Received: 29 April 2020 | Revised: 15 September 2020 | Accepted: 9 October 2020

DOI: 10.1111/mec.15723

SPECIAL ISSUE ARTICLE

MOLECULAR ECOLOGY WILEY

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] | Gentile 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]

[1]Université de Lorraine, CNRS, LIEC, Metz, France

[2]Université Grenoble Alpes, CNRS, LECA, Laboratoire d'Ecologie Alpine, Grenoble, France

[3]SPYGEN, Le Bourget du Lac, France

[4]Department of Environmental Science and Policy, Università degli Studi di Milano, Milano, Italy

[5]Université de Strasbourg, CNRS, ENGEES, LIVE UMR 7362, Strasbourg, France

[6]University of Duisburg-Essen, Aquatic Ecosystem Research, Essen, Germany

[7]UiT – The Arctic University of Norway, Tromsø Museum, Tromsø, Norway

Correspondence
Philippe Usseglio-Polatera, Université de Lorraine, CNRS, LIEC, F-57000 Metz, France.
Email: philippe.usseglio-polatera@univ-lorraine.fr

Funding information
Agence Nationale de la Recherche, Grant/Award Number: ANR-13-ECOT-0002-01

Abstract

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 streams 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 bulk material of the same macroinvertebrate samples, using read numbers (expressed as relative frequencies) as a proxy for abundances. In particular, we 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. In total, 140 taxa were identified based on morphological criteria, and 127 were identified based on DNA metabarcoding using 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 us to improve 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.

KEYWORDS

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

1 | INTRODUCTION

Human activities have major negative impacts on freshwater ecosystems with drastic consequences for 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, namely 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, 2019; Terlizzi et al., 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 bulk material (Hering et al., 2018; Taberlet et al., 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 et al., 2011; Rimet et al., 2016), in order to obtain a list of taxa potentially present in the monitored ecosystem or in the bulk sample of the surveyed community.

Compared to traditional morphology-based methods, the metabarcoding approach is usually considered as noninvasive (e.g., when directly extracting DNA from water samples; Dejean et al., 2012; Valentini et al., 2009), and comparably inexpensive and rapid (Baird & Hajibabaei, 2012; Ji et al., 2013; Taberlet et al., 2012). It can also allow for better taxon identification than morphological expertise (Sweeney et al., 2011), and an overall better detection of all the species in aquatic ecosystems (Civade et al., 2016; Valentini et al., 2016). Moreover, metabarcoding also allows a sound estimate of beta diversity (Ji et al., 2013; Serrana et al., 2019; Sweeney et al., 2011; Yu et al., 2012), and is a reliable source of information for policy-making (Ji et al., 2013). Therefore, metabarcoding has been considered as a potential and credible alternative to morphology-based monitoring for both terrestrial and aquatic ecosystems (Baird & Hajibabaei, 2012; Elbrecht et al., 2017; Shaw et al., 2017). In freshwaters, several studies have highlighted the biomonitoring potential of metabarcoding, which could efficiently discriminate streams according to their ecological quality (Gibson et al., 2015; Hajibabaei et al., 2011; Ji et al., 2013; Kuntke et al., 2020; Mächler et al., 2019; Serrana et al., 2019; Sweeney et al., 2011; Zizka et al., 2020). For instance, environmental DNA (eDNA) information on eukaryotic communities in bottom sediments has been strongly associated with land-use pressure types (Xie et al., 2017), and macroinvertebrate bulk data have been used to infer key gradients of stream condition, including dissolved oxygen, dissolved organic carbon, total nitrogen and conductivity (Emilson et al., 2017).

The main objective of the present study was to test the ability of metabarcoding, applied to standardized bulk samples of benthic macroinvertebrates, to assess the ecological status of streams based on a large-scale biomonitoring programme performed within the WFD context. Using metabarcoding on bulk samples could allow us to bypass the organism identification step, which is time-consuming and a source of uncertainty due to operator misidentifications and interoperator identification variability (Metzeling et al., 2003). We also investigated how to improve the bioassesment ability of metabarcoding by testing:

1. 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 et al., 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 et al., 2014).
2. 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; Serrana et al., 2019; Yu et al., 2012; 97.5% in Carew et al., 2018; 98% in Lobo et al., 2017). Low-abundance (in terms of number of reads) MOTUs, for example 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).

The selection of genetic markers (e.g., mitochondrial vs. nuclear, single vs. multicopy, fast evolving vs. conserved, protein-coding vs. ribosomal), the respective primer sets as well as the threshold values

are important decisions to take in a metabarcoding experiment, but for which no clear guidelines exist (but see Bokulich et al., 2013). Such decisions may have a strong impact on the DNA-based bioassessment, which could be exacerbated by the high phylogenetic diversity of benthic invertebrates in a single sample. For example, taxonomic groups that are not as well amplified as other groups by a given primer set or marker could be filtered out of the data set due to their low numbers of reads. The amplification rate can indeed vary greatly among the major taxonomic groups in benthic invertebrate assemblages, even when accounting for biomass (Elbrecht et al., 2017).

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

2 | MATERIAL AND METHODS

2.1 | Sampling sites and data acquisition

Sampling was conducted at 18 sites on 15 streams belonging to five different stream types defined according to stream order (Strahler, 1957) and French hydroecoregions (Wasson et al., 2002, 2006) (Figure 1; see also Table S1). 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 the gathering of large amounts of information on chemical and hydromorphological 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 geographical origin (two different hydro-eco-regions × 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 et al., 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

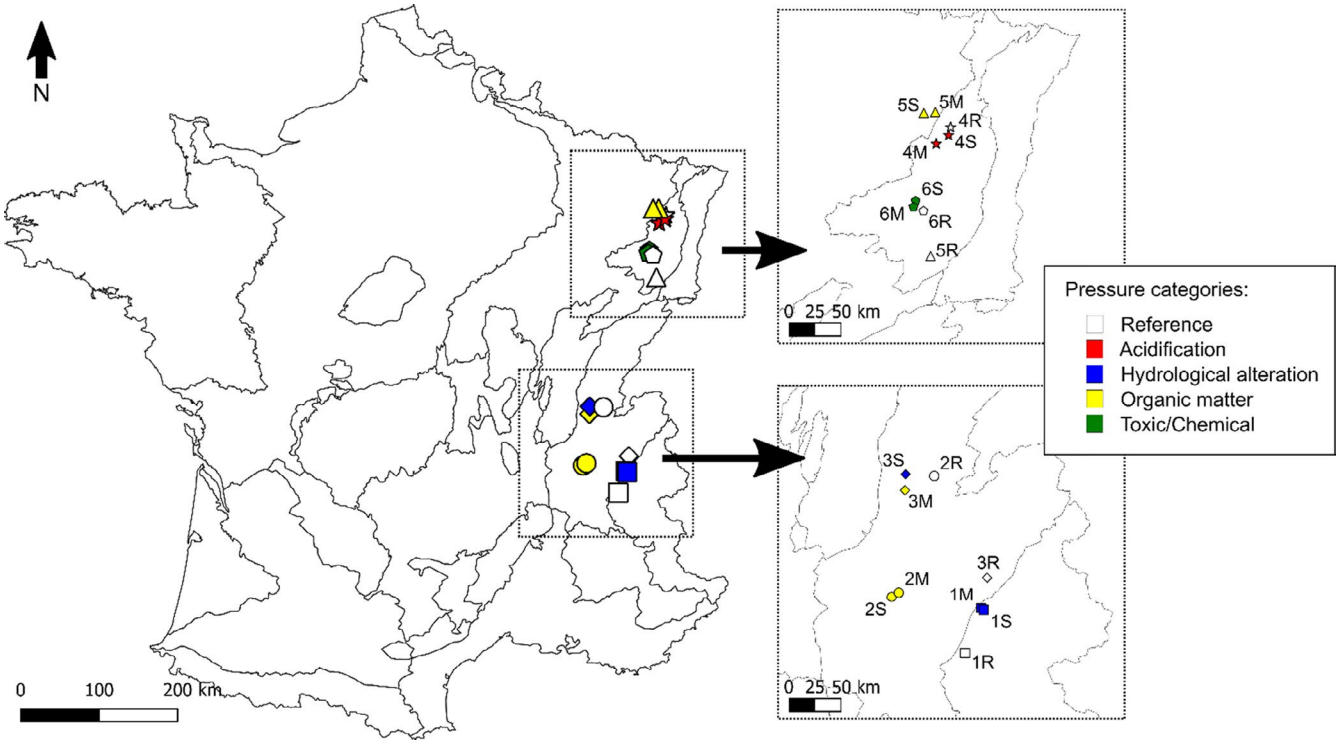

FIGURE 1 Location of the sampled sites. Delineations indicate the main French hydro-eco-regions (Wasson et al., 2002). The number in the code name of each sampled site gives information about the triplet of sites to which it belongs (from 1 to 6), whereas the letter gives information about its status, either LIRR ("Reference"; "R") or IRR, exhibiting either a Moderate "M" pressure intensity or a Strong "S" pressure intensity (see Materials and Methods for further details). All three sites found in a triplet belong to the same stream type according to the French typology of streams (Wasson et al., 2002). Further information about the sites is available in Table S1

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 (Figure 2a). Each habitat is characterized by its substrate type (among 12 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 m^2) was used to sample macroinvertebrates in each habitat. Samples were preserved in undenatured alcohol (~70% final concentration), for up to 1 year. In the 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

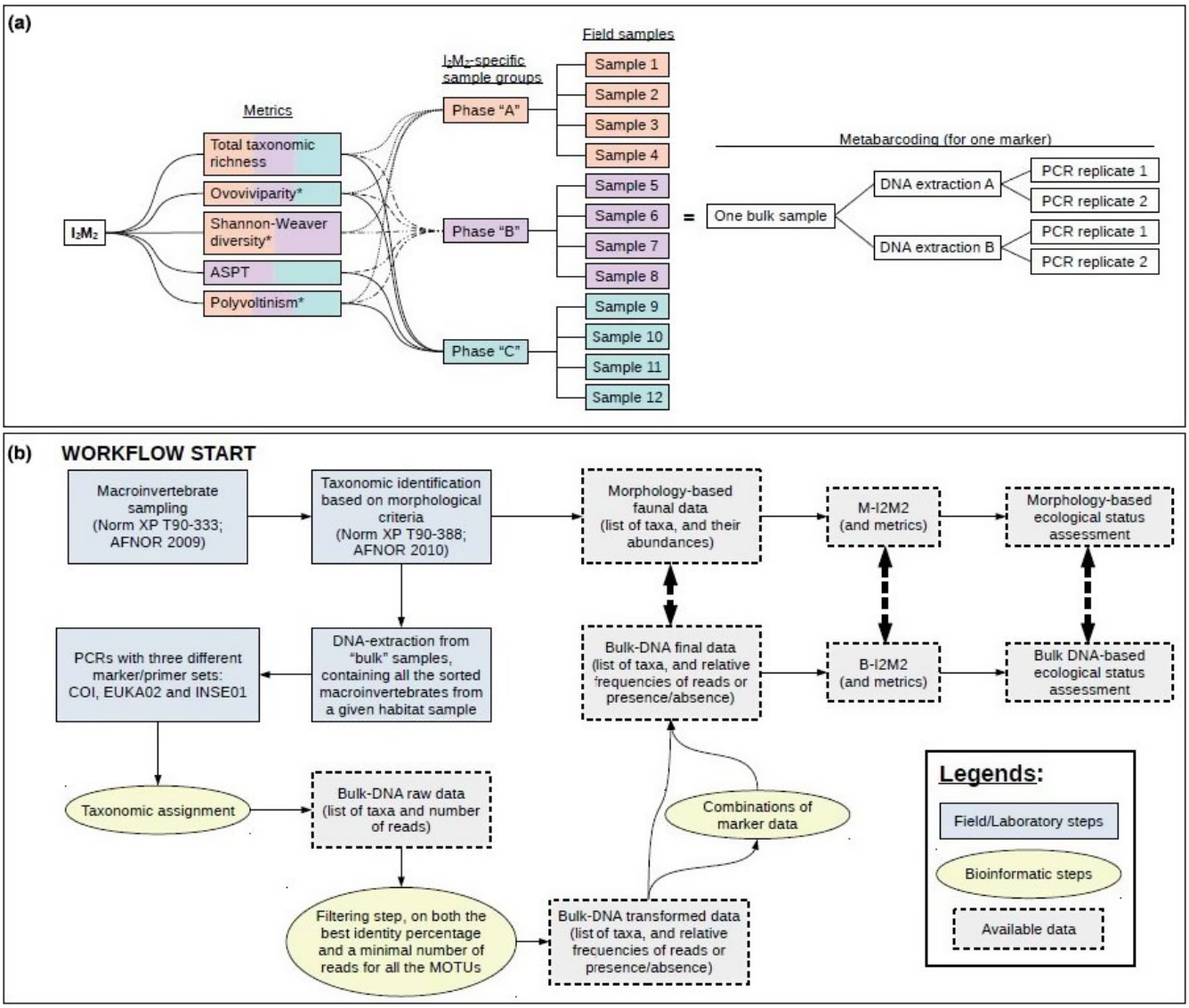

FIGURE 2 (a) Details about how samples of a given site sampling event were pooled for metric calculation based on the results obtained with both the morphology-based standard approach and the bulk-DNA based approach, and details about DNA extractions and PCR replications for each bulk sample. *Metrics which take into account taxon abundance (or reads, for bulk data) in their calculation. (b) Workflow. Dashed double-arrowed lines indicate comparisons of the results obtained with both the standard approach and the bulk-DNA based approach

or organisms altered during the sampling process), were identified at the best taxonomic level possible (i.e., 29 individuals from eight taxa over 89,157 individuals caught during the whole study). After taxonomic identification based on morphological criteria, all the sorted organisms, that is the "bulk samples" ($N = 432$), were stored in 95°C undenatured alcohol (for up to 1 year) and sent for metabarcoding.

2.2 | I_2M_2 index calculation

Faunal data (i.e., abundance per taxon × phase × site) were combined per site, as indicated in Figure 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 total taxonomic richness, Shannon–Weaver diversity index (Shannon & Weaver, 1963), average score per taxon (ASPT; Armitage et al., 1983), and the proportions of ovoviviparous (Ovoviviparity) and polyvoltine (Polyvoltinism) organisms in the invertebrate assemblage. The I_2M_2 index was constructed and calibrated over 10 chemical and seven hydromorphological pressure categories (see Mondy et al., 2012, for further details). One subindex per pressure category is first calculated as the weighted mean of the values of the five metrics, with the weights equal to the ability of each metric to discriminate between LIRRs and IRRs for this pressure category (quantified by its "discrimination efficiency" [DE]; Ofenböck et al., 2004). The final I_2M_2 index value is then calculated as the arithmetic mean of the 17 subindex values (Mondy et al., 2012).

2.3 | Metabarcoding

Bulk samples were homogenized using an IKA ULTRA-TURRAX Tube Drive control system with sterile 20-ml tubes and 10 steel beads (5 mm Ø) by grinding at 4000 rpm for 15 min (IKA). Complete samples were then incubated overnight at 56°C in 5 ml lysis buffer (Tris-HCl 0.1 m, EDTA 0.1 m, NaCl 0.01 m and N-lauroyl sarcosine 1%, pH 7.5–8.0). Extractions were then completed using the DNeasy Blood Tissue Kit (Qiagen), according to the manufacturer's instructions. DNA extracts were recovered in a total volume of 300 µl. Two DNA extractions were performed per bulk sample. Negative extractions without samples were systematically performed to monitor possible contamination.

Three primer pairs respectively corresponding to three different markers were used for each sample, Inse01 for a mitochondrial 16S rDNA region (Elbrecht et al., 2016; Taberlet et al., 2018), Euka02 for a nuclear 18S rDNA region (Guardiola et al., 2015) and BF1-BR2 for the cytochrome *c* oxidase I (COI) region (Elbrecht & Leese, 2017). 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 bovine serum albumin (BSA, Roche Diagnostic) and 0.5 µm of each primer for COI and 16S primers or 0.2 µm for 18S primers. Two PCR (polymerase chain reaction) 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′-labelled 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 (90 s 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 analysed in parallel to the samples to monitor possible contamination during the PCR step.

After PCR amplification, PCR products from the same marker were combined in equal volumes and purified using the MinElute (Qiagen) purification kit. Purified amplicons were checked by high-resolution capillary electrophoresis (QIAxcel System, Qiagen) and sent to Fasteris for library preparation and sequencing. Libraries were prepared according to the PCR-free MetaFast protocol (Taberlet et al., 2018; for further details, see also: 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) with a paired-end approach (2 × 125 bp), while the COI amplicons (two libraries) were sequenced on a MiSeq platform (Illumina) producing 2 × 250-bp paired-end reads.

2.4 | 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 aligning 10 bp of maximal quality on both read ends) were discarded. For the Euka02 primer, as the read length did not allow us 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 (Figure 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 intrasample distances (between PCR replicates of a given bulk sample) while maximizing the intersample 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 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 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), and then, after further sample pooling (Figure 2a), they were filtered based on their total number of reads (Figure 2b). For the filtering step based on best identity values, we tested several thresholds, between 80% and 100%. After this first filtering step, we standardized the MOTU 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, subfamily 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 to be used in the bioevaluation process were removed from the data set (e.g., MOTUs such as Metazoa, Neoptera or Holometabola). Figure S1 gives 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 ≤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, the results presented in this study do not include this optional filtering step.

According to the standard used for index calculation (XP T90-333; AFNOR, 2010; Figure 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; Figure 2a). A summary of the main identified MOTUs by each marker is available in Table S2. Taxa were then filtered according to their total number of reads in each pooled sample for each marker. We opted 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 data sets 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 (RFs) or in presence/absence (PA).

2.5 | 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. 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 data sets already filtered with the three best identity thresholds previously identified. As a result, we obtained 15 data sets (one per combination of best identity threshold × read minimal number threshold) for each individual marker, 225 data sets (15 × 15 combinations) for each pair of markers (e.g., [COI + Euka02]), and 3375 data sets (i.e., 15 × 15 × 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 in at least one of the two or three data sets 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 data sets.

2.6 | 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, 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 (EQRs; 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 data set (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 data set. 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 larger available data set (with 10,074 sampling events, vs. only 4132 sampling events in the original data set used in Mondy et al., 2012). Based on this new data set, more robust reference and DE values were (re)calculated, and combinations of phases used for individual metric calculation were revised to optimize the discrimination efficiency of metrics (see Figure 2a). In this study we used this revised version of the I_2M_2 (Usseglio-Polatera et al., 2016).

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). Two other ways of calculating R^2 values, based on "perfect match" ($y = x$) regressions, were also explored, but discarded (for further details, see Figure S2 and related comments in the Supporting Information).

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 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).

3 | RESULTS

3.1 | Taxonomic identification

In total, 140 morphotaxa were identified in the 36 site sampling events (Figure 3). The bulk data filtered with the lowest threshold for best identity (80%) allowed us 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 based on morphological criteria were not recovered by at least one of the markers. Of these 41 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 nine taxa for Inse01.

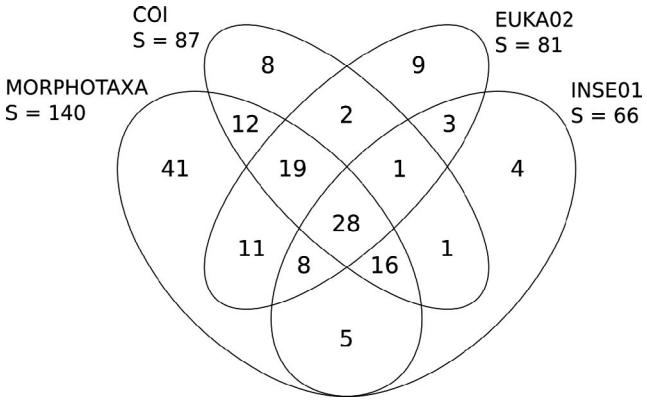

FIGURE 3 Venn diagram showing overlaps among the taxonomic lists provided by the standard approach (140 morphotaxa) and by each primer set. Comparisons were made for each primer set at the lowest threshold for both the best identity (80%) and the number of reads (1). S = number of different taxa

3.2 | Relative frequencies versus 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; $p < .001$ for both primer sets; Figure 4a). For Inse01, adjusted R^2 values were not significantly different between RF and PA data (Wilcoxon signed rank test for paired data; $p > .05$). Nevertheless, the primer set Inse01 exhibited the highest adjusted R^2 values with PA data, with a maximum of .738. The maximum adjusted R^2 values for the primer sets COI and Euka02 were respectively .524 and .671, with RF data. Based on these preliminary results, we chose to focus on data expressed in RFs for the remaining analyses.

3.3 | Threshold selection for markers

We used adjusted R^2 to examine 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; $p < .001$ for each primer set; Figure 5). For COI, the thresholds ≤84% of identity provided the best values of adjusted R^2 (medians > .50; Figure 5a). Adjusted R^2 values decreased with increasing identity thresholds from 85% to 100%, albeit with a small increase of R^2 for identity between 96% and 99%. For Euka02, the adjusted R^2 increased slowly with the threshold values until 96% of identity, with all the medians above .61 (Figure 5b), while decreased rapidly (median values close to .49) at very high values of identity (97%–100%). For Inse01, the adjusted R^2 median values globally ranged between .58 and .62 for thresholds between 80% and 97% of identity. The highest R^2 values were observed with identity values of

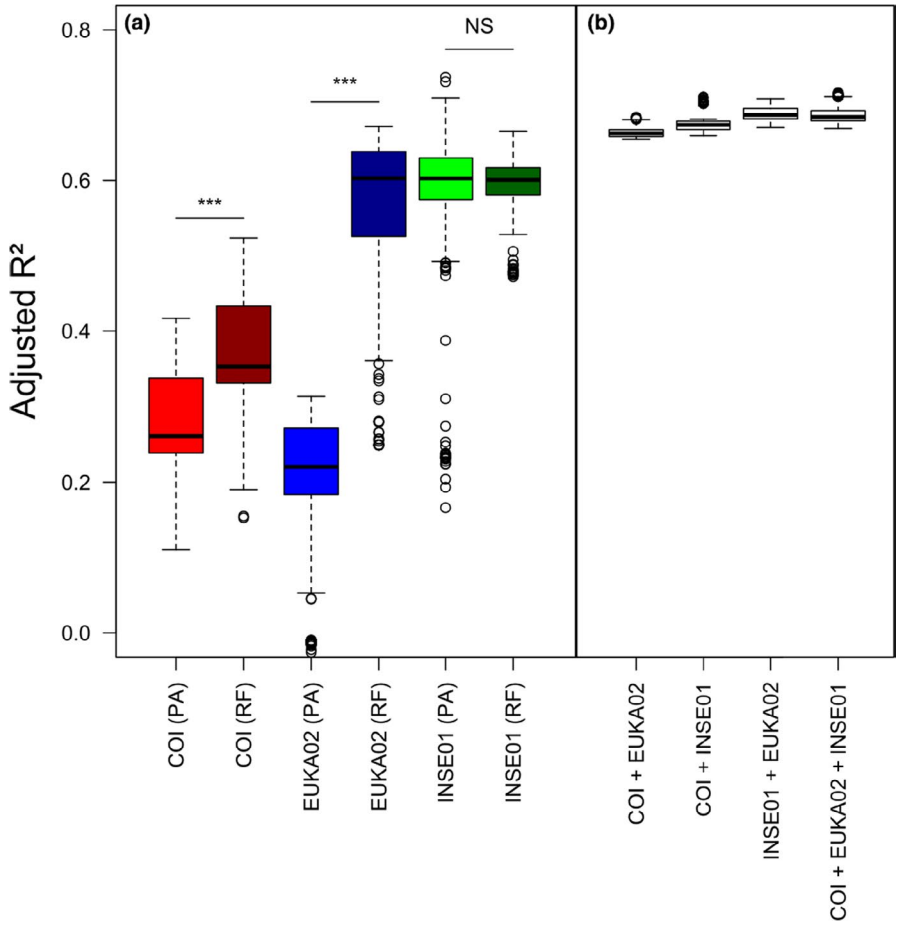

FIGURE 4 (a) Boxplots describing the distributions of the adjusted R^2 values derived from the linear regressions between the I_2M_2 index values calculated based on the relative abundances of taxa (standard approach) versus the I_2M_2 index values calculated based on the bulk-DNA data: (a) based on the presence/absence (PA) or the relative frequencies of reads (RF) of taxa provided by each marker. $N = 21$ thresholds for best identity × 21 thresholds for minimal number of reads = 441 values per "marker × data type" pair (i.e., boxplot). Statistical differences between PA and RF, for a given marker, were investigated with Wilcoxon tests for paired data, with *** if $p < .001$ and NS if $p > .05$ (not significant). Each boxplot represents the 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. (b) Based on each combination of markers. Read numbers were first transformed to relative frequencies within data provided by each marker separately, before being averaged, for each taxon, over the number of combined markers. $N = 225$ for each pair of markers, and $N = 3375$ for the combination of the three markers. Each boxplot represents the 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

98%–100% (Figure 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 (Figure 5).

The relationships between the adjusted R^2 values 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 (Figure 6). The adjusted R^2 varied significantly according to the tested thresholds for the minimal number of reads (Friedman rank sum tests; $p < .001$ for each primer set). For COI, adjusted R^2 was >.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 was between 15 and 45 (Figure 6a). For Euka02, adjusted R^2 increased with increasing minimal number of reads until a plateau at about 65 reads (Figure 6b). For both markers, the relationships between adjusted R^2 and thresholds for the minimum number of reads were similar for all the tested thresholds for best identity. The patterns of response of Inse01 were similar for the 98% and 99% thresholds for best identity (Figure 6c), but differed from that obtained with a threshold of 100% identity. The highest adjusted R^2 values were observed for all the identity thresholds using thresholds of between 50 and 70 reads. Based on these results, the threshold values 5, 60, 65, 70 and 100 for COI, 80, 85, 90, 95 and 100 for Euka02, and 50, 55, 60, 65 and 70 for Inse01 were selected (Figure 6).

3.4 | Biomonitoring efficiency of the bulk-DNA approach

After having identified, for each marker, the best mode of data expression (relative frequencies) and the best ranges of identity

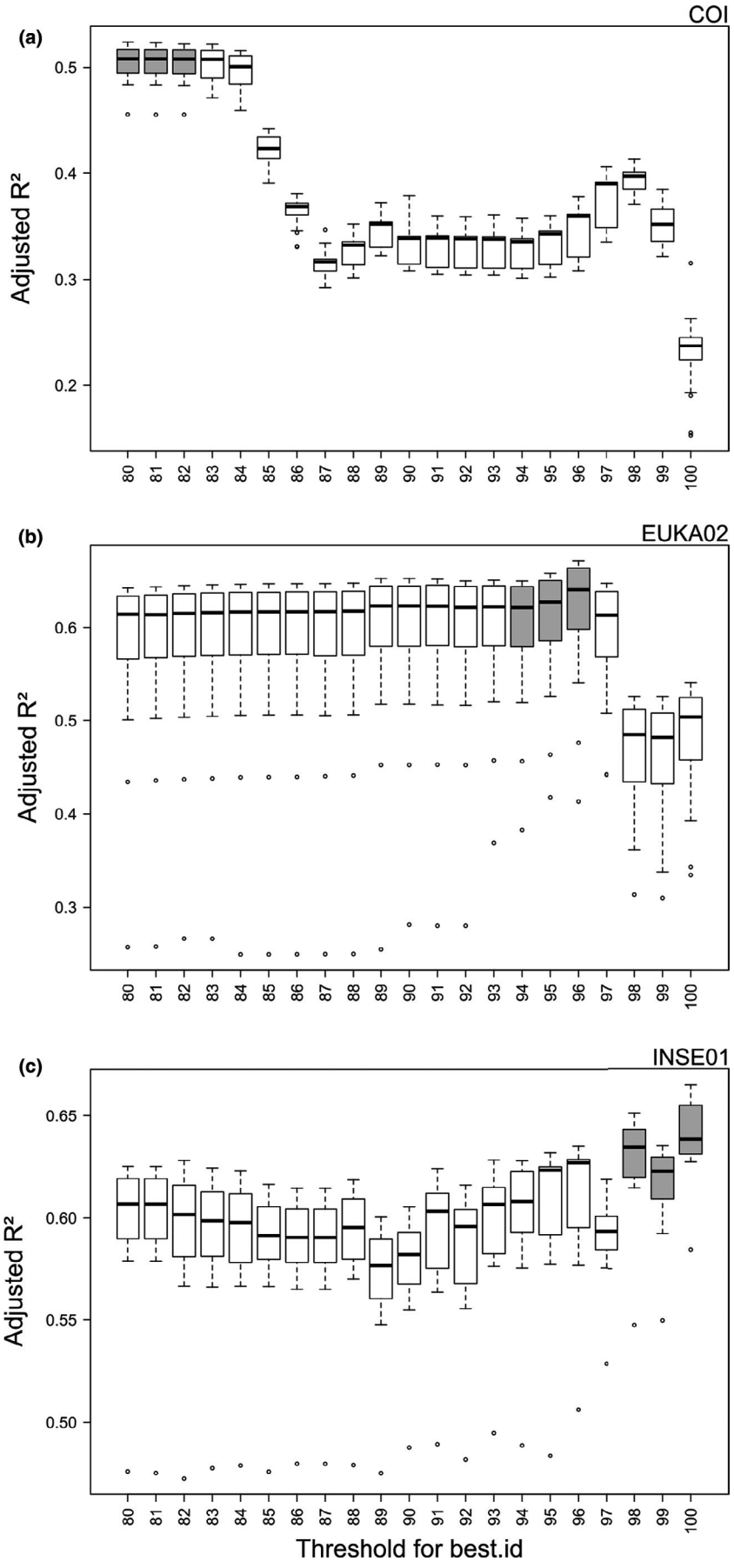

FIGURE 5 Boxplots of the adjusted R^2 values provided by the linear regressions between the values of the standard I_2M_2 and the values of the bulk-DNA based I_2M_2, for each marker (a = COI, b = Euka02, c = Inse01) and for different thresholds for best identity. Reads were expressed as relative frequencies. $N = 21$ for each boxplot. Each boxplot represents the 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. Boxplots of the thresholds selected for the following analyses are in grey

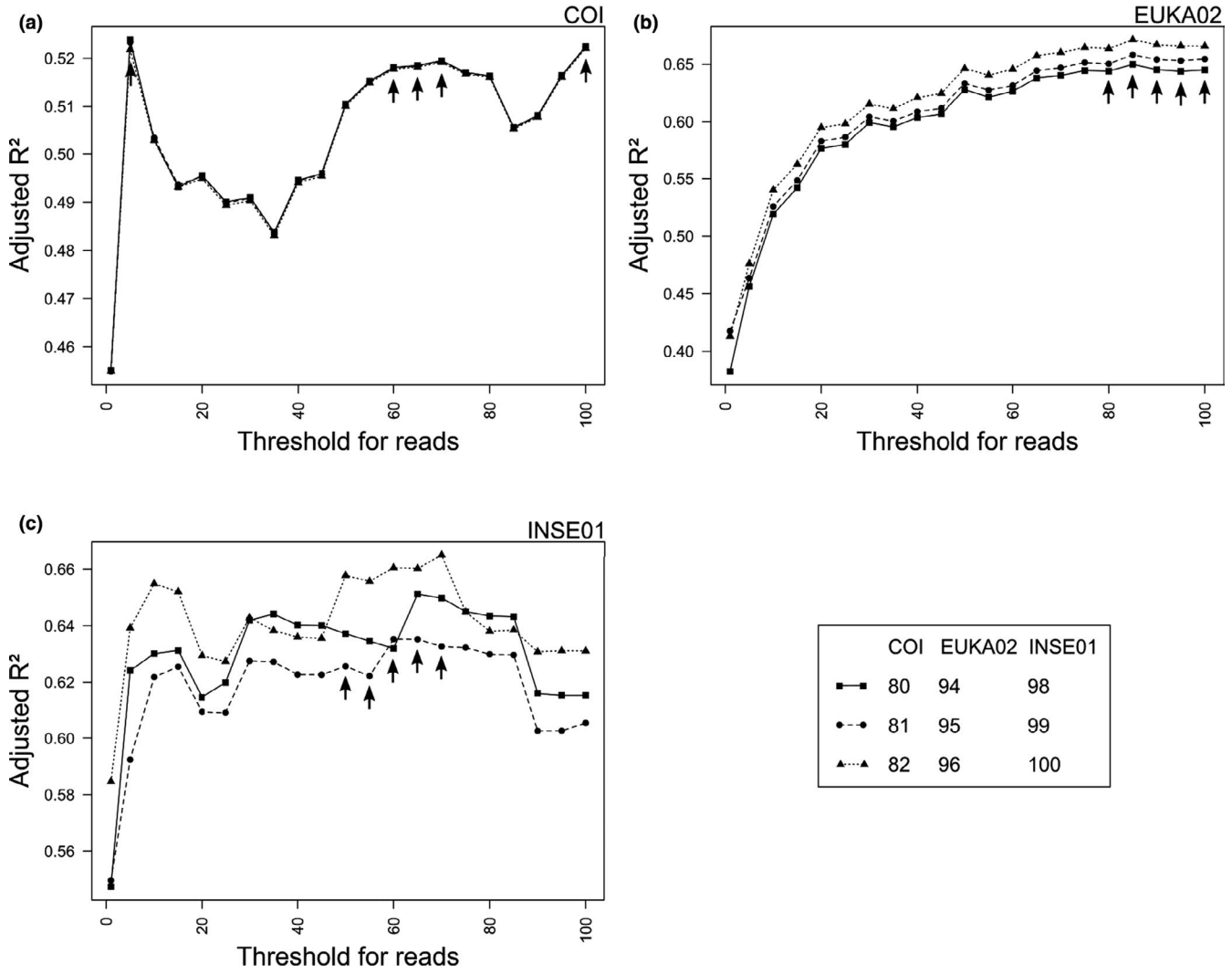

FIGURE 6 Values of the adjusted R^2 of the linear regressions between the values of the standard I_2M_2 and the values of the bulk-DNA based I_2M_2 for each marker according to different values of threshold for the minimal number of reads, and calculated for the three previously selected thresholds for best identity (cf. Figure 5). Reads were expressed in relative frequencies for the calculation of bulk-DNA based I_2M_2. Arrows indicate the five selected thresholds for the minimal number of reads for the following analyses

threshold and minimal number of reads, we analysed the global bioassessment efficiency of each individual marker and each combination of markers. We compared the values of the B-I_2M_2, its related metrics and the allocated ecological quality classes with those provided by the standard approach. We also compared the discrimination efficiency of the B-I_2M_2 calculated on each tested combination.

3.4.1 | Biomonitoring efficiency of individual markers

The marker Euka02 provided the best regression between M-I_2M_2 and B-I_2M_2 (Figure 7a) with a best adjusted R^2 of .671 for the combination of thresholds "96%:85" (96% best identity and a minimum of 85 reads, hereafter abbreviated min.r), among the 15 combinations of selected thresholds for Euka02. The best adjusted R^2 values were .524 and .665 for COI (80%:5 min.r) and Inse01 (100%:70 min.r), respectively and among the 15 combinations 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; $p < .05$; Table 1). The ecological quality classes allocated to sites based on both approaches differed also in many cases: 5/36, 16/36 and 18/36 times for Inse01, Euka02 and COI respectively (Table 1). Both COI and Euka02 often allocated a worse ecological status to sites compared with the morphology-based approach, while Inse01 tended to provide a good match between ecological classes assigned with both approaches (Table 1). The discrimination efficiency of the I_2M_2 index values provided by the bulk-DNA approach with COI (DE = 0.875; Table 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 for ASPT and the relative frequency of polyvoltine organisms in the invertebrate assemblage (Polyvoltinism

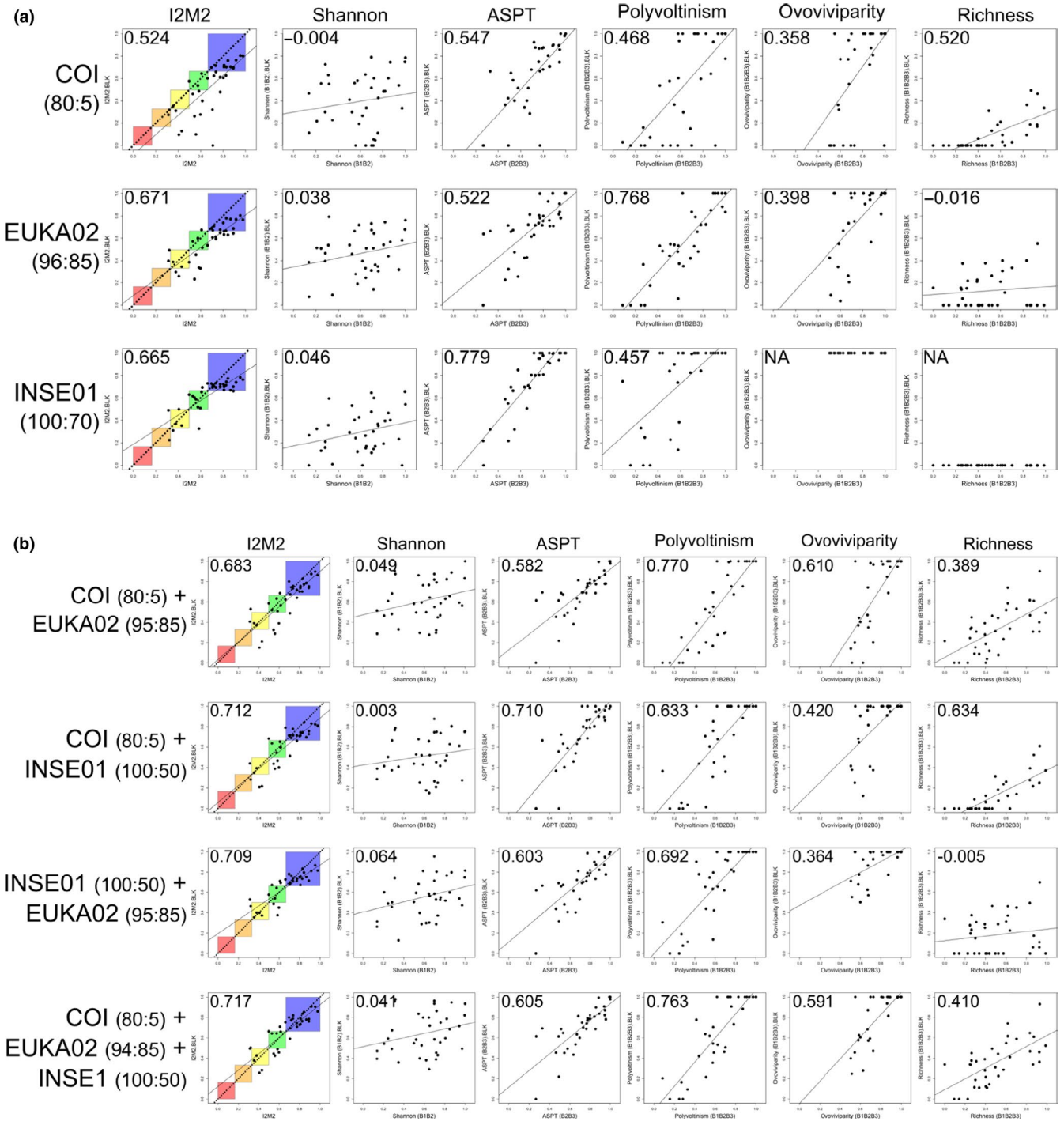

FIGURE 7 Linear regressions (solid lines) between the values of the I_2M_2 index (and its associated metrics) calculated from the standard versus the bulk-DNA based data, for each marker (a) and combination of markers (b). Results presented are obtained for the best combination (= maximum adjusted R^2) of thresholds for best identity (first value in parentheses) and minimum number of reads (second value), within the range of tested values. All the metric values are expressed in ecological quality ratios (EQRs). Dashed lines (in right-hand graphs) are lines with a slope equal to 1. $N = 36$ (18 streams × 2 years). Top-left values are adjusted R^2

in Figure 7) were correctly modelled from bulk-DNA data for all the markers. Adjusted R^2 ranged from .522 (Euka02) to .779 (Inse01) for ASPT, and between .457 (Inse01) and .768 (Euka02) for Polyvoltinism. The relative frequency of ovoviviparous organisms (Ovoviviparity in Figure 7) was correctly modelled by Euka02 and COI, with adjusted R^2 equal to .398 and .358, respectively, but not by Inse01. Richness was correctly modelled by COI (adj-R^2 = .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 −.004 (COI) to .046 (Inse01).

TABLE 1 Summary of the results corresponding to the best match between the standard and the bulk-DNA approaches, obtained with each marker and each combination of markers

				Misclassifications							
			Wilcoxon test	Worse predicted state				No change	Better predicted state		Total
Bulk data	Adjusted R^2	DE	(p-value)	−4	−3	−2	−1	0	+1	+2	
COI (80:5)	.524	0.875	.000	1	2	3	12	18			18/36
EUKA02 (96:85)	.671	0.750	.000			1	13	20	2		16/36
INSE01 (100:70)	.665	0.750	.011			1	2	31	2		5/36
COI (80:5) + EUKA02 (95:85)	.683	0.958	.004			2	5	27	1	1	9/36
COI (80:5) + INSE01 (100:50)	.712	0.833	.002			1	6	28	1		8/36
INSE01 (100:50) + EUKA02 (95:85)	.709	0.917	.181				1	31	4		5/36
COI (80:5) + EUKA02 (94:85) + INSE01 (100:50)	.717	0.958	.432				5	27	3	1	9/36

Note: Wilcoxon signed rank tests for paired data and adjusted R^2 (from linear regressions) are applied/calculated between the values of the I_2M_2 index obtained with the standard and the bulk-DNA approaches. The deviation between the quality class allocated by both methods, expressed in number of classes, is negative if the bulk-DNA-based approach provides a worse evaluation than the standard approach and positive in the opposite case. The thresholds for the best identity and for the minimum number of reads corresponding to the best match are respectively provided in parentheses for each primer set or combination of markers. DE = discrimination efficiency (higher is better).

3.4.2 | 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 (Figures 4b and 7b). The best adjusted R^2 between M-I_2M_2 and B-I_2M_2 increased from .671 (Euka02) to .717 when we combined all marker information (Figures 4b and 7b). DE values of marker combinations ranged from 0.833 (for [COI + Inse01]) to 0.958 (for [COI + Euka02] and [COI + Inse01 + Euka02]) (Table 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 (Figure 7). The adjusted R^2 of total richness and the relative frequency of ovoviviparous organisms in the assemblage greatly increased from .520 (COI) to .634 [COI + Inse01], and from .398 (Euka02) to .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 .046 (Inse01) to .064 [Inse01 + Euka02] and from .768 (Euka02) to .770 [COI + Euka02], respectively. The adjusted R^2 of the ASPT index decreased for combinations of primer sets from .779 (Inse01) to .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 (Table 1). Globally, the combinations of markers allocated a rather worse ecological status than the standard approach (Table 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 $p < .01$; Table 1), but not for [Inse01 + Euka02] and for the combinations of the three markers (both $p > .05$; Table 1).

4 | DISCUSSION

4.1 | 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 et al. (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 et al. (2020) have suggested

that macroinvertebrate presence/absence data can lead to similar bioassesment results when compared to abundance-based data.

The close 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 have confirmed that bulk data would readily allow efficient retrieval of 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).

4.2 | Combining data from multiple markers

Our results also confirmed that combining data from different markers improves the appraisal of biodiversity based on bulk samples (Clarke et al., 2014; Hajibabaei et al., 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 streams 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., Ephemeroptera, Plecoptera and Trichoptera; 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 indeed contributes highly 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).

4.3 | 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 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, 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).

4.4 | 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 et al. (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 EQRs, 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.

4.5 | On individual metrics

Combining data obtained with at least two different markers has allowed us 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, total taxonomic richness, and relative frequencies of 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 to 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 when applying both approaches, even if calculated on only two-thirds of the available data on the sampled invertebrate assemblage (both the Shannon–Weaver index and the ASPT are calculated on eight sample units over 12, corresponding to only two of the three phases; see Figure 2b and Mondy et al., 2012, for further details).

As in previous studies (Hajibabaei et al., 2012; Serrana et al., 2019), morphotaxa not recovered with metabarcoding were often "rare" taxa, that is taxa with fewer than 10 individuals over the whole sampling design (observed for 29 out of the 41 missing taxa). Among the five individual metrics combined in the I_2M_2, the Shannon–Weaver index and relative frequencies of polyvoltine and ovoviviparous invertebrates within the assemblage are the only metrics that take into account taxon abundances. Missing the rarest taxa seems to have only a limited impact on the calculated values of the trait-based metrics, but probably not on those of the Shannon–Weaver diversity index.

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

4.6 | 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 data set (values currently in use were defined on a data set containing more than 10,000 site sampling events), whereas "BEST" and "WORST" values could only have been defined on our modest bulk data set. 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, in terms of both correlation and ecological status assessment.

Similarly, the I_2M_2 index is calculated as the arithmetic average of 17 subindices (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 DE of individual metrics has been calculated at the national scale. Here, the sampling design was too small to allow 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.

4.7 | 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 et al., 2014; Beentjes et al., 2019; Carew et al., 2013, 2018; Elbrecht et al., 2017; Emilson et al., 2017; Gibson et al., 2014, 2015; Hajibabaei et al., 2011, 2019; Kuntke et al., 2020; Serrana et al., 2019; Zizka et al., 2020). However, more work is needed before implementing bulk-metabarcoding in the routine monitoring of streams. Indeed, we have highlighted a series of biases leading to the reclassification of several site sampling events in terms of ecological quality class. Solutions exist for reducing these biases, and we have focused on several of them in the previous sections. For instance, a reference DNA barcoding database, including 578 different taxa (62% identified at the species level) has been specifically built for this study (Ficetola et al., 2020), in order to work with a database as complete as possible, as recommended by several authors (Aylagas et al., 2014; Hering et al., 2018). Several studies (Gibson et al., 2015; Ji et al., 2013; Mächler et al., 2019; Serrana et al., 2019; Sweeney et al., 2011) have also suggested that new indices could be constructed specifically on bulk-DNA information for stream bioassessment based on macroinvertebrate assemblages, as has already been done for benthic diatoms (e.g., Vasselon et al., 2017). On the one hand, even if retrieved MOTUs are not assigned to taxa, MOTUs can still be used to efficiently discriminate between impaired and reference situations (e.g., Emilson et al., 2017). However, this approach would need a huge preliminary sampling phase in order to construct a new index based on a database that would include the majority of the MOTUs that could be recovered, for instance, for any stream found at the national scale. On the other hand, when molecular methods prove to be mature enough, they could be used to obtain

species-level taxonomic lists. Based on these lists, the observed and reference values of taxonomy-based and trait-based metrics could be refined to further improve the discrimination efficiency of DNA-based indices. Other promising sources of DNA for stream bioassessment are the ethanol used for sample/invertebrate preservation (e.g., Hajibabaei et al., 2012; Martins et al., 2020; Zizka et al., 2019) or even DNA directly extracted from unsorted samples (Pereira-da-Conceicoa et al., 2019). Both sources would avoid the time-consuming steps of invertebrate sorting and counting. Supervised machine learning is another promising approach, as it could be used to directly model a standard biotic index value based on metabarcoding data (Cordier et al., 2019; Frühe et al., 2020).

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

ACKNOWLEDGMENTS

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

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 the 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. Analysed the 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 contributions from all the authors.

DATA AVAILABILITY STATEMENT

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

ORCID

Albin Meyer https://orcid.org/0000-0002-0419-205X
Alice Valentini https://orcid.org/0000-0001-5829-5479
Aurélie Bonin https://orcid.org/0000-0001-7800-8609
Gentile Francesco Ficetola https://orcid.org/0000-0003-3414-5155

REFERENCES

AFNOR (2009). *Qualité de l'eau – Prélèvement des macro-invertébrés aquatiques en rivières peu profondes*. XP T90–333.

AFNOR (2010). *Qualité de l'eau – Traitement au laboratoire d'échantillons contenant des macro-invertébrés de cours d'eau*. XP T90–388.

Armitage, P., Moss, D., Wright, J., & Furse, M. (1983). The performance of a new biological water quality score system based on macroinvertebrates over a wide range of unpolluted running-water sites. *Water Research*, *17*(3), 333–347. https://doi.org/10.1016/0043-1354(83)90188-4

Aylagas, E., Borja, Á., Muxika, I., & Rodríguez-Ezpeleta, N. (2018). Adapting metabarcoding-based benthic biomonitoring into routine marine ecological status assessment networks. *Ecological Indicators*, *95*, 194–202. https://doi.org/10.1016/j.ecolind.2018.07.044

Aylagas, E., Borja, Á., & Rodríguez-Ezpeleta, N. (2014). Environmental status assessment using DNA Metabarcoding: Towards a genetics based marine biotic index (gAMBI). *PLoS One*, *9*(3), e90529. https://doi.org/10.1371/journal.pone.0090529

Baird, D. J., & Hajibabaei, M. (2012). Biomonitoring 2.0: A new paradigm in ecosystem assessment made possible by next-generation DNA sequencing. *Molecular Ecology*, *21*(8), 2039–2044. https://doi.org/10.1111/j.1365-294X.2012.05519.x

Baird, D. J., Pascoe, T. J., Zhou, X., & Hajibabaei, M. (2011). Building freshwater macroinvertebrate DNA-barcode libraries from reference collection material: Formalin preservation vs specimen age. *Journal of the North American Benthological Society*, *30*(1), 125–130. https://doi.org/10.1899/10-013.1

Beentjes, K. K., Speksnijder, A. G. C. L., Schilthuizen, M., Hoogeveen, M., Pastoor, R., & van der Hoorn, B. B. (2019). Increased performance of DNA metabarcoding of macroinvertebrates by taxonomic sorting. *PLoS One*, *14*(12), e0226527. https://doi.org/10.1371/journal.pone.0226527

Bennett, C., Owen, R., Birk, S., Buffagni, A., Erba, S., Mengin, N., Murray-Bligh, J., Ofenböck, G., Pardo, I., van de Bund, W., & Wagner, F. Wasson, J.-G. (2011). Bringing European river quality into line: An exercise to intercalibrate macro-invertebrate classification methods. *Hydrobiologia*, *667*(1), 31–48. https://doi.org/10.1007/s10750-011-0635-2

Bokulich, N. A., Subramanian, S., Faith, J. J., Gevers, D., Gordon, J. I., Knight, R., Mills, D. A., & Caporaso, J. G. (2013). Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nature Methods*, *10*(1), 57–59. https://doi.org/10.1038/nmeth.2276

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., Taberlet, P., & 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

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

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

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), 1–20. https://doi.org/10.18637/jss.v022.i04

Dudgeon, D., Arthington, A. H., Gessner, M. O., Kawabata, Z.-I., Knowler, D. J., Lévêque, C., Naiman, R. J., Prieur-Richard, A.-H., Soto, D., & Stiassny, M. L. J., & Sullivan, C. A. (2006). Freshwater biodiversity: Importance, threats, status and conservation challenges. *Biological Reviews*, *81*(2), 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, *5*, 1–11. 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*. 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., Hallwachs, W., & 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

Guardiola, M., Uriz, M. J., Taberlet, P., Coissac, E., Wangensteen, O. S., & Turon, X. (2015). Deep-sea, deep-sequencing: Metabarcoding extracellular DNA from sediments of marine canyons. *PLoS One*, *10*(10), e0139633. https://doi.org/10.1371/journal.pone.0139633

Hajibabaei, M., Porter, T. M., Robinson, C. V., Baird, D. J., Shokralla, S., & Wright, M. T. G. (2019). Watered-down biodiversity? A comparison of metabarcoding results from DNA extracted from matched water and bulk tissue biomonitoring samples. *PLoS One*, *14*(12), e0225409. https://doi.org/10.1371/journal.pone.0225409

Hajibabaei, M., Shokralla, S., Zhou, X., Singer, G. A. C., & Baird, D. J. (2011). Environmental barcoding: A next-generation sequencing approach for biomonitoring applications using river benthos. *PLoS One*, *6*(4), e17497. https://doi.org/10.1371/journal.pone.0017497

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. https://doi.org/10.1186/1472-6785-12-28

Hering, D., Borja, A., Jones, J. I., Pont, D., Boets, P., Bouchez, A., Bruce, K., Drakare, S., Hänfling, B., Kahlert, M., Leese, F., Meissner, K., Mergen, P., Reyjol, Y., Segurado, P., Vogler, A., & Kelly, M. (2018). Implementation options for DNA-based identification into ecological status assessment under the European Water Framework Directive. *Water Research*, *138*, 192–205. https://doi.org/10.1016/j.watres.2018.03.003

Hutchings, P. (2019). An advocate for taxonomic research in Australia. *Pacific Conservation Biology*, *25*(1), 34. https://doi.org/10.1071/PC17033

Ji, Y., Ashton, L., Pedley, S. M., Edwards, D. P., Tang, Y., Nakamura, A., Kitching, R., Dolman, P. M., Woodcock, P., Edwards, F. A., Larsen, T. H., Hsu, W. W., Benedick, S., Hamer, K. C., Wilcove, D. S., Bruce, C., Wang, X., Levi, T., Lott, M., … Yu, D. W. (2013). Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding. *Ecology Letters*, *16*(10), 1245–1257. https://doi.org/10.1111/ele.12162

Kuntke, F., de Jonge, N., Hesselsøe, M., & Lund Nielsen, J. (2020). Stream water quality assessment by metabarcoding of invertebrates. *Ecological Indicators*, *111*, 105982. https://doi.org/10.1016/j.ecolind.2019.105982

Larras, F., Coulaud, R., Gautreau, E., Billoir, E., Rosebery, J., & Usseglio-Polatera, P. (2017). Assessing anthropogenic pressures on streams:

A random forest approach based on benthic diatom communities. *Science of the Total Environment*, *586*, 1101–1112. https://doi.org/10.1016/j.scitotenv.2017.02.096

Lobo, J., Shokralla, S., Costa, M. H., Hajibabaei, M., & Costa, F. O. (2017). DNA metabarcoding for high-throughput monitoring of estuarine macrobenthic communities. *Scientific Reports*, *7*(1), 15618. https://doi.org/10.1038/s41598-017-15823-6

Mächler, E., Little, C. J., Wüthrich, R., Alther, R., Fronhofer, E. A., Gounand, I., Harvey, E., Hürlemann, S., Walser, J.-C., & Altermatt, F. (2019). Assessing different components of diversity across a river network using eDNA. *Environmental DNA*, *1*(3), 290–301. https://doi.org/10.1002/edn3.33

Martins, F. M. S., Porto, M., Feio, M. J., Egeter, B., Bonin, A., Serra, S. R. Q., Taberlet, P., & Beja, P. (2020). Modelling technical and biological biases in macroinvertebrate community assessment from bulk preservative using multiple metabarcoding markers. *Molecular Ecology*, https://doi.org/10.1111/mec.15620

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 I. D. Hering, P. F. M. Verdonschot, O. Moog, & L. Sandin (Eds.), *Integrated assessment of running waters in Europe* (pp. 251–268). Springer. 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. https://www.R-project.org/

Rimet, F., Chaumeil, P., Keck, F., Kermarrec, L., Vasselon, V., Kahlert, M., Franc, A. & 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.

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

Usseglio-Polatera, P., Larras, F., Coulaud, R., & Billoir, E. (2016). *Finalisation I_2M_2, chantier Diagnostic multi-EQB et développement d'un indice pour les grands cours d'eau* (p. 147). Scientific Report, Lorraine University.

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., Bellemain, E., Besnard, A., Coissac, E., Boyer, F., Gaboriaud, C., Jean, P., Poulet, N., Roset, N., Copp, G. H., Geniez, P., Pont, D., Argillier, C., Baudoin, J.-M., … 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., Glidden, S., Bunn, S. E., Sullivan, C. A., Liermann, C. R., & 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

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Meyer A, Boyer F, Valentini A, et al. 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. *Mol Ecol*. 2020;00:1–18. https://doi.org/10.1111/mec.15723