the presence/absence of taxa identified at the family level (Armitage et al., 1983). It probably explains the better robustness of this metric, and why ASPT still exhibited good correlation between values calculated applying both approaches, even if calculated on only 2/3 of the available data on the sampled invertebrate assemblage

575 (both the Shannon-Weaver index and the ASPT are calculated on eight sample units over
576 twelve, corresponding to only two of the three phases; see Fig. 2B and Mondy et al., 2012, for
577 further details).

578 As in previous studies (Hajibabaei et al., 2012; Serrana et al., 2019), morphotaxa not
579 recovered with metabarcoding were often “rare” taxa, *i.e.* taxa with less than 10 individuals
580 over the whole sampling design (observed for 29 out of the 41 missing taxa). Among the five
581 individual metrics combined in the I_2M_2, the Shannon-Weaver index and the relative
582 frequencies of polyvoltine and ovoviviparous invertebrates within the assemblage are the only
583 ones taking into account the taxon abundances. Missing the rarest taxa seems to have only a
584 limited impact on the calculated values of the trait-based metrics, but probably not on those of
585 the Shannon-Weaver diversity index.

586 Metabarcoding also identified taxa absent in the morphotaxa list. Their overall
587 frequency was moderate (20% of the global faunal list, including both identified morphotaxa
588 and DNA-based taxa) when compared, for instance, to the results provided by Serrana et al.
589 (2019), who had observed nine taxa identified only with metabarcoding for a total of 20 taxa.
590 Such taxa could especially lead to bias in values of metrics closely depending on taxonomic
591 richness; *i.e.* total richness and Shannon-Weaver diversity in the I_2M_2. Nevertheless, in our
592 study, a non-negligible part of these taxa corresponded to taxa identified at heterogeneous
593 taxonomic levels among samples; *i.e.* at the family level in some samples (*e.g.* Ephemerellidae)
594 and the genus level (*e.g. Ephemerella*) in other samples from the same site. This bias could
595 be partially addressed by standardizing the taxonomic list at the site scale, instead of at the
596 sample scale as it is currently done in the data preparation step of the I_2M_2 calculation.

597

598 *On reference values and discrimination efficiency*

For best managing the potential bias in taxonomic identification and abundance estimation when using the bulk-DNA approach (Hering et al., 2018), stream-type reference values, “BEST” and “WORST”, could have been specifically redefined for the calculation of the B-I_2M_2 index. During preliminary analyses we tested such an approach, but preliminary results showed that it was not actually pertinent: performances were similar, but we actually observed more misclassifications in ecological quality class allocation when using metabarcoding-specific “BEST” and “WORST” reference values. “BEST” and “WORST” values used routinely for the calculation of the M-I_2M_2 were indeed defined per stream type (for “BEST” values) and on the available national dataset (values currently in use were defined on a dataset containing more than 10,000 site sampling events), whereas “BEST” and “WORST” values could only have been defined on our modest bulk dataset. Therefore “BEST” and “WORST” values defined for metabarcoding data would have been less robust than those used for the M-I_2M_2, and would have represented an additional source of discrepancy between B-I_2M_2 and M-I_2M_2 approaches, both in terms of correlation and ecological status assessment.

Similarly, the I_2M_2 index is calculated as the arithmetic average of 17 sub-indices (one per pressure category), each corresponding to the mean of the individual metrics weighed by their respective DE for the corresponding category of pressure (Mondy et al., 2012). The discrimination efficiency of individual metrics has been calculated at the national scale. Here, the sampling design was too small to have the possibility of calculating robust values of DE specifically allocated to the calculation of bulk-based I_2M_2 values. Defining such DE values in the future should also improve the relationship between M-I_2M_2 and B-I_2M_2.

Perspectives and conclusion

This study supports the bulk metabarcoding approach as a promising method for stream biomonitoring based on bulk-DNA from benthic macroinvertebrate samples (Aylagas,

624 Borja, Rodríguez-Ezpeleta, & Consuegra, 2014; Beentjes et al., 2019; Carew et al. 2018;
625 Carew, Pettigrove, Metzeling, & Hoffmann, 2013; Elbrecht et al., 2017; Emilson et al., 2017;
626 Gibson et al., 2014, 2015; Hajibabaei et al., 2011, 2019; Kuntke et al., 2020; Serrana et al.,
627 2019; Zizka et al., 2020). However, more work is needed before implementing bulk-
628 metabarcoding in the routine monitoring of streams. Indeed, we have highlighted a series of
629 biases leading to the reclassification of several site sampling events in terms of ecological
630 quality class. Solutions exist for reducing these biases, and we have focused on several of
631 them in the previous sections. For instance, a reference DNA barcoding database, including
632 578 different taxa (62% identified at the species level) has been specifically built for this study
633 (Ficetola et al., 2020), in order to work with a database as complete as possible, as
634 recommended by several authors (Aylagas et al., 2014; Hering et al., 2018). Several studies
635 (Gibson et al., 2015; Ji et al., 2013; Mächler et al., 2019; Serrana et al., 2019; Sweeney et al.,
636 2011) have also suggested that new indices could be constructed specifically on bulk-DNA
637 information for stream bioassessment based on macroinvertebrate assemblages, as already
638 been done for benthic diatoms (*e.g.* Vasselon, Domaizon, Rimet, Kahlert, & Bouchez, 2017).
639 On one hand, even if retrieved MOTUs are not assigned to taxa, MOTUs can still be used to
640 efficiently discriminate between impaired and reference situations (*e.g.* Emilson et al., 2017).
641 However, this approach would need a huge preliminary sampling phase in order to construct
642 a new index based on a database that would include the majority of the MOTUs that could be
643 recovered, for instance, for any stream found at the national scale. On the other hand, when
644 molecular methods will prove to be mature enough, they could be used to obtain species-level
645 taxonomic lists. Based on these lists, the observed and reference values of taxonomy-based
646 and trait-based metrics could be refined to further improve the discrimination efficiency of
647 DNA-based indices. Other promising sources of DNA for stream bioassessment are the
648 ethanol used for sample/invertebrate preservation (*e.g.* Hajibabaei et al., 2012; Martins et al.,
649 2020; Zizka, Leese, Peinert, & Geiger, 2019) or even DNA directly extracted from unsorted
650 samples (Pereira-da-Conceicoa et al., 2019). Both sources would allow to avoid the time-

651 consuming steps of invertebrate sorting and counting. Supervised machine learning is also
652 another promising approach, as it could be used to directly model a standard biotic index value
653 based on metabarcoding data (Cordier, Lanzén, Apothéloz-Perret-Gentil, Stoeck, &
654 Pawlowski, 2019; Frühe et al., 2020).

655 Last, one major challenge for an optimal match between standard and bulk-DNA
656 results in a stream bioassessment context is the optimization of taxon abundance estimations.
657 Taxon abundances are taken into account in many invertebrate-based bioassessment
658 methods *(e.g.* in all the intercalibrated European methods; Bennet et al., 2011). In this study,
659 we did not investigate how to better correlate the relative abundances of morphotaxa in
660 samples with information provided by the relative numbers of sequence reads. However,
661 several studies have shown that the numbers of reads could be correlated with taxon
662 abundances or biomasses, albeit frequently with a poor fit (Carew et al., 2013; Deagle,
663 Thomas, Shaffer, Trites, & Jarman, 2013; Dowle et al., 2016; Elbrecht & Leese, 2015, 2017;
664 Serrana et al., 2019), strengthening the idea that better estimating the relative abundances of
665 taxa based on their relative numbers of reads in a bulk sample could improve the modelled
666 values of a biotic index.

667

668 **Acknowledgments**

669 This study was supported by a grant from Agence Nationale de la Recherche (aquaDNA;
670 ANR-13-ECOT-0002-01).
671 We thank SPYGEN staff for its help for DNA extraction.
672 We thank Vasco Elbrecht for his insightful comments. Thanks also go to Jan Pawlowski,
673 Christian Moritz and three anonymous reviewers for improving a previous version of the paper.
674

675 **References**

676
677 AFNOR. (2009). Qualité de l'eau – Prélèvement des macro-invertébrés aquatiques en
678 rivières peu profondes. XP T90-333.
679
680 AFNOR. (2010). Qualité de l'eau – Traitement au laboratoire d'échantillons contenant des
681 macro-invertébrés de cours d'eau. XP T90-388.
682
683 Armitage, P., Moss, D., Wright, J., & Furse, M. (1983). The performance of a new biological
684 water quality score system based on macroinvertebrates over a wide range of unpolluted
685 running-water sites. Water Research, 17(3), 333-347. https://doi.org/10.1016/0043-
686 1354(83)90188-4
687
688 Aylagas, E., Borja, Á., Muxika, I., & Rodríguez-Ezpeleta, N. (2018). Adapting metabarcoding-
689 based benthic biomonitoring into routine marine ecological status assessment networks.
690 *Ecological Indicators*, *95*, 194-202. https://doi.org/10.1016/j.ecolind.2018.07.044
691
692 Aylagas, E., Borja, Á., & Rodríguez-Ezpeleta, N. (2014). Environmental Status Assessment
693 Using DNA Metabarcoding : Towards a Genetics Based Marine Biotic Index (gAMBI). *PLoS*
694 *ONE*, *9*(3), e90529. https://doi.org/10.1371/journal.pone.0090529
695
696 Baird, D. J., & Hajibabaei, M. (2012). Biomonitoring 2.0 : A new paradigm in ecosystem
697 assessment made possible by next-generation DNA sequencing. *Molecular Ecology*, *21*(8),
698 2039-2044. https://doi.org/10.1111/j.1365-294X.2012.05519.x
699
700 Baird, D. J., Pascoe, T. J., Zhou, X., & Hajibabaei, M. (2011). Building freshwater
701 macroinvertebrate DNA-barcode libraries from reference collection material : Formalin
702 preservation vs specimen age. *Journal of the North American Benthological Society*, *30*(1),
703 125-130. https://doi.org/10.1899/10-013.1
704
705 Beentjes, K. K., Speksnijder, A. G. C. L., Schilthuizen, M., Hoogeveen, M., Pastoor, R., & van
706 der Hoorn, B. B. (2019). Increased performance of DNA metabarcoding of macroinvertebrates
707 by taxonomic sorting. *PLOS ONE*, *14*(12), e0226527.
708 https://doi.org/10.1371/journal.pone.0226527
709
710 Bennett, C., Owen, R., Birk, S., Buffagni, A., Erba, S., Mengin, N., … Wasson, J.-G. (2011).
711 Bringing European river quality into line : An exercise to intercalibrate macro-invertebrate
712 classification methods. *Hydrobiologia*, *667*(1), 31-48. https://doi.org/10.1007/s10750-011-
713 0635-2
714
715 Birk, S., Bonne, W., Borja, A., Brucet, S., Courrat, A., Poikane, S., … Hering, D. (2012). Three
716 hundred ways to assess Europe's surface waters : An almost complete overview of biological
717 methods to implement the Water Framework Directive. *Ecological Indicators*, *18*, 31-41.
718 https://doi.org/10.1016/j.ecolind.2011.10.009
719
720 Bokulich, N. A., Subramanian, S., Faith, J. J., Gevers, D., Gordon, J. I., Knight, R., …
721 Caporaso, J. G. (2013). Quality-filtering vastly improves diversity estimates from Illumina
722 amplicon sequencing. *Nature Methods*, *10*(1), 57-59. https://doi.org/10.1038/nmeth.2276
723

Boyer, F., Mercier, C., Bonin, A., Le Bras, Y., Taberlet, P., & Coissac, E. (2016). obitools : A unix-inspired software package for DNA metabarcoding. *Molecular Ecology Resources*, *16*(1), 176-182. https://doi.org/10.1111/1755-0998.12428

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. https://doi.org/10.1371/journal.pone.0226547

Carew, M. E., Kellar, C. R., Pettigrove, V. J., & Hoffmann, A. A. (2018). Can high-throughput sequencing detect macroinvertebrate diversity for routine monitoring of an urban river? *Ecological Indicators*, *85*, 440-450. https://doi.org/10.1016/j.ecolind.2017.11.002

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. https://doi.org/10.1186/1742-9994-10-45

Civade, R., Dejean, T., Valentini, A., Roset, N., Raymond, J.-C., Bonin, A., … Pont, D. (2016). Spatial Representativeness of Environmental DNA Metabarcoding Signal for Fish Biodiversity Assessment in a Natural Freshwater System. *PLOS ONE*, *11*(6), e0157366. https://doi.org/10.1371/journal.pone.0157366

Clarke, L. J., Soubrier, J., Weyrich, L. S., & Cooper, A. (2014). Environmental metabarcodes for insects : *In silico* PCR reveals potential for taxonomic bias. *Molecular Ecology Resources*, *14*(6), 1160-1170. https://doi.org/10.1111/1755-0998.12265

Crawley, M. J. (2007). *The R Book*. John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470515075

Cordier, T., Lanzén, A., Apothéloz-Perret-Gentil, L., Stoeck, T., & Pawlowski, J. (2019). Embracing Environmental Genomics and Machine Learning for Routine Biomonitoring. *Trends in Microbiology*, *27*(5), 387-397. https://doi.org/10.1016/j.tim.2018.10.012

Deagle, B. E., Thomas, A. C., Shaffer, A. K., Trites, A. W., & Jarman, S. N. (2013). Quantifying sequence proportions in a DNA-based diet study using Ion Torrent amplicon sequencing : Which counts count? *Molecular Ecology Resources*, *13*(4), 620-633. https://doi.org/10.1111/1755-0998.12103

Dejean, T., Valentini, A., Miquel, C., Taberlet, P., Bellemain, E., & Miaud, C. (2012). Improved detection of an alien invasive species through environmental DNA barcoding : The example of the American bullfrog *Lithobates catesbeianus*: *Alien invasive species detection using eDNA*. *Journal of Applied Ecology*, *49*(4), 953-959. https://doi.org/10.1111/j.1365-2664.2012.02171.x

Dolédec, S., & Statzner, B. (2008). Invertebrate traits for the biomonitoring of large European rivers : An assessment of specific types of human impact. *Freshwater Biology*, *53*(3), 617-634. https://doi.org/10.1111/j.1365-2427.2007.01924.x

Dowle, E. J., Pochon, X., Banks, J., Shearer, K., & Wood, S. A. (2016). Targeted gene enrichment and high-throughput sequencing for environmental biomonitoring : A case study using freshwater macroinvertebrates. *Molecular Ecology Resources*, *16*(5), 1240-1254. https://doi.org/10.1111/1755-0998.12488

Dray, S., & Dufour, A.-B. (2007). The ade4 Package : Implementing the Duality Diagram for Ecologists. *Journal of Statistical Software*, *22*(4). https://doi.org/10.18637/jss.v022.i04

Dudgeon, D., Arthington, A. H., Gessner, M. O., Kawabata, Z.-I., Knowler, D. J., Lévêque, C., … Sullivan, C. A. (2006). Freshwater biodiversity : Importance, threats, status and conservation challenges. *Biological Reviews*, *81*(02), 163. https://doi.org/10.1017/S1464793105006950

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. https://doi.org/10.1371/journal.pone.0130324

Elbrecht, V., & Leese, F. (2017). *Validation and development of COI metabarcoding primers for freshwater macroinvertebrate bioassessment* [Preprint]. PeerJ Preprints. https://doi.org/10.7287/peerj.preprints.2044v5

Elbrecht, V., Taberlet, P., Dejean, T., Valentini, A., Usseglio-Polatera, P., Beisel, J.-N., Coissac, E., Boyer, F., & Leese, F. (2016). Testing the potential of a ribosomal 16S marker for DNA metabarcoding of insects. *PeerJ*, *4*, e1966. https://doi.org/10.7717/peerj.1966

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. https://doi.org/10.1111/2041-210X.12789

Emilson, C. E., Thompson, D. G., Venier, L. A., Porter, T. M., Swystun, T., Chartrand, D., … Hajibabaei, M. (2017). DNA metabarcoding and morphological macroinvertebrate metrics reveal the same changes in boreal watersheds across an environmental gradient. *Scientific Reports*, *7*(1), 12777. https://doi.org/10.1038/s41598-017-13157-x

Ficetola, G. F., Boyer, F., Valentini, A., Bonin, A., Meyer, A., Dejean, T., Gaboriaud, C., Usseglio- Polatera, P., & Taberlet, P. (2020). Comparison of markers for the monitoring of freshwater benthic biodiversity through DNA metabarcoding. *Molecular Ecology*. https://doi.org/10.1111/mec.15632

Frühe, L., Cordier, T., Dully, V., Breiner, H., Lentendu, G., Pawlowski, J., Martins, C., Wilding, T. A., & Stoeck, T. (2020). Supervised machine learning is superior to indicator value inference in monitoring the environmental impacts of salmon aquaculture using eDNA metabarcodes. *Molecular Ecology*, mec.15434. https://doi.org/10.1111/mec.15434

Gibson, J. F., Shokralla, S., Curry, C., Baird, D. J., Monk, W. A., King, I., & Hajibabaei, M. (2015). Large-Scale Biomonitoring of Remote and Threatened Ecosystems via High-Throughput Sequencing. *PLOS ONE*, *10*(10), e0138432. https://doi.org/10.1371/journal.pone.0138432

Gibson, J., Shokralla, S., Porter, T. M., King, I., van Konynenburg, S., Janzen, D. H., … Hajibabaei, M. (2014). Simultaneous assessment of the macrobiome and microbiome in a bulk sample of tropical arthropods through DNA metasystematics. *Proceedings of the National Academy of Sciences*, *111*(22), 8007-8012. https://doi.org/10.1073/pnas.1406468111

828 Guardiola, M., Uriz, M. J., Taberlet, P., Coissac, E., Wangensteen, O. S., & Turon, X. (2015).
829 Deep-Sea, Deep-Sequencing : Metabarcoding Extracellular DNA from Sediments of Marine
830 Canyons. *PLOS ONE*, *10*(10), e0139633. https://doi.org/10.1371/journal.pone.0139633
831
832 Hajibabaei, M., Porter, T. M., Robinson, C. V., Baird, D. J., Shokralla, S., & Wright, M. T. G.
833 (2019). Watered-down biodiversity? A comparison of metabarcoding results from DNA
834 extracted from matched water and bulk tissue biomonitoring samples. *PLOS ONE*, *14*(12),
835 e0225409. https://doi.org/10.1371/journal.pone.0225409
836
837 Hajibabaei, M., Shokralla, S., Zhou, X., Singer, G. A. C., & Baird, D. J. (2011). Environmental
838 Barcoding : A Next-Generation Sequencing Approach for Biomonitoring Applications Using
839 River Benthos. *PLoS ONE*, *6*(4), e17497. https://doi.org/10.1371/journal.pone.0017497
840
841 Hajibabaei, M., Spall, J. L., Shokralla, S., & van Konynenburg, S. (2012). Assessing
842 biodiversity of a freshwater benthic macroinvertebrate community through non-destructive
843 environmental barcoding of DNA from preservative ethanol. *BMC Ecology*, *12*(1), 28.
844 https://doi.org/10.1186/1472-6785-12-28
845
846 Hering, D., Borja, A., Jones, J. I., Pont, D., Boets, P., Bouchez, A., … Kelly, M. (2018).
847 Implementation options for DNA-based identification into ecological status assessment under
848 the European Water Framework Directive. *Water Research*, *138*, 192-205.
849 https://doi.org/10.1016/j.watres.2018.03.003
850
851 Hutchings, P. (2019). An advocate for taxonomic research in Australia. *Pacific Conservation*
852 *Biology*, *25*(1), 34. https://doi.org/10.1071/PC17033
853
854 Ji, Y., Ashton, L., Pedley, S. M., Edwards, D. P., Tang, Y., Nakamura, A., … Yu, D. W. (2013).
855 Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding. *Ecology Letters*,
856 *16*(10), 1245-1257. https://doi.org/10.1111/ele.12162
857
858 Kuntke, F., de Jonge, N., Hesselsøe, M., & Lund Nielsen, J. (2020). Stream water quality
859 assessment by metabarcoding of invertebrates. *Ecological Indicators*, *111*, 105982.
860 https://doi.org/10.1016/j.ecolind.2019.105982
861
862 Larras, F., Coulaud, R., Gautreau, E., Billoir, E., Rosebery, J., & Usseglio-Polatera, P. (2017).
863 Assessing anthropogenic pressures on streams : A random forest approach based on benthic
864 diatom communities. *Science of The Total Environment*, *586*, 1101-1112.
865 https://doi.org/10.1016/j.scitotenv.2017.02.096
866
867 Lobo, J., Shokralla, S., Costa, M. H., Hajibabaei, M., & Costa, F. O. (2017). DNA
868 metabarcoding for high-throughput monitoring of estuarine macrobenthic communities.
869 *Scientific Reports*, *7*(1), 15618. https://doi.org/10.1038/s41598-017-15823-6
870
871 Mächler, E., Little, C. J., Wüthrich, R., Alther, R., Fronhofer, E. A., Gounand, I., … Altermatt,
872 F. (2019). Assessing different components of diversity across a river network using eDNA.
873 *Environmental DNA*, *1*(3), 290-301. https://doi.org/10.1002/edn3.33
874
875 Martins, F. M. S., Porto, M., Feio, M. J., Egeter, B., Bonin, A., Serra, S. R. Q., Taberlet, P., &
876 Beja, P. (2020). Modelling technical and biological biases in macroinvertebrate community
877 assessment from bulk preservative using multiple metabarcoding markers. *Molecular Ecology*.
878 https://doi.org/10.1111/mec.15620
879

Metzeling, L., Chessman, B., Hardwick, R., & Wong, V. (2003). Rapid assessment of rivers using macroinvertebrates : The role of experience, and comparisons with quantitative methods. *Hydrobiologia*, *510*(1-3), 39-52. https://doi.org/10.1023/B:HYDR.0000008500.34301.a0

Mondy, C. P., Villeneuve, B., Archaimbault, V., & Usseglio-Polatera, P. (2012). A new macroinvertebrate-based multimetric index (I2M2) to evaluate ecological quality of French wadeable streams fulfilling the WFD demands : A taxonomical and trait approach. *Ecological Indicators*, *18*, 452-467. https://doi.org/10.1016/j.ecolind.2011.12.013

Ofenböck, T., Moog, O., Gerritsen, J., & Barbour, M. (2004). A Stressor Specific Multimetric Approach for Monitoring Running Waters in Austria Using Benthic Macro-Invertebrates. In D. Hering, P. F. M. Verdonschot, O. Moog, & L. Sandin (Éds.), *Integrated Assessment of Running Waters in Europe* (p. 251-268). Springer Netherlands. https://doi.org/10.1007/978-94-007-0993-5_15

Pereira-da-Conceicoa, L., Elbrecht, V., Hall, A., Briscoe, A., Barber-James, H., & Price, B. (2019). *Metabarcoding unsorted kick-samples facilitates macroinvertebrate-based biomonitoring with increased taxonomic resolution, while outperforming environmental DNA* [Preprint]. https://doi.org/10.1101/792333

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*(4), 819-830. https://doi.org/10.1111/1755-0998.12355

R Core Team. (2020). *R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.* https://www.R-project.org/

Rimet, F., Chaumeil, P., Keck, F., Kermarrec, L., Vasselon, V., Kahlert, M., … Bouchez, A. (2016). R-Syst::diatom : An open-access and curated barcode database for diatoms and freshwater monitoring. *Database*, *2016*, baw016. https://doi.org/10.1093/database/baw016

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. https://doi.org/10.1016/j.ecolind.2019.02.008

Serrana, J. M., Yaegashi, S., Kondoh, S., Li, B., Robinson, C. T., & Watanabe, K. (2018). Ecological influence of sediment bypass tunnels on macroinvertebrates in dam-fragmented rivers by DNA metabarcoding. *Scientific Reports*, *8*(1), 10185. https://doi.org/10.1038/s41598-018-28624-2

Shannon, C. E., & Weaver, W. (1963). The mathematical theory of communication. Univ. Illinois Press, Urbana, USA.

Shaw, J. L. A., Weyrich, L., & Cooper, A. (2017). Using environmental (e)DNA sequencing for aquatic biodiversity surveys : A beginner's guide. *Marine and Freshwater Research*, *68*(1), 20. https://doi.org/10.1071/MF15361

Statzner, B., Bady, P., Dolédec, S., & Schöll, F. (2005). Invertebrate traits for the biomonitoring of large European rivers : An initial assessment of trait patterns in least impacted river reaches. *Freshwater Biology*, *50*(12), 2136-2161. https://doi.org/10.1111/j.1365-2427.2005.01447.x

Strahler, A. N. (1957). Quantitative analysis of watershed geomorphology. *Transactions, American Geophysical Union*, *38*(6), 913. https://doi.org/10.1029/TR038i006p00913

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. https://doi.org/10.1899/10-016.1

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

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. https://doi.org/10.1111/j.1365-294X.2012.05470.x

Tapolczai, K., Vasselon, V., Bouchez, A., Stenger-Kovács, C., Padisák, J., & Rimet, F. (2019). 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. https://doi.org/10.1002/ece3.4701

Terlizzi, A., Bevilacqua, S., Fraschetti, S., & Boero, F. (2003). Taxonomic sufficiency and the increasing insufficiency of taxonomic expertise. *Marine Pollution Bulletin*, *46*(5), 556-561. https://doi.org/10.1016/S0025-326X(03)00066-3

Valentini, A., Pompanon, F., & Taberlet, P. (2009). DNA barcoding for ecologists. *Trends in Ecology & Evolution*, *24*(2), 110-117. https://doi.org/10.1016/j.tree.2008.09.011

Valentini, A., Taberlet, P., Miaud, C., Civade, R., Herder, J., Thomsen, P. F., … Dejean, T. (2016). Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. *Molecular Ecology*, *25*(4), 929-942. https://doi.org/10.1111/mec.13428

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*(1), 162-177. https://doi.org/10.1086/690649

Vörösmarty, C. J., McIntyre, P. B., Gessner, M. O., Dudgeon, D., Prusevich, A., Green, P., … Davies, P. M. (2010). Global threats to human water security and river biodiversity. *Nature*, *467*(7315), 555-561. https://doi.org/10.1038/nature09440

Wasson, J.-G., Chandesris, A., Pella, H., & Blanc, L. (2002). *Les hydro-écorégions de France métropolitaine — Approche régionale de la typologie des eaux courantes et éléments pour la définition des peuplements de référence d'invertébrés*.

Wasson, J.-G., Chandesris, A., Pella, H., Sauquet, E., & Mengin, N. (2006). *Appui scientifique à la mise en œuvre de la Directive Européenne Cadre sur l'Eau. Typologie des cours d'eau de France métropolitaine*.

Xie, Y., Wang, J., Yang, J., Giesy, J. P., Yu, H., & Zhang, X. (2017). Environmental DNA metabarcoding reveals primary chemical contaminants in freshwater sediments from different land-use types. *Chemosphere*, *172*, 201-209. https://doi.org/10.1016/j.chemosphere.2016.12.117

Yu, D. W., Ji, Y., Emerson, B. C., Wang, X., Ye, C., Yang, C., & Ding, Z. (2012). Biodiversity soup : Metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring: Biodiversity soup. *Methods in Ecology and Evolution*, *3*(4), 613-623. https://doi.org/10.1111/j.2041-210X.2012.00198.x

Zizka, V. M. A., Geiger, M. F., & Leese, F. (2020). DNA metabarcoding of stream invertebrates reveals spatio-temporal variation but consistent status class assessments in a natural and urban river. *Ecological Indicators*, 115, 106383. https://doi.org/10.1016/j.ecolind.2020.106383

Zizka, V. M. A., Leese, F., Peinert, B., & Geiger, M. F. (2019). DNA metabarcoding from sample fixative as a quick and voucher-preserving biodiversity assessment method. *Genome*, *62*(3), 122-136. https://doi.org/10.1139/gen-2018-0048

Data Accessibility

All datasets will be available at figshare upon manuscript acceptance: https://doi.org/10.6084/m9.figshare.13110899 (bulk-DNA data) and https://doi.org/10.6084/m9.figshare.13110692 (faunal data)

Author Contributions

Designed the research: Pierre Taberlet, Tony Dejean, Gentile Francesco Ficetola, Frédéric Boyer, Alice Valentini, Aurélie Bonin, Albin Meyer, Philippe Usseglio-Polatera, Jean-Nicolas Beisel, Florian Leese.

Performed research: Alice Valentini, Aurélie Bonin, Albin Meyer, Coline Gaboriaud, Pierre Taberlet, Tony Dejean, Gentile Francesco Ficetola, Frédéric Boyer, Philippe Usseglio-Polatera, Philippe Wagner, Jonathan Bouquerel, Jean-Nicolas Beisel.

Analyzed data: Albin Meyer, Philippe Usseglio-Polatera, Gentile Francesco Ficetola, Frédéric Boyer, Aurélie Bonin.

Wrote the paper: Albin Meyer wrote the first draft of the paper, with subsequent contribution of all the authors.

Tables

Tab.1 Total abundance and number of reads for all the taxa identified with the morphology- and/or bulk-based approaches. Bulk data were filtered with a homogeneous best identity threshold of 0.80.

Higher rank	Family	Taxon	Abundance	Reads (COI)	Reads (EUKA02)	Reads (INSE01)
Oligochaeta			6971	87902	4023314	2091067
Hirudinea	Erpobdellidae		158	4169	313919	1567
	Glossiphoniidae		92	4274	27341	0
	Piscicolidae		6	181	4380	0
Hydracarina			86	32	102	0
Crustacea	Gammaridae	*Echinogammarus*	0	0	985	0
	Gammaridae	*Gammarus*	20318	209364	943643	0
	Astacidae	*Pacifastacus*	1	0	0	0
	Cambaridae	*Procambarus*	0	65	0	0
	Asellidae		70	943	0	1975
Coleoptera	Chrysomelidae		0	0	0	1039
	Curculionidae		27	0	44087	15211
	Dytiscidae	Colymbetinae	14	0	100	0
	Dytiscidae		1	0	0	0
	Dytiscidae	Hydroporinae	127	5041	79	0
	Dytiscidae	Laccophilinae	2	0	0	0
	Elmidae		0	0	516	0
	Elmidae	*Elmis*	1171	5488	25160	227850
	Elmidae	*Esolus*	963	2911	23131	68410
	Elmidae	*Limnius*	1309	23073	142941	1017298
	Elmidae	*Macronychus*	2	0	0	128
	Elmidae	*Oulimnius*	581	2792	1315	18411
	Elmidae	*Riolus*	716	12770	22435	96015
	Gyrinidae	*Orectochilus*	31	128	0	1144
	Haliplidae	*Haliplus*	8	142	0	0
	Hydraenidae	*Hydraena*	336	829	61998	16424
	Hydrophilidae		3	0	0	0
	Hydrophilidae	Hydrophilinae	0	0	383	0
	Hydrophilidae	Sphaeridiinae	1	0	0	0
	Scirtidae	*Cyphon*	3	0	0	0
	Scirtidae	*Helodes*	70	0	0	0
	Scirtidae	*Hydrocyphon*	3	0	0	0
Diptera	Anthomyidae		3	539	0	0
	Athericidae		624	20069	0	2895
	Blephariceridae		16	0	0	0
	Ceratopogonidae		296	3989	0	0
	Chaoboridae		1	0	0	0
	Chironomidae		18992	14658	0	18513
	Culicidae		0	1450	0	582
	Dixidae		5	0	0	0
	Empididae		218	414	0	0
	Ephydridae		4	0	0	0
	Limoniidae		508	8357	375057	428638
	Psychodidae		147	534	0	0

1020

1021 Tab.1 (continued)

Higher rank	Family	Taxon	Abundance	Reads (COI)	Reads (EUKA02)	Reads (INSE01)
Diptera	Ptychopteridae		0	13	0	0
	Rhagionidae		1	0	0	0
	Simuliidae		4287	27203	54558	780508
	Stratiomyidae		9	0	0	0
	Tabanidae		8	14036	0	0
	Tipulidae		14	5924	0	18095
Ephemeroptera	Baetidae		0	0	0	21088
	Baetidae	*Baetis*	7063	282764	3766695	3664057
	Baetidae	*Cloeon*	4	187	0	0
	Caenidae	*Brachycercus*	1	0	0	0
	Caenidae	*Caenis*	330	0	25517	529
	Ephemerellidae	*Ephemerella*	2757	33602	204825	31769
	Ephemerellidae		0	57357	686	0
	Ephemerellidae	*Torleya*	99	0	576400	0
	Ephemeridae	*Ephemera*	53	25640	57405	566199
	Heptageniidae	*Ecdyonurus*	89	63253	489674	1250160
	Heptageniidae	*Electrogena*	0	0	0	649
	Heptageniidae	*Epeorus*	300	5092	264696	83753
	Heptageniidae	*Heptagenia*	1	2	681	259
	Heptageniidae	*Rhithrogena*	780	75947	1149041	2130344
	Leptophlebiidae	*Habroleptoides*	708	22839	0	741183
	Leptophlebiidae	*Habrophlebia*	100	1280	291009	115316
	Leptophlebiidae	*Leptophlebia*	0	0	171	0
	Leptophlebiidae		0	0	15370	0
	Leptophlebiidae	*Paraleptophlebia*	6	0	223154	21646
	Potamanthidae	*Potamanthus*	0	0	21	10
	Siphlonuridae	*Siphlonurus*	0	0	936	11
Hemiptera	Aphelocheiridae	*Aphelocheirus*	49	2492	34159	129833
	Corixidae	Corixinae	1	0	692	1336
	Corixidae	*Micronecta*	119	0	559	0
	Mesoveliidae	*Mesovelia*	5	0	0	0
Hymenoptera	Agriotypidae	*Agriotypus*	2	0	0	0
Lepidoptera	Crambidae		1	0	1917	114
Megaloptera	Sialidae	*Sialis*	8	467	0	0
Odonata	Aeshnidae	*Aeshna*	0	2316	0	0
	Aeshnidae	*Anax*	2	0	0	0
	Calopterygidae	*Calopteryx*	24	270	0	66203
	Coenagrionidae		2	0	345	0
	Cordulegasteridae	*Cordulegaster*	16	283	0	49980
	Gomphidae	*Gomphus*	1	0	0	0
	Gomphidae	*Onychogomphus*	1	982	0	0
	Platycnemididae	*Platycnemis*	1	804	0	825

1022

1023 Tab.1 (continued)

Higher rank	Family	Taxon	Abundance	Reads (COI)	Reads (EUKA02)	Reads (INSE01)
Plecoptera	Capniidae	*Capnia*	2	0	24	0
	Capniidae		0	0	0	32
	Capniidae	*Capnioneura*	14	0	0	0
	Chloroperlidae	*Chloroperla*	27	278	0	26272
	Chloroperlidae		0	28	0	0
	Chloroperlidae	*Siphonoperla*	559	7380	1711584	436685
	Leuctridae	*Leuctra*	1563	2704	779840	1232887
	Nemouridae	*Amphinemura*	1018	4517	0	362313
	Nemouridae	*Nemoura*	278	355	0	2335
	Nemouridae		0	76	941894	7
	Nemouridae	*Nemurella*	2	0	0	10378
	Nemouridae	*Protonemura*	3469	129089	0	2148653
	Perlidae	*Dinocras*	13	3938	31865	27300
	Perlidae	*Perla*	36	0	0	0
	Perlodidae	*Dictyogenus*	5	0	0	32162
	Perlodidae	*Diura*	2	0	132202	225296
	Perlodidae	*Isoperla*	175	20271	0	89548
	Perlodidae	*Perlodes*	3	11896	0	78653
	Perlodidae		1	0	0	0
	Taeniopterygidae	*Brachyptera*	4	0	1845	4901
	Taeniopterygidae	*Rhabdiopteryx*	0	0	712	0
	Taeniopterygidae	*Taeniopteryx*	62	148	3123	0
Trichoptera	Beraeidae	*Beraeodes*	0	0	13651	0
	Beraeidae	*Ernodes*	1	0	0	0
	Brachycentridae		0	4465	0	0
	Brachycentridae	*Brachycentrus*	8	0	210	0
	Brachycentridae	*Micrasema*	670	8496	17315	0
	Glossosomatidae	*Agapetus*	581	0	1230	0
	Glossosomatidae	*Glossosoma*	157	3385	68511	0
	Goeridae	*Goera*	5	0	0	0
	Goeridae		1	0	1	0
	Goeridae	*Lithax*	60	0	0	0
	Goeridae	*Silo*	92	1890	26334	398
	Hydropsychidae	*Cheumatopsyche*	79	1247	1127	0
	Hydropsychidae	*Hydropsyche*	3120	91288	7170	378
	Hydroptilidae	*Hydroptila*	208	435	0	0
	Hydroptilidae	*Orthotrichia*	0	1474	0	0
	Hydroptilidae	*Ptilocolepus*	1	0	0	0
	Hydroptilidae	*Stactobia*	1	0	0	0
	Lepidostomatidae	*Crunoecia*	6	0	0	0
	Lepidostomatidae	*Lasiocephala*	4	0	0	0
	Lepidostomatidae	*Lepidostoma*	316	1463	0	0

1024

1025 Tab.1 (continued)

Higher rank	Family	Taxon	Abundance	Reads (COI)	Reads (EUKA02)	Reads (INSE01)
Trichoptera	Leptoceridae	*Adicella*	20	326	269	0
	Leptoceridae	*Athripsodes*	7	0	0	122
	Leptoceridae	*Mystacides*	40	258	187	0
	Leptoceridae	*Oecetis*	4	178	304	0
	Limnephilidae	Drusinae	80	4998	25279	0
	Limnephilidae		10	0	0	0
	Limnephilidae	Limnephilinae	1133	66589	100892	186822
	Odontoceridae	*Odontocerum*	253	0	0	0
	Philopotamidae	*Philopotamus*	72	970	0	0
	Philopotamidae	*Wormaldia*	1	0	0	0
	Polycentropodidae	*Holocentropus*	40	0	0	0
	Polycentropodidae	*Plectrocnemia*	105	0	0	8023
	Polycentropodidae	*Polycentropus*	137	171	2654	0
	Psychomyiidae	*Lype*	1	0	0	0
	Psychomyiidae	*Psychomyia*	530	1041	12772	119483
	Psychomyiidae	*Tinodes*	70	0	4177	173
	Rhyacophilidae	*Rhyacophila*	677	8312	108836	910479
	Sericostomatidae	*Sericostoma*	821	51100	0	546599
	Sericostomatidae		0	0	87001	0
Hydrozoa			0	2	0	0
Bivalvia	Sphaeriidae	*Pisidium*	444	0	0	0
	Sphaeriidae	*Sphaerium*	0	0	323460	2810
Gastropoda	Acroloxidae	*Acroloxus*	1	0	0	0
	Ancylidae	*Ancylus*	342	5113	699087	425
	Bithyniidae	*Bithynia*	6	114	4477	0
	Hydrobiidae	*Potamopyrgus*	145	961	273236	0
	Lymnaeidae	*Galba*	3	0	0	0
	Lymnaeidae	*Lymnaea*	1	0	0	0
	Lymnaeidae		1	0	14	0
	Lymnaeidae	*Radix*	76	11975	484006	0
	Neritidae	*Theodoxus*	0	82	0	0
	Physidae	*Physa stricto-sensu*	17	0	0	0
	Physidae	*Physella*	10	0	0	0
	Physidae		0	530	223269	0
	Planorbidae		10	0	960	0
	Valvatidae	*Valvata*	7	0	209	0
Nemathelmintha			1	0	0	0
Turbellaria	Dugesiidae		296	221	270076	0
	Planariidae		527	1847	1671815	0
Porifera	Spongillidae		0	0	108	0

1026

Tab. 2. Summary of the results corresponding to the best match between morphology-based and bulk-DNA based approaches, obtained with each individual primer set and each combination of primer sets. Wilcoxon tests and adjusted R^2 (from linear regressions) are applied/calculated between the values of the I_2M_2 index obtained with morphology-based versus bulk-DNA based approaches. The thresholds for best identity and for minimum number of reads are respectively provided in brackets for each primer set or combination of primer sets. DE = Discrimination Efficiency.

				Reclassifications							
			Wilcoxon tests	Worse predicted state				No changes	Better predicted state		
Bulk data	Adjusted R^2	DE	(p-value)	-4	-3	-2	-1	0	+1	+2	Total
COI (83:0)	0.396	0.833	0.119	1			4	27	4		9/36
EUKA02 (96:85)	0.544	0.750	0.001			3	11	19	2	1	17/36
INSE01 (100:10)	0.611	0.875	0.010		1		7	26	2		10/36
COI (82:5) + EUKA02 (95:85)	0.648	0.875	0.002		1	1	7	25	1	1	11/36
COI (84:10) + INSE01 (100:10)	0.740	0.875	0.003			1	9	23	2		12/36
INSE01 (100:5) + EUKA02 (96:85)	0.697	0.875	0.000			2	9	23	2		13/36
COI (82:5) + EUKA02 (95:85) + INSE01 (100:5)	0.698	0.958	0.002			1	8	25	2		11/36

Figures

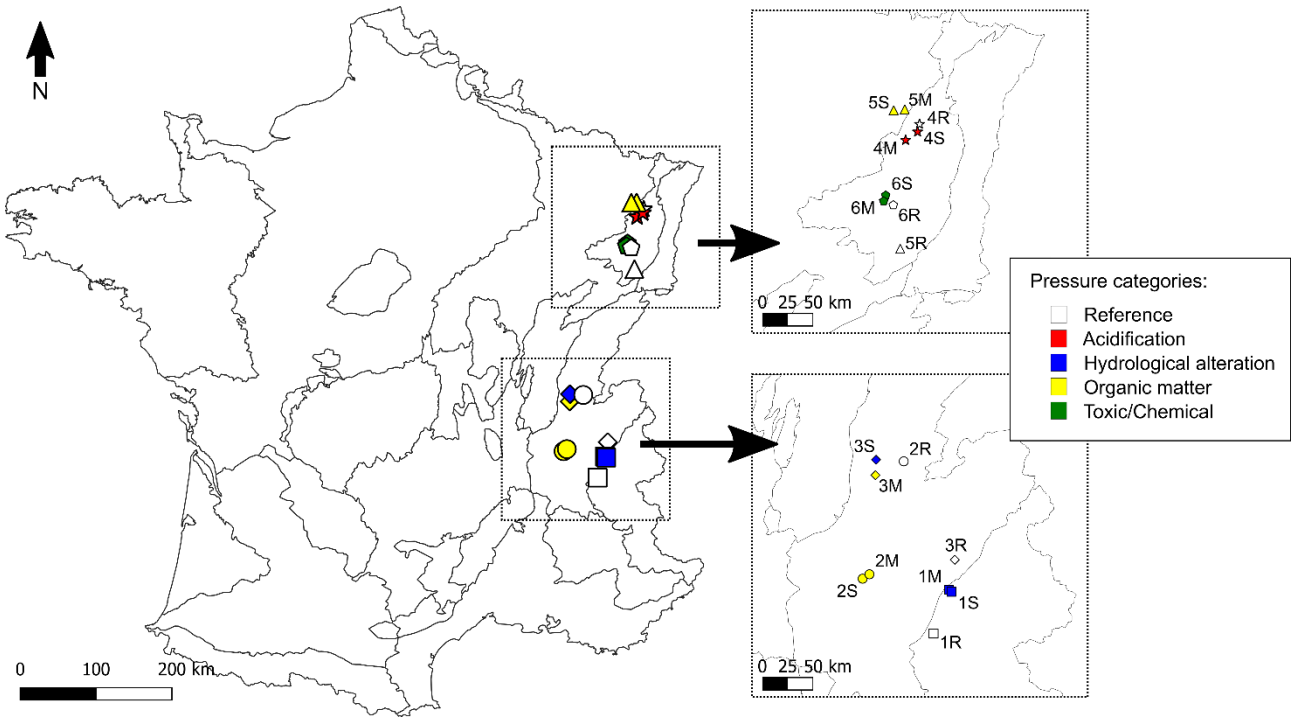


Fig. 1: Location of the sampled sites. Delineations indicate the main French hydro-eco-regions (Wasson et al. 2002). The number in the code name of each sampled site gives the information about its group membership (from 1 to 6), whereas the letter gives information about its status, either a LIRR ("Reference"; "R") or a IRR, exhibiting either a Moderate "M" pressure intensity or a Strong "S" pressure intensity (see first section of Material & Methods for further details). All three sites found in a group exhibit the same stream type. Further information about the sites are available in the Supplementary Material (Table S1).

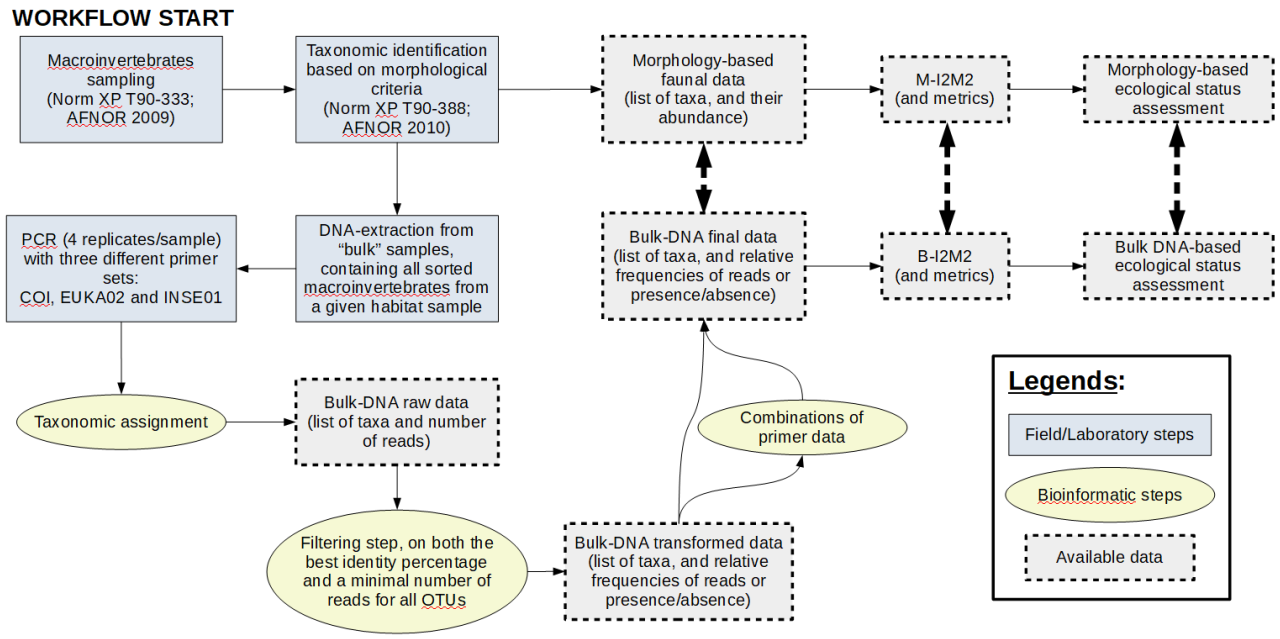


Fig. 2: Workflow. Dashed double-arrowed lines indicate comparisons of the results obtained with both the morphological-based approach and the bulk DNA one.

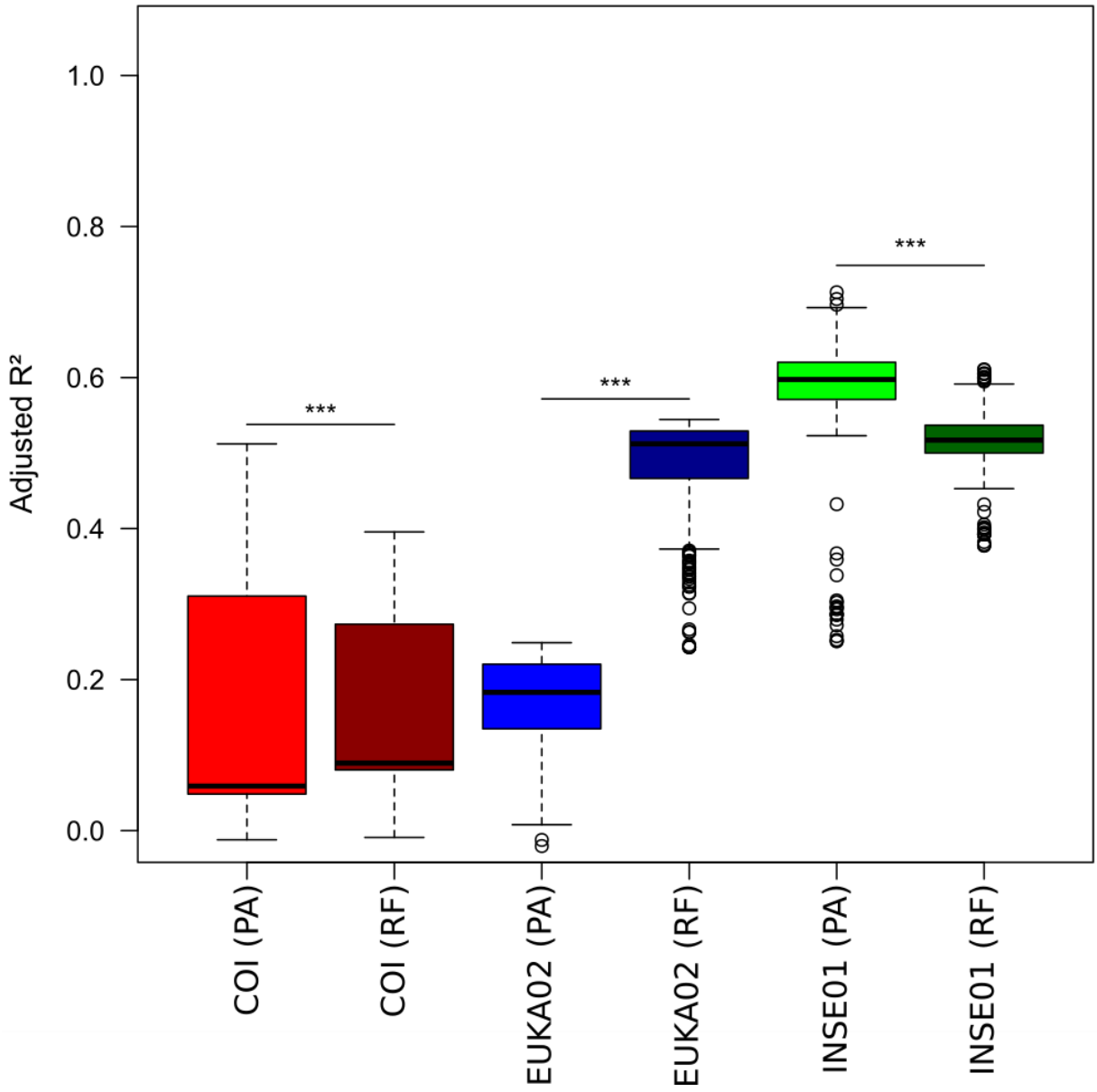
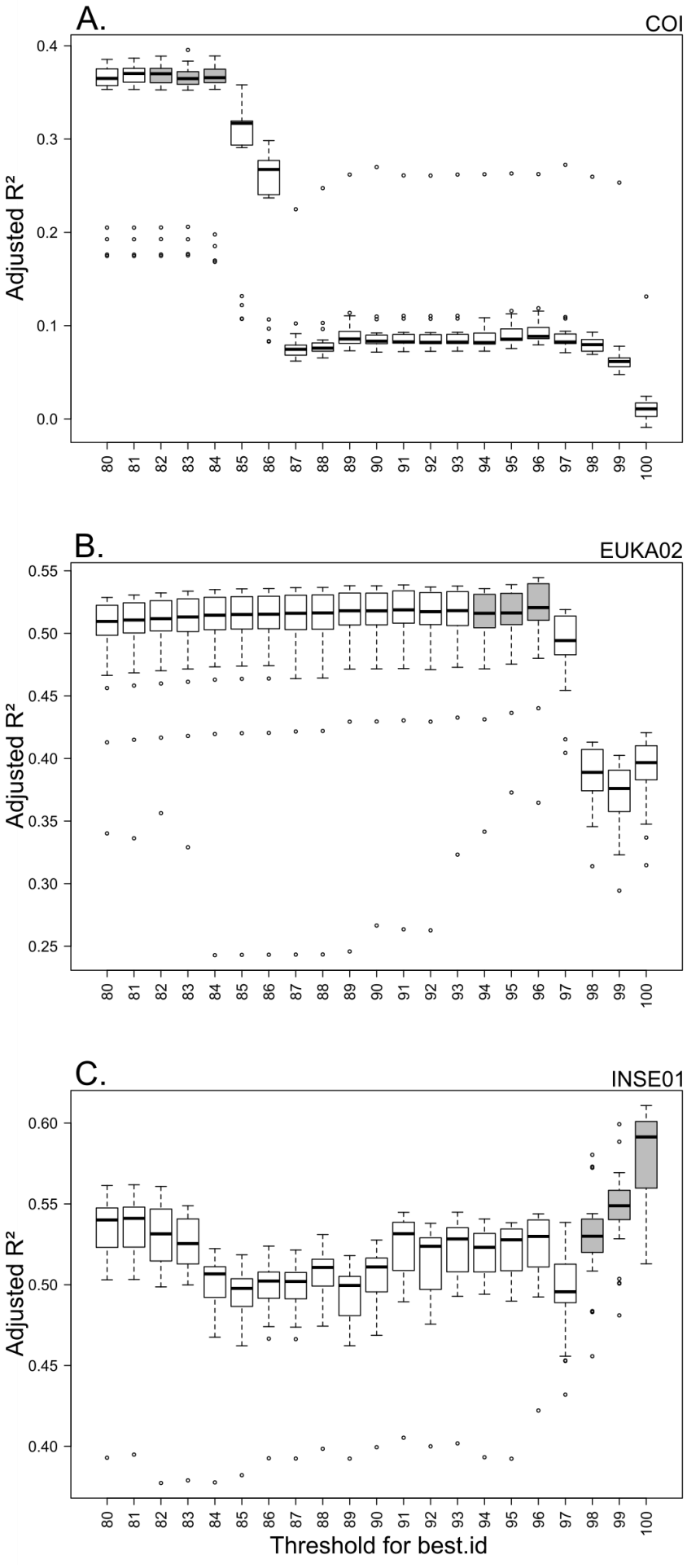


Fig. 3: Boxplots describing the distributions of adjusted R^2 values derived from linear regressions between I_2M_2 index values calculated on relative abundances (morphology-based approach) versus presence/absence (PA) or relative frequencies of reads (RF) for each primer set. N = 21 thresholds for best identity x 21 thresholds for minimal number of reads = 441 values per couple of "primer x data type". Statistical differences between PA vs RF, for a given primer, were investigated with Wilcoxon tests for paired data, with *** if $p < 0.001$. Each boxplot

represents minimum/Q25/median/Q75/maximum values, respectively. Outliers (open circles) are outside the 1.5 interquartile range of the corresponding adjusted R^2 value distribution.


A.
COI
Adjusted R²
0.0
0.1
0.2
0.3
0.4
80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100
B.
EUKA02
Adjusted R²
0.25
0.30
0.35
0.40
0.45
0.50
0.55
80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100
C.
INSE01
Adjusted R²
0.40
0.45
0.50
0.55
0.60
80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100
Threshold for best.id

1063

Fig. 4: Boxplots of the values of adjusted R^2 of linear regression between values of I_2M_2 based on morphology-based data and I_2M_2 calculated on bulk DNA for each primer set (A = COI, B = EUKA02, C = INSE01) and for different thresholds for best identity. Reads were expressed as relative frequencies. N = 21 for each boxplot. Each boxplot represents minimum/Q25/median /Q75/maximum values, respectively. Outliers (open circles) are outside the 1.5 interquartile range of the corresponding adjusted R^2 value distribution. In grey, selected thresholds for the following analyses.

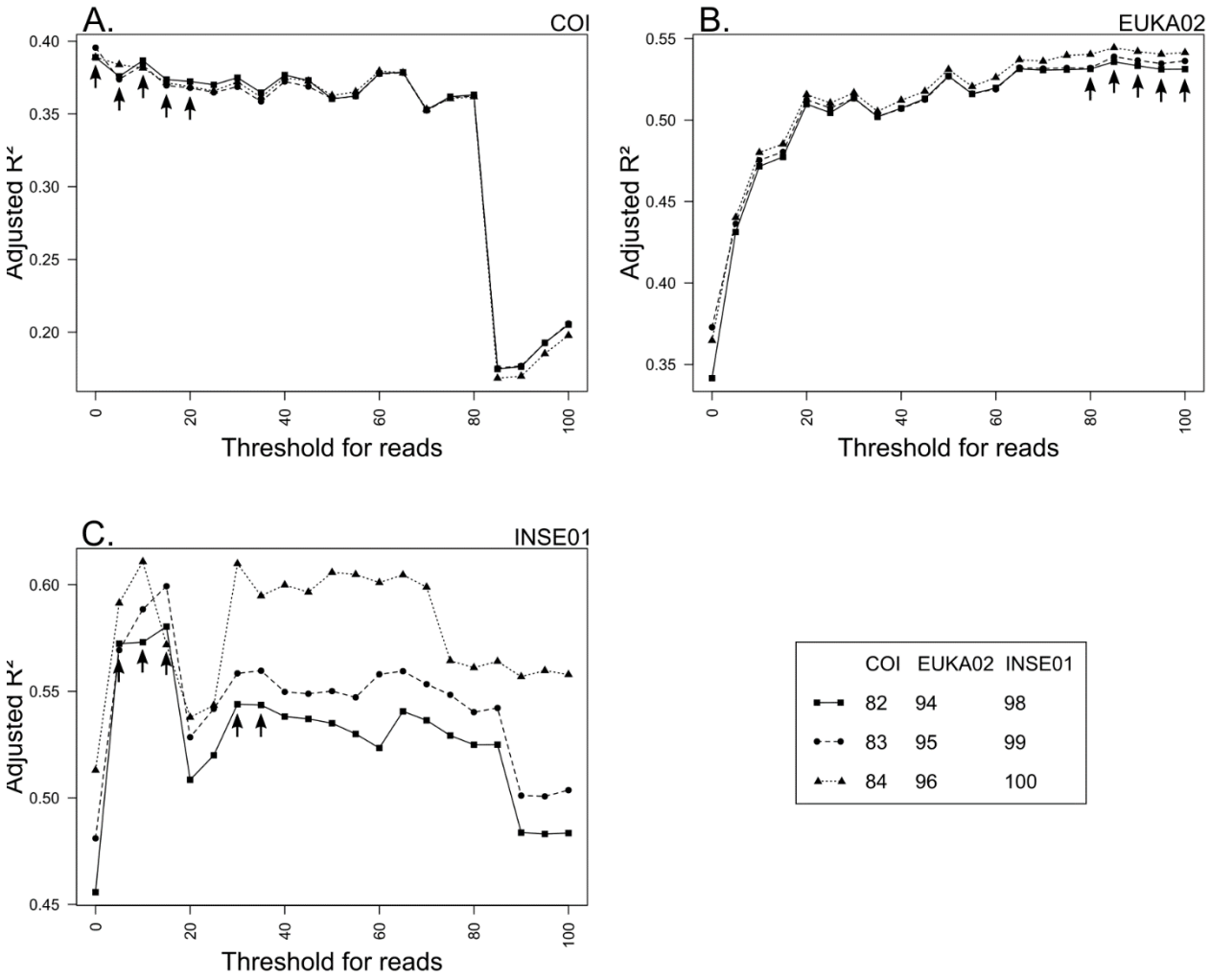


Fig. 5: Values of the adjusted R^2 of linear regressions between I_2M_2 calculated on morphology-based data and I_2M_2 values calculated on bulk DNA-based data for each primer set (reads are expressed in relative frequencies) according to different values of threshold for the minimal number of reads, and calculated for the three previously selected thresholds for best identity (cf. Fig. 4). Arrows indicate the five selected thresholds for minimal number of reads for the following analyses.

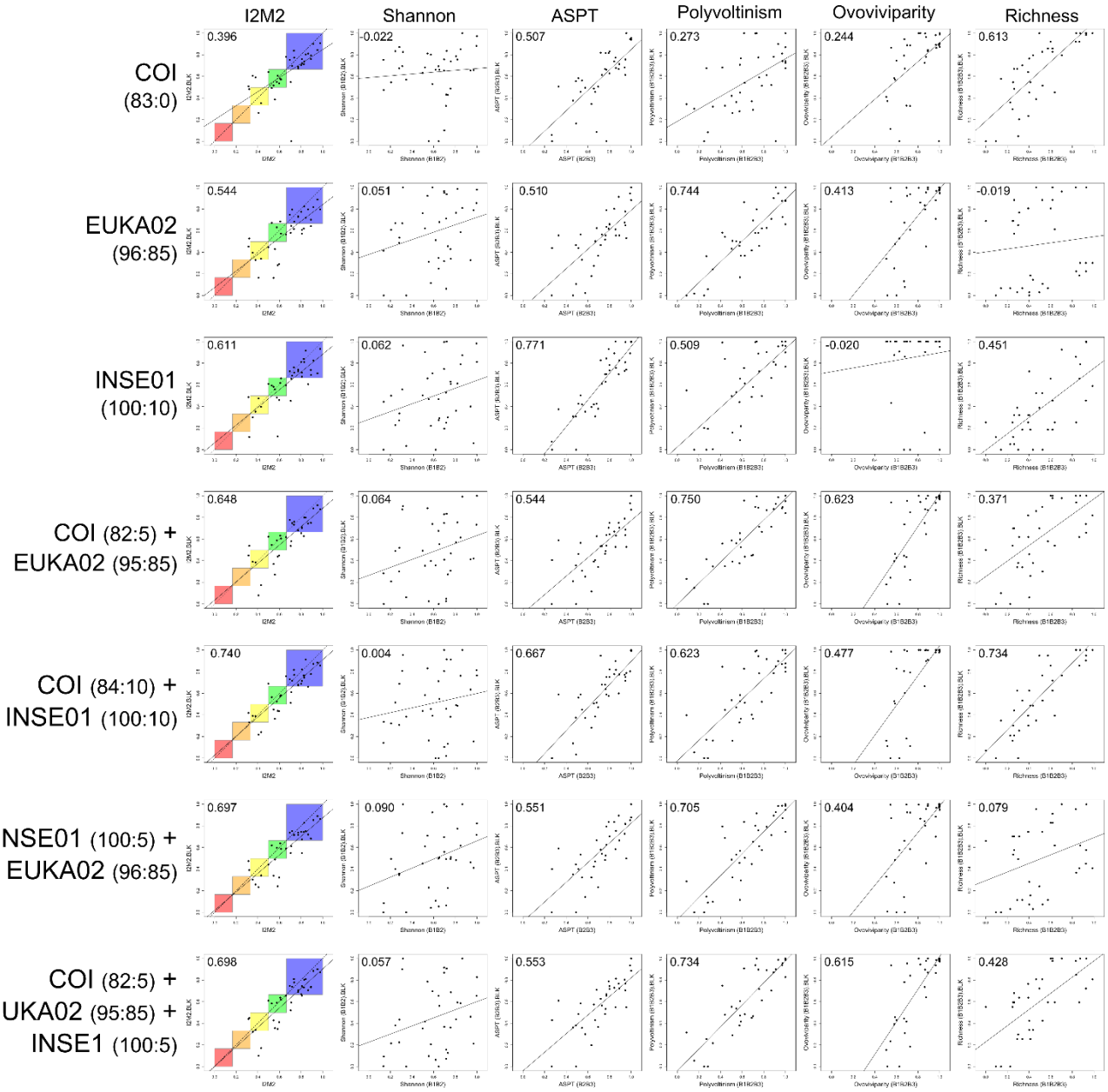


Fig. 6: Linear regressions (solid lines) between the values of the I_2M_2 index (and its associated metrics) calculated from the morphology-based versus bulk DNA-based data, for each primer set and combinations of primer sets. Results presented here are for the best combination (= maximum adjusted R^2) in terms of thresholds for best identity and minimum number of reads within the range of tested values. All metric values are expressed in Ecological Quality Ratios (EQR). Dashed lines (in right graphs) are lines with a slope equals to 1. *N = 36* (18 streams x 2 years). Top-left values are adjusted R^2.

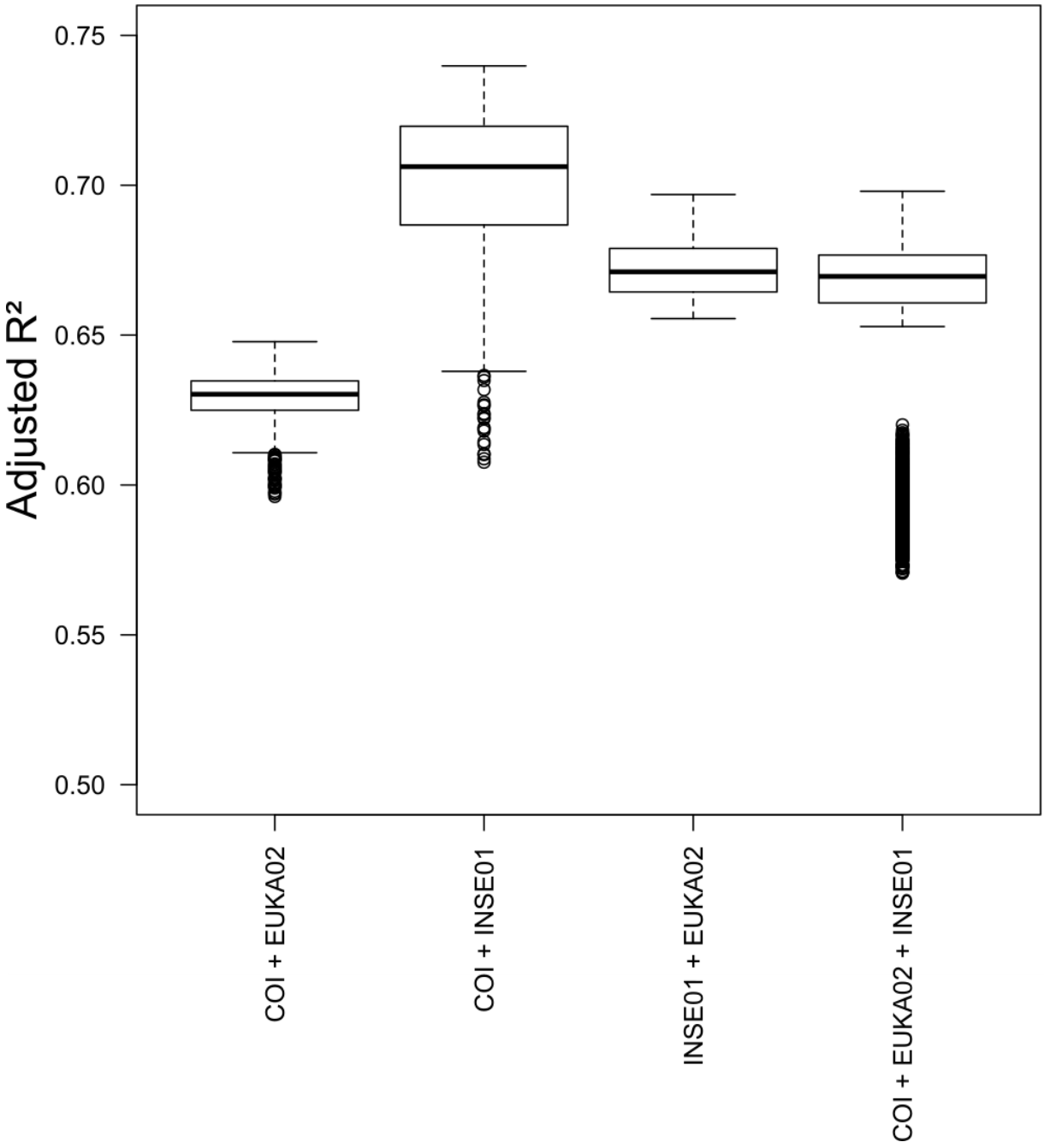


Fig. 7: Boxplots of the values of adjusted R^2 corresponding to linear regressions between I_2M_2 values calculated on morphology-based and bulk DNA-based data, for each combination of primer sets. Read numbers were first transformed in relative frequencies within data provided by each primer set separately, before being summed for each primer set combination and once again transformed in relative frequencies. *N = 225* for each pair of primer sets, and *N = 3375* for the combination of the three primer sets.

1094

Supplementary material

Tab. S1. Detailed location of the sampled sites. Stream types are according to Wasson et al. (2002). [a] this site was not considered as a reference for the definition of bulk-DNA based "BEST" values, as it was an acidic stream (pH = 6.4). Pressure intensity: "M" = moderate, "S" strong.

Code	Stream	Site	Longitude (WGS84)	Latitude (WGS84)	Stream type	Pressure category
1M	GELON	VILLARD	6.1215540	45.4770249	TP2	Hydrological alteration (M)
1R	CROP	VILLARD-BONNOT	5.9749229	45.2224414	TP2	Reference
1S	GELON	VERNEIL	6.1417991	45.4646720	TP2	Hydrological alteration (S)
2M	BOURBRE	ST-CLAIR	5.4574619	45.5892812	P5	Organic matter (M)
2R	SEMINE	BELLEYDOUX	5.7885934	46.2489425	P5	Reference
2S	BOURBRE	CESSIEU	5.3987061	45.5647726	P5	Organic matter (S)
3M	OIGNIN	BRION	5.5419970	46.1744351	M5	Organic matter (M)
3R	CHERAN	CARLET	6.1829477	45.6528586	M5	Reference
3S	OIGNIN	MATAFELON	5.5525989	46.2655298	M5	Hydrological alteration (S)
4M	RAVINE	RAON L'ETAPE	6.9292740	48.4087970	TP4	Acidification (M)
4R	PLAINE	RAON SUR PLAINE	7.0612097	48.4986129	P4	Reference
4S	GENTIL SAPIN	SENONES	7.0400186	48.4536782	TP4	Acidification (S)
5M	VEZOUZE	CIREY	6.9373024	48.5896737	P4	Organic matter (M)
5R	SAVOUREUSE	LEPUIX	6.8261808	47.7677172	P4	Reference
5S	VEZOUZE	BLAMONT	6.8384378	48.5884879	P4	Organic matter (S)
6M	CLEURIE	LA FORGE	6.7042109	48.0559488	P4	Toxic/Chemical (M)
6R	MENAURUPT	SAPOIS	6.7868727	48.0290126	P4	Reference[a]
6S	CLEURIE	LE THOLY	6.7252253	48.0891638	P4	Toxic/Chemical (S)