Metabarcodes for the monitoring of freshwater benthic biodiversity through environmental DNA

Gentile Francesco Ficetola[1], Aurélie Bonin[2], Frédéric Boyer[2], Alice Valentini[3], Albin Meyer[4], Tony Dejean[5], Coline Gaboriaud[3], Philippe Usseglio-Polatera[4], and Pierre Taberlet[6]

[1]Università degli Studi di Milano
[2]Universite Grenoble Alpes
[3]SPYGEN
[4]Université de Lorraine Bibliothèque Ingénieurs Enim
[5]Spygen
[6]Université Joseph Fourier

April 28, 2020

Abstract

Environmental DNA and metabarcoding have great potential for the biomonitoring of freshwater environments. However, successful application of metabarcoding to biodiversity monitoring requires universal primers with high taxonomic coverage that amplify highly-variable, short metabarcodes with high taxonomic resolution. Moreover, reliable and extensive reference databases are essential to match the outcome of metabarcoding analyses with the available taxonomy and biomonitoring indices. Benthic invertebrates, particularly insects, are key taxa for freshwater biomonitoring. Nevertheless, so far, no formal comparison has assessed primers for metabarcoding of freshwater macrobenthos. Here we combined in vitro and in silico analyses to test the performance of metabarcoding primers amplifying regions in the 18S rDNA (Euka02 metabarcode), 16S rDNA (Inse01), and COI (BF1_BR2-COI) genes, and developed an extensive database of benthic invertebrates of France and Europe, with a special focus on three key insect orders (Ephemeroptera, Plecoptera and Trichoptera). In vitro analyses on 1514 individuals, belonging to 578 different taxonomic units showed very different amplification rates across primer combinations. The Euka02 marker showed the highest universality, while the Inse01 marker showed excellent performance for the amplification of insects. The BF1_BR2-COI metabarcode showed the highest resolution, while the resolution of Euka02 was often limited. By combining in vitro data with GenBank information, we developed a curated database including sequences representing 822 genera. The heterogeneous performance of the different metabarcodes highlights the complexity of the identification of the best markers, and advocates for the integration of multiple metabarcodes for a more comprehensive and accurate understanding of ecological impacts on freshwater biodiversity.

Metabarcodes for the monitoring of freshwater benthic biodiversity through environmental DNA

Gentile Francesco Ficetola[1,2], Frédéric Boyer[1], Alice Valentini[3], Aurélie Bonin[1,2], Albin Meyer[4], Tony Dejean[3], Coline Gaboriaud[3], Philippe Usseglio-Polatera[4], Pierre Taberlet[1,5]

[1] Univ. Grenoble Alpes, CNRS, Univ. Savoie Mont Blanc, LECA, Laboratoire d'Ecologie Alpine, F-38000 Grenoble

[2] Department of Environmental Sciences and Policy; University of Milano, Milano, Italia.

[3] SPYGEN, Le Bourget du Lac, France

[4] Université de Lorraine, CNRS, LIEC, F-57000 Metz, France

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

Abstract

Environmental DNA and metabarcoding have great potential for the biomonitoring of freshwater environments. However, successful application of metabarcoding to biodiversity monitoring requires universal primers with high taxonomic coverage that amplify highly-variable, short metabarcodes with high taxonomic resolution. Moreover, reliable and extensive reference databases are essential to match the outcome of metabarcoding analyses with the available taxonomy and biomonitoring indices. Benthic invertebrates, particularly insects, are key taxa for freshwater biomonitoring. Nevertheless, so far, no formal comparison has assessed primers for metabarcoding of freshwater macrobenthos. Here we combined *in vitro* and *in silico*analyses to test the performance of metabarcoding primers amplifying regions in the 18S rDNA (Euka02 metabarcode), 16S rDNA (Inse01), and COI (BF1_BR2-COI) genes, and developed an extensive database of benthic invertebrates of France and Europe, with a special focus on three key insect orders (Ephemeroptera, Plecoptera and Trichoptera). *In vitro* analyses on 1514 individuals, belonging to 578 different taxonomic units showed very different amplification rates across primer combinations. The Euka02 marker showed the highest universality, while the Inse01 marker showed excellent performance for the amplification of insects. The BF1_BR2-COI metabarcode showed the highest resolution, while the resolution of Euka02 was often limited. By combining *in vitro* data with GenBank information, we developed a curated database including sequences representing 822 genera. The heterogeneous performance of the different metabarcodes highlights the complexity of the identification of the best markers, and advocates for the integration of multiple metabarcodes for a more comprehensive and accurate understanding of ecological impacts on freshwater biodiversity.

Keywords: freshwater biodiversity; biomonitoring; biotic indices; DNA metabarcoding; primer bias; invertebrates; cytochrome c oxidase I; amplification rate; universality; taxonomic resolution.

1 INTRODUCTION

Freshwater environments are essential providers of clean water and other services for human society. They also host a substantial biodiversity, still they are globally subjected to the joint impact of multiple stressors such as pollution, eutrophication, climate change and hydrological and hydromorphological modifications (Noges *et al.*2016; Iversen *et al.* 2019). As a consequence, numerous regulations have been adopted at both the national and international level for the protection of water resources, such as the European Water Framework Directive (Directive 2000/60/EC) and the Clean Water Act of the US Environmental Protection Agency (33 U.S.C. §§1251-1387 1972) (Pawlowski *et al.* 2018). These regulations generally require the monitoring of freshwater environments through a combination of physicochemical, hydrological, and biotic parameters, to obtain prompt measurements of water quality and of the ecological status of ecosystems.

Multiple approaches exist to assess freshwater quality using aquatic organisms. Benthic invertebrates are perhaps the most frequently used biological group in aquatic bioassessment (Birk *et al.* 2012), because they are (i) taxonomically, biologically and functionally diverse (Usseglio-Polatera *et al.* 2000; Usseglio-Polatera*et al.* 2001), (ii) rather easy to identify at the genus or family levels (Tachet *et al.* 2010), (iii) often sedentary and reacting rapidly to anthropogenic pressures in all types of freshwater bodies (Hering *et al.* 2006b; Archaimbault *et al.* 2010; Hering*et al.* 2013), and (iv) their occurrence integrates the effects of environmental changes over several months (Floury *et al.* 2013). Invertebrate assemblages are thus a tool of choice to assess the ecological status of water bodies *(e.g.* Marzin *et al.*2012; Hering *et al.* 2013; Mondy & Usseglio-Polatera 2013) and to demonstrate environmental degradation (Miler *et al.* 2013; Mondy & Usseglio-Polatera 2013; Theodoropoulos *et al.* 2020) or restoration (Arce *et al.* 2014; Kupilas *et al.* 2016; Camargo 2017; Carlson *et al.* 2018).

Generally, bioassessment indices relying on benthic communities are based on the standardized collection of invertebrate assemblages from monitored sites, followed by organism sorting and taxonomic identification using morphological criteria. Then, quality scores can be attributed on the basis of the presence and/or

abundance of certain taxa (Friberg *et al.* 2006; Birk *et al.* 2012). As morphological identification is often challenging, in many case protocols do not require species-level identification, and identification at the genus or family level (and, in some cases, even at coarser levels) can be enough for the calculation of many biotic indices evaluating the ecological status of rivers (Bailey *et al.* 2001; Chessman *et al.*2007; Birk *et al.* 2012). Nevertheless, the morphological identification of hundreds collected specimens, including young, small-sized, larval stages and organisms damaged during sampling, remains time-consuming and requires a substantial taxonomic expertise, increasing the cost and time required for in-depth assessment of water quality (Haase *et al.* 2004; Hering *et al.* 2018).

Environmental DNA (eDNA) and metabarcoding are revolutionizing the monitoring of biodiversity at all levels, because they circumvent the challenge of morphological identification and allow the efficient detection of many taxa that are difficult to capture and detect using traditional methods (Taberlet *et al.* 2018). eDNA and metabarcoding are therefore extremely promising for the assessment of freshwater communities (Hering *et al.* 2018). DNA can be extracted from the tissue of pooled invertebrate communities, amplified using universal primers, sequenced, and identified on the basis of reference databases (Baird & Hajibabaei 2012; Yu *et al.* 2012; Andújar*et al.* 2018). This approach uses the same starting material than traditional biomonitoring, but allows skipping the complexity of morphology-based taxonomy (Baird & Hajibabaei 2012). Alternatively, DNA can be obtained directly from the water (Ficetola *et al.* 2008). Environmental DNA extracted from freshwaters allows the detection of many taxa that are difficult to capture and detect using traditional methods, but also poses new challenges compared to metabarcoding performed on the tissues of captured individuals. When in aquatic environments, DNA undergoes rapid degradation (Eichmiller *et al.*2016; Buxton *et al.* 2017); therefore eDNA is generally characterized by small fragment sizes (Jo *et al.* 2017; Bylemans*et al.* 2018), but see also (Sigsgaard *et al.* 2017). This generally precludes the use of "standard" barcode primers, which often amplify long DNA fragments (e.g. >300 bp in the most frequently used COI markers; Andújar *et al.* 2018). Furthermore, highly degenerated primers increase the risk of non-specific amplification, thus this kind of primers is not really suitable for the amplification of the complex mix of DNA extracted from the environment. As a consequence, the monitoring of benthic invertebrates using eDNA requires the development and assessment of primers with appropriate features.

Besides the length of the amplified region, three main characteristics are essential for satisfactory eDNA metabarcodes. First, the eDNA amplification rate generally decreases with the number of mismatches between target fragments and primers. Primers must therefore be designed in order to have a consistently low number of mismatches within sequences of the target group (high universality or taxonomic coverage; Ficetola *et al.* 2010; Piñol *et al.* 2015; Marquina *et al.* 2019). Taxonomic coverage can be assessed through both *in silico* and *in vitro* analyses. *In silico* analyses can allow the rapid assessment of all the taxa for which information is publicly available in databases, but *in vitro* tests are still needed to confirm the conditions under which primers work in the real world. Second, the amplified region must be highly variable, to ensure the identification of amplified organisms at the desired taxonomic level (high resolution; Ficetola *et al.* 2010; Tang *et al.* 2012; Marquina *et al.* 2019). Finally, extensive databases are essential if we want to assign the amplified sequences to known taxa. Even though attempts have been made for the assessment of environmental quality without a taxonomic assignment of DNA fragments (Ji *et al.* 2013; Apothéloz-Perret-Gentil *et al.* 2017), taxonomic assignment is essential if we want to produce data comparable with traditional indices of water quality, or if we want to combine eDNA data with information obtained through traditional methods (*e.g.* to analyse long-term series of water body surveys). Despite several attempts to assess freshwater quality using eDNA (Hering *et al.* 2018; Serrana*et al.* 2019; Czechowski *et al.* 2020; Pont *et al.*2020; Yang & Zhang 2020), so far no formal comparison has been performed among short primers suitable for eDNA metabarcoding of freshwater macrobenthos. In addition, there is a pressing need of exhaustive reference databases for taxonomic assignment.

In this study we combined *in vitro* and *in silico* analyses to compare the performance of three primer pairs potentially suitable for the analysis of eDNA from freshwater invertebrates (macrobenthos), and we developed an extensive reference database for benthic invertebrates living in European freshwaters. We mostly focused on three insect orders (Ephemeroptera, Plecoptera and Trichoptera), which are among the most frequently

used invertebrates for the bioassessment of streams (e.g. Brabec *et al.* 2004; Hering *et al.* 2006a; Gabriels *et al.* 2010; Arman *et al.* 2019; but see also Cox*et al.* 2019). We also considered a broad range of organisms belonging to other orders of insects and to other classes. We first produced the metabarcodes on the broadest available number of taxa from France, and then combined metabarcodes obtained *in vitro* with sequences available in public database, to obtain extensive and reliable measures of metabarcode performance, and to produce a extensive reference database for the monitoring of freshwaters through eDNA.

2 MATERIAL AND METHODS

We used the standardized database of European freshwater organisms (Schmidt-Kloiber & Hering 2015; download on 01 March 2018) as taxonomic reference for our analyses, considering all the benthic macroinvertebrates. Although in some cases this database considers non-monophyletic groups (*e.g.* Crustacea), it provides an exhaustive checklist of benthic macroinvertebrates that serve as an essential basis for monitoring bioassessment.

2.1 *In vitro* analyses of reference specimens

Most of the reference specimens were provided by OPIE-Benthos which is a working group of OPIE (Office Pour les Insectes et leur Environnement) especially dedicated to aquatic insect studies and aquatic ecosystem protection in France. OPIE-Benthos has developed a national inventory and reference collection of aquatic insects, including Ephemeroptera, Plecoptera, Trichoptera, and more recently aquatic Coleoptera, aquatic and semi-aquatic Heteroptera, aquatic larval stages of Megaloptera, Nevroptera and Diptera (Ptychopteridae) (http://www.opie-benthos.fr/opie/insecte.php). Corresponding organisms, identified at the highest possible level (species, if possible) by experienced taxonomists, were provided in triplicates (*i.e.* three specimens per taxon). The collection was completed by additional taxa (*e.g.* non-insect taxa) specifically sampled by the authors for this reference database.

Specimens were stored in 99% ethanol before DNA extraction. Total DNA was extracted from the entire organism. Samples (constituted of one specimen) were initially incubated overnight at 56 degC in 0.5 ml of 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 GmbH, Hilden, Germany), according to the manufacturer's instructions. DNA extracts were recovered in a total volume of 300 µl of elution buffer. Negative extractions without specimens were systematically performed to monitor possible contaminations. Three DNA amplifications were carried out for each sample using the following primer pairs: Inse01, amplifying a ˜155 bp region of the 16S mitochondrial rDNA (Taberlet*et al.* 2018); Euka02, amplifying a ˜123 bp region of the 18S rDNA (Guardiola *et al.* 2015; Taberlet *et al.*2018); and the BF1 and BR2 primers, which amplify a ˜316 bp region of the cytochrome c oxidase I (Elbrecht & Leese 2017). Inse01 has been developed mostly to amplify insects, Euka02 to amplify all eukaryotes, while BF1 and BR2 were designed to amplify freshwater macroinvertebrates (Elbrecht & Leese 2017; Taberlet *et al.*2018). DNA amplifications were performed in a final volume of 20 µL, using 2 µL of DNA extract 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 Inse01, or 0.2 µM of each primer for Euka2. Forward and reverse primers were 5'-labeled with eight-nucleotide tags with at least three differences between any pair of tags, so that each PCR replicate was identified by a unique combination of tags. This allowed the assignment of each sequence to the corresponding replicate during sequence analysis (Coissac 2012; Taberlet*et al.* 2018). 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 Euka2, and 1 min at 72°C (1m 30s for COI), and followed by a final elongation at 72°C for 7 min. Negative DNA extraction and PCR controls (ultrapure water, with 3 replicates as well) were analysed in parallel with the samples to monitor possible contaminations during the PCR step.

For Euka02 and Inse01, sequencing was performed by 2 × 125-bp pair-end sequencing on Illumina HiSeq 2500 platform, while for BF1_BR2-COI sequencing was performed by 2 × 250-bp pair-end sequencing on Illumina MiSeq platform at Fasteris (Geneva, Switzerland). Sequencing data were processed using the OBITools

(Boyer *et al.* 2016). Raw sequences were first aligned (illuminapairedend) to recover the amplicon sequence and then demultiplex (ngsfilter) to assign them to the samples. This was followed by dereplication (obiuniq) keeping track for each sequence of its count in the samples. Then for each sample, the ratio of counts for the most abundant sequence and the second most abundant sequence was calculated. Only the most abundant sequences having a count greater than 1000 and a ratio above 1/10 were considered to get rid of badly amplified samples and samples were several product were amplified.

As a further validation step, all the retrieved metabarcodes were matched against NCBI using BLAST, to identify eventual cases in which the obtained metabarcode is a spurious amplification of a non-target organism (*e.g.* fungi or algae). The *in vitro* amplification rate was measured for each taxon as the proportion of specimens for which we obtained valid metabarcodes.

2.2 Setting up the composite reference databases

For each species within the database of European freshwater organisms (Schmidt-Kloiber & Hering 2015), we matched the binomial name with the NCBI taxonomy database to retrieve their NCBI taxonomic code (taxid). All the available metabarcodes for the three regions of interest, together with their associated taxid, were extracted from the EMBL sequence data repository (release 136) using the ecoPCR program (Ficetola *et al.* 2010) by matching the primer sequences with up to 3 errors and restricting the metabarcodes to relevant lengths (length >30 bp for Euka02, length 70-270 bp for Inse01, length 100-500 for BF1_BR2-COI). The three composite reference databases (one for each metabarcode region) were then built by aggregating metabarcodes for each genus with those obtained from specimens analysed *in vitro* . In order to obtain the most complete coverage of genera found in France, we obtained the taxid of all metabarcodes produced through*in vitro* analyses as well as metabarcodes extracted from EMBL and associated to the taxid of a species found in France. For genera for which no such metabarcode existed, we included the metabarcodes extracted from EMBL and associated to the taxid of a species of the same genus found in Europe. If no such metabarcode existed, we included all the metabarcodes extracted from EMBL, and associated to a taxid belonging to this genus, also considering species that are not native in Europe.

2.3 Assessing the resolution of metabarcodes

We assessed the resolution of each metabarcoding region with the same procedure. First, the metabarcodes obtained as described above were compared to each other to find identical metabarcodes; this allowed producing a list of unique metabarcodes. For each unique metabarcode, we obtained the list of all the associated taxids. We tested taxonomic resolution at four levels: order, family, genus, and species. More specifically, we tested if, at a given taxonomic level, the list of associated taxids would collapse to a unique taxid or not (*i.e.*all taxids have the same ancestor taxid at that level). If a list would not collapse to one unique taxid for the tested taxonomic level, it meant that this metabarcode was not discriminant for this taxonomic level. Consider for instance a given metabarcode associated to multiple species within multiple genera within one single family. This particular metabarcode showed a family-level resolution, but not a species- or a genus-level resolution. It must be remarked that these measures of taxonomic resolution are heavily dependent on the available database. For example, if the database includes the metabarcode of only one species within a genus, this analysis could return a species-level resolution, even though it is possible that unanalysed species within the same genus share the same metabarcode.

3 RESULTS

3.1 *In vitro* analyses of reference specimens

We extracted and amplified DNA from 1514 individuals, belonging to 578 different taxa (Table 1). The majority of specimens were insects, and three insect orders with macrobenthic larvae (Ephemeroptera, Plecoptera and Trichoptera) altogether accounted for 80% of analysed specimens. Out of these specimens, 99% were morphologically identified at the family level or higher, 95% at the genus level or higher, and 62% at the species level. The average number of sampled individuals was 2.6 individuals per taxon (range: 1-12; median: 3). For Ephemeroptera, Plecoptera, Trichoptera and Megaloptera the analysed specimens covered well the

diversity of French and European benthic fauna (100%, 74%, 78% and 100% of genera recorded in France for Ephemeroptera, Plecoptera, Trichoptera and Megaloptera, respectively; 70%, 52%, 65% and 100% of all the genera recorded in Europe; Table 2). Representation was relatively good for Coleoptera, Hemiptera and Neuroptera, whereas coverage was weaker for the remaining orders of insects and for non-insects.

The amplification rate using the three metabarcodes was highly heterogeneous among taxonomic groups (Fig. 1). Euka02 (18S) showed the highest average amplification success (88%), with consistently high amplification success in all the taxa except Malacostraca (Fig. 1). Within insects, Euka02 showed excellent amplification success in most of orders, but its amplification success was poor with Diptera (Fig. 1b).

As expected, Inse01 showed good amplification success for insects (82%), while it showed a limited amplification of the remaining taxa (Fig. 1a). Within insects, Inse01 showed excellent amplification success in all the orders except Trichoptera, where amplification success was 71% (Fig. 1b).

Finally, BF1_BR2-COI showed an average amplification rate of 48%, with highly variable results among taxa (Fig. 1a). BF1_BR2-COI showed a good amplification rate with Gastropoda, Clitellata and Malacostraca, while the rate was lower for several orders of insects. Within insects, BF1_BR2-COI showed good performance in Coleoptera and Diptera (amplification success [?] 74%), while it amplified less than 50% of specimens from Ephemeroptera, Plecoptera and Trichoptera (Fig. 1b).

3.2 Combined database

When we combined sequences obtained *in vitro* with sequences obtained from GenBank, we obtained a total of 18 834 metabarcodes (3 441 for Euka02, 9 715 for Inse01 and 5 678 for BF1_BR2-COI). Insects accounted for the majority of metabarcodes, followed by Crustacea and Clitellata (Table 3). The combined database showed a good coverage of the diversity of European benthic fauna. For the Euka02 primer pair, the completeness of the database was particularly good (>80%) for Turbellaria, Coleoptera and Odonata. For Inse01, the completeness was particularly good for Coleoptera, Ephemeroptera and Odonata, while BF1_BR2-COI showed a relatively homogeneous completeness across taxa, with values between 50 and 70% for most of taxa (Fig. 2).

3.3 Taxonomic resolution of metabarcodes

The taxonomic resolution was strongly different among metabarcodes. For Euka02, 21% of metabarcodes were associated with more than one species in the database (Fig. 3a). The best resolution was observed for BF1_BR2-COI, with just 3% of metabarcodes associated with more than one species, while Inse01 showed an intermediate resolution (10% of metabarcodes associated with more than one species; Fig. 3a). The taxonomic resolutions of these metabacodes were clearly better if we consider the identification at the genus level (Fig. 3b). Euka02 showed the weakest performance, with around 6% of metabarcodes associated with more than one genus, while BF1_BR2-COI showed the best performance, with less than 1% of metabarcodes associated with more than one genus. Inse01 showed a generally good performance, with less than 1% of metabarcodes associated with more than one genus for most taxa. The performance was slightly poorer for Plecoptera and Trichoptera, with around 4% of metabarcodes associated with more than one genus. Family level identification was very good for all the metabarcodes, with a slightly poorer performance of Euka02 (Fig. 3c). It must be remarked that these values of resolution are calculated on an incomplete set of data, since our database did not include the sequences of many species and genera (Table 3), and all resolution estimates would probably be poorer if calculated on a complete database.

4 DISCUSSION

Metabarcoding-based biomonitoring requires the availability of primers with high performance, as they must amplify all the relevant target taxa, have sufficient resolution to identify them at the desired taxonomic level, and amplify short sequences usable with eDNA (Ficetola*et al.* 2010; Taberlet *et al.* 2018). Finding primers with all these features is challenging, and the identification of "perfect" metabarcodes has often been labelled as a "search for the Holy Grail" (Rubinoff *et al.* 2006). By combining an extensive *in vitro*analysis with the assessment of publicly-available sequences, our study highlights the complexity of finding all these desired

features in one single metabarcode. It also provides a comparison of performances, allowing the identification of most appropriate primers for different aims and taxonomic groups, and it produced a reference database for the taxonomic identification of a large number of benthic insects.

4.1 The importance of good reference databases

Metabarcoding enables biodiversity monitoring either with or without the taxonomic identification of the retrieved taxa. Taxonomic identification clearly requires appropriate reference databases that can be obtained*ad hoc* (*e.g.* by amplifying sequences from all the taxa from the target group) (Cilleros *et al.* 2019; Moriniere *et al.* 2019) or by searching public databases such as GenBank. Public databases offer an ever-growing resource, given that they combine the outcome of thousands of studies and produce a sheer amount of data that would be unreachable by *ad hoc* studies. Public databases are not error-free, still analyses showed that for animals, the error rate of GenBank for genus-level identification is generally low (˜0.7 / 3.5%), suggesting that it can be a formidable data source for applications relying on molecular data to understand the impact of environmental changes on biodiversity (Leray *et al.*2019). However, public databases are opportunistic collections of the material from multiple studies, thus they do not have the ambition of a taxonomic completeness. Ad-hoc databases (see also Ratnasingham & Hebert 2007) are thus essential resources to obtain the taxonomic coverage required if we want to identify most of benthic macroinvertebrates.

Several researchers advocated that COI-based markers should be favoured for metabarcoding because they are standard barcodes for animals, and thus we can expect a very large availability of sequences in reference databases (Andujar *et al.* 2018; Leray *et al.* 2019). For benthic macroinvertebrates, a very large number of COI sequences is available in GenBank (Table 3). For instance, BF1_BR2-COI is largely the marker with the highest number of sequences of benthic Diptera, with nearly 3,000 sequences of BF1_BR2-COI available against only 1000 sequences of Inse01 (16S rDNA), still the number of available sequences is surprisingly variable across taxa. Nevertheless, a very large number of sequences does not necessarily allow a better taxonomic coverage. In fact, most of genera of benthic Diptera do not have sequences in reference database for COI, and Inse01 sequences represent slightly more genera than BF1_BR2-COI (25% for Inse01 against just 15% for BF1_BR2-COI; Fig. 3). The mismatch between number of sequences and database completeness could be related to the different scopes of studies employing the different markers. In fact, COI is the most used marker by standard barcoding studies, which often aim at unveiling diversity among closely related, cryptic taxa, thus studies often consider many individuals from closely related, morphologically similar species within genera (Hebert *et al.* 2004). Conversely, the 16S and 18S rDNA genes are often used to build phylogenies (*e.g.*Alvarez-Presas *et al.* 2008; Criscione & Ponder 2013), and many phylogenetic studies aim at representing the largest number of genera and families. Such process could also explain the strong differences among taxa (*e.g.* a very high completeness for Euka02 with Turbellaria, and a much better coverage for Inse01 with Gastropoda; Fig. 3). If the aim is the species-level identification, databases should be exhaustive at the species-level, and markers should have a species-level resolution. Likely, for freshwater biomonitoring a genus-level identification is often enough (Bailey *et al.* 2001; Chessman*et al.* 2007), thus our database provides a good completeness that can allow the identification of most of genera, particularly with the markers Euka02 and Inse01.

Matching metabarcodes with reliable reference databases can allow obtaining metabarcoding-based biomonitoring data, that should be comparable with historical data obtained through traditional (*e.g.* morphological) approaches. Freshwater environments are highly sensitive to human impacts, and the availability of long-term time series is pivotal to identify trends of occupancy and the ecological quality of environments (Outhwaite *et al.* 2020).

4.2 Metabarcoding without taxonomic identification

Metabarcoding can provide ecological information even if reference databases are not available, as molecular taxonomic units can allow the comparison of communities among sites with environmental differences (Ji*et al.* 2013; Apotheloz-Perret-Gentil *et al.* 2017). The taxonomy-free approach allows overcoming the fact that, despite intensive efforts, databases remain incomplete for many taxa (Fig. 3). Primers with high taxonomic

coverage and resolution are essential also in this case. High taxonomic coverage is needed to avoid under-representation of some taxa, while resolution allows teasing apart related taxa. Related taxa can have very different ecological properties, and some widespread taxa, tolerant to human disturbance, can be closely related to highly sensitive specialists (Caro *et al.*2005). Therefore, ecological responses of communities can remain obscured if metabarcodes are not able to resolve related taxa with different ecology. Our study focused on European taxa, where taxonomic knowledge is particularly good (Moustakas & Karakassis 2005; Rodrigues*et al.* 2010; Brewer *et al.* 2012) and, with targeted studies, we could envisage an improvement of database completeness in the next years. However, our results on primer performance can be also useful in megadiverse, tropical areas, where taxonomy-free biomonitoring can be a viable option (Andersen *et al.* 2019).

4.3 Universality and resolution of primers

Our analysis did not identify one single outperforming metabarcode. The universality of primers was variable among taxa, with Euka02 showing the highest performance for some phyla (platyhelminthes, molluscs, annelids and even some arthropods), and Inse01 showing a generally good performance for insects. However, each of these metabarcodes has some drawbacks. For instance, Euka02 amplifies very long sequences for some taxa of crustaceans (Isopoda and Amphipoda; Guardiola *et al.*2015; Taberlet *et al.* 2018) thus their eDNA metabarcoding with this marker is problematic. Conversely, Inse01 is a metabarcode developed specifically for insects, and fails to amplify key freshwater taxa such as Turbellaria and molluscs (Fig. 1). In our *in vitro*analysis, BF1_BR2-COI showed a moderate amplification rate, still for insects a relevant proportion of specimens were not amplified (Fig. 1). This is in contrast with previous analyses, that successfully amplified 100% of tested insects using BF1_BR2-COI (Elbrecht & Leese 2017). Differences might be due to DNA quality, as this primer amplifies relatively long metabarcodes (>300 bp). Some of our >1500 specimens were old, and this can cause DNA degradation, while the starting material of Elbrecht and Leese (2017) was probably of better quality. Furthermore, in several cases BF1_BR2-COI did not amplify the DNA of our target organisms, but amplified the DNA of contaminants, i.e. other organisms for which small body fragments were probably present in the tube, and that perhaps showed excellent match with the primers. Unfortunately, these conditions (degraded DNA, and contemporaneous presence of many organisms) are typical of eDNA metabarcoding studies, stressing the complexity of finding appropriate primers.

Differences in performance were also strong when considering the resolution of the metabarcodes. BF1_BR2-COI clearly showed the best resolution while Euka02 showed a very poor performance, as in many cases it failed even at the family level (Fig. 3; see also Tang *et al.*2012). COI is a highly variable region, and this has promoted its use as standard barcode for animals (Hebert *et al.* 2003; Hebert *et al.* 2004; Andujar *et al.* 2018). The excellent performance of BF1_BR2-COI can also be explained by the relatively long amplified region. Inse01 showed an intermediate performance, as its resolution was insufficient for species-level identification, while genus level identification was good for most of taxa (Fig. 3). It must be remarked that these are optimistic values of resolution, given that our database was far from complete, particularly at the genus-level and for some taxa, therefore a more complete database could yield poorer resolution values.

4.4 No Holy Grail for macrobenthos metabarcodes?

The heterogeneous performance of the different metabarcodes highlights the complexity of the identification of the best markers. No primer showed the best performance for all the considered metrics, as the most "universal" marker (Euka02) showed a generally poor resolution, while the marker with the highest resolution (BF1_BR2-COI) did not successfully amplify many taxa. The selection of metabarcodes for biomonitoring is therefore a trade-off, depending on the aims of studies. Euka02 can allow a good assessment of overall biodiversity, but it is unable to tease apart closely related taxa, thus it might be not enough to define the ecological status of environments. Furthermore, the poor resolution would hamper the comparison with historical data for most of taxa. Conversely, the excellent resolution of BF1_BR2-COI could allow species-level identification, and might have more power to distinguish different communities. However, this comes at a cost. Many taxa did not amplify either because the level of DNA degradation compromises the amplification of a relatively long metabarcode, or because the poor match of the primer(s) with their target. In fact, the

relatively long amplified metabarcode could limit its usefulness for application with environmental DNA extracted from water. Finally, Inse01 showed a generally good performance, but it is not appropriate for many non-insect taxa.

Given these limitations, it is unlikely that one single metabarcode will be able to fully replace the traditional biomonitoring using macrobenthic invertebrates. Nevertheless, the data obtained through multiple metabarcodes can be integrated for a more comprehensive and accurate understanding of ecological impacts on freshwater biodiversity. For instance, highly universal markers, providing a complete but coarse picture of animal biodiversity (*e.g.* Euka02) can be combined with markers providing a specific focus on key taxa (*e.g.* Inse01) or a high-resolution level (*e.g.* BF1_BR2-COI). The integration of multiple metabarcodes certainly increases the cost and complexity of studies, still it has the potential to provide an unprecedented amount of data, thus opening unexplored avenues to biodiversity assessment.

Acknowledgments

This study was funded by Agence Nationale de la Recherche (ANR-13-ECOT-0002 aquaDNA). GFF is partially funded by the European Research Council under the European CommunityHorizon 2020 Programme, Grant Agreement no. 772284 (IceCommunities). We thank SPYGEN staff for its help in laboratory, and S. Jolivet, M. Brulin, J. La Doare, G. Coppa, M. Ferran and OPIE volunteers for the collection of reference specimens".

Competing interest

A.V. and T.D. are research scientists at a private company specialising in the use of eDNA for species detection.

References

Alvarez-Presas M, Baguna J, Riutort M (2008) Molecular phylogeny of land and freshwater planarians (Tricladida, Platyhelminthes): From freshwater to land and back. *Molecular Phylogenetics and Evolution* ,**47** , 555-568.

Andersen JC, Oboyski P, Davies N*, et al.* (2019) Categorization of species as native or nonnative using DNA sequence signatures without a complete reference library. *Ecological Applications* , **29** , 11.

Andujar C, Arribas P, Yu DW, Vogler AP, Emerson BC (2018) Why the COI barcode should be the community DNA metabarcode for the metazoa.*Molecular Ecology* , **27** , 3968-3975.

Apotheloz-Perret-Gentil L, Cordonier A, Straub F*, et al.* (2017) Taxonomy-free molecular diatom index for high-throughput eDNA biomonitoring. *Molecular Ecology Resources* , **17** , 1231-1242.

Arce E, Archaimbault V, Mondy CP, Usseglio-Polatera P (2014) Recovery dynamics in invertebrate communities following water-quality improvement: taxonomy- vs trait-based assessment. *Freshwater Science* , **33** , 1060-1073.

Archaimbault V, Usseglio-Polatera P, Garric J, Wasson JG, Babut M (2010) Assessing pollution of toxic sediment in streams using bio-ecological traits of benthic macroinvertebrates. *Freshwater Biology* ,**55** , 1430-1446.

Arman NZ, Salmiati S, Said MIM, Aris A (2019) Development of macroinvertebrate-based multimetric index and establishment of biocriteria for river health assessment in Malaysia. *Ecological Indicators* , **104** , 449-458.

Bailey RC, Norris RH, Reynoldson TB (2001) Taxonomic resolution of benthic macroinvertebrate communities in bioassessments. *Journal of the North American Benthological Society* , **20** , 280-286.

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

Birk S, Bonne W, Borja A*, et al.* (2012) Three hundred ways to assess Europe's surface waters: An almost complete overview of biological methods to implement the Water Framework Directive.*Ecological Indicators* , **18** , 31-41.

Boyer F, Mercier C, Bonin A*, et al.* (2016) obitools: a unix-inspired software package for DNA metabarcoding. *Molecular Ecology Resources* , **16** , 176-182.

Brabec K, Zahradkova S, Nemejcova D*, et al.* (2004) Assessment of organic pollution effect considering differences between lotic and lentic stream habitats. *Hydrobiologia* , **516** , 331-346.

Brewer MS, Sierwald P, Bond JE (2012) Millipede Taxonomy after 250 Years: Classification and Taxonomic Practices in a Mega-Diverse yet Understudied Arthropod Group. *Plos One* , **7** , 12.

Buxton AS, Groombridge JJ, Griffiths RA (2017) Is the detection of aquatic environmental DNA influenced by substrate type? *Plos One* ,**12** .

Bylemans J, Furlan EM, Gleeson DM, Hardy CM, Duncan RP (2018) Does Size Matter? An Experimental Evaluation of the Relative Abundance and Decay Rates of Aquatic Environmental DNA. *Environmental Science & Technology* , **52** , 6408-6416.

Camargo JA (2017) Multimetric assessment of macroinvertebrate responses to mitigation measures in a dammed and polluted river of Central Spain.*Ecological Indicators* , **83** , 356-367.

Carlson PE, Donadi S, Sandin L (2018) Responses of macroinvertebrate communities to small dam removals: Implications for bioassessment and restoration. *Journal of Applied Ecology* , **55** , 1896-1907.

Caro T, Eadie J, Sih A (2005) Use of substitute species in conservation biology. *Conservation Biology* , **1821-1826** .

Chessman B, Williams S, Besley C (2007) Bioassessment of streams with macroinvertebrates: effect of sampled habitat and taxonomic resolution.*Journal of the North American Benthological Society* , **26** , 546-565.

Cilleros K, Valentini A, Allard L*, et al.* (2019) Unlocking biodiversity and conservation studies in high-diversity environments using environmental DNA (eDNA): A test with Guianese freshwater fishes.*Molecular Ecology Resources* , **19** , 27-46.

Coissac E (2012) OligoTag: a program for designing sets of tags for nextgeneration sequencing of multiplexed samples. *Methods in Molecular Biology* , **888** , 13–31.

Cox B, Oeding S, Taffs K (2019) A comparison of macroinvertebrate-based indices for biological assessment of river health: A case example from the sub-tropical Richmond River Catchment in northeast New South Wales, Australia. *Ecological Indicators* , **106** .

Criscione F, Ponder WF (2013) A phylogenetic analysis of rissooidean and cingulopsoidean families (Gastropoda: Caenogastropoda). *Molecular Phylogenetics and Evolution* , **66** , 1075-1082.

Czechowski P, Stevens MI, Madden C, Weinstein P (2020) Steps towards a more efficient use of chironomids as bioindicators for freshwater bioassessment: Exploiting eDNA and other genetic tools. *Ecological Indicators* , **110** , 11.

Eichmiller JJ, Best SE, Sorensen PW (2016) Effects of Temperature and Trophic State on Degradation of Environmental DNA in Lake Water.*Environmental Science and Technology* , **50** , 1859-1867.

Elbrecht V, Leese F (2017) Validation and Development of COI Metabarcoding Primers for Freshwater Macroinvertebrate Bioassessment.*Frontiers in Environmental Science* , **5** , 11.

Ficetola GF, Coissac E, Zundel S*, et al.* (2010) An In silico approach for the evaluation of DNA barcodes. *BMC Genomics* ,**11** , 434.

Ficetola GF, Miaud C, Pompanon F, Taberlet P (2008) Species detection using environmental DNA from water samples. *Biology Letters* ,**4** , 423-425.

Floury M, Usseglio-Polatera P, Ferreol M, Delattre C, Souchon Y (2013) Global climate change in large European rivers: long-term effects on macroinvertebrate communities and potential local confounding factors.*Global Change Biology* , **19** , 1085-1099.

Friberg N, Sandin L, Furse MT*, et al.* (2006) Comparison of macroinvertebrate sampling methods in Europe. *Hydrobiologia* ,**566** , 365-378.

Gabriels W, Lock K, De Pauw N, Goethals PLM (2010) Multimetric Macroinvertebrate Index Flanders (MMIF) for biological assessment of rivers and lakes in Flanders (Belgium). *Limnologica* , **40** , 199-207.

Guardiola M, Uriz MJ, Taberlet P*, et al.* (2015) Deep-Sea, Deep-Sequencing: Metabarcoding Extracellular DNA from Sediments of Marine Canyons. *Plos One* , **10** , e0139633.

Haase P, Lohse S, Pauls S*, et al.* (2004) Assessing streams in Germany with benthic invertebrates: development of a practical standardised protocol for macro invertebrate sampling and sorting.*Limnologica* , **34** , 349-365.

Hebert PDN, Penton EH, Burns JM, Janzen DH, Hallwachs W (2004) Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator* . *Proceedings of the National Academy of Sciences of the United States of America* ,**101** , 14812-14817.

Hebert PDN, Ratnasingham S, deWaard JR (2003) Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society B-Biological Sciences* ,**270** , S96-S99.

Hering D, Borja A, Carvalho L, Feld CK (2013) Assessment and recovery of European water bodies: key messages from the WISER project.*Hydrobiologia* , **704** , 1-9.

Hering D, Borja A, Jones JI*, et al.* (2018) Implementation options for DNA-based identification into ecological status assessment under the European Water Framework Directive. *Water Research* , **138** , 192-205.

Hering D, Feld CK, Moog O, Ofenbock T (2006a) Cook book for the development of a Multimetric Index for biological condition of aquatic ecosystems: Experiences from the European AQEM and STAR projects and related initiatives. *Hydrobiologia* , **566** , 311-324.

Hering D, Johnson RK, Kramm S*, et al.* (2006b) Assessment of European streams with diatoms, macrophytes, macroinvertebrates and fish: a comparative metric-based analysis of organism response to stress.*Freshwater Biology* , **51** , 1757-1785.

Iversen LL, Winkel A, Baastrup-Spohr L*, et al.* (2019) Catchment properties and the photosynthetic trait composition of freshwater plant communities. *Science* , **366** , 878.

Ji Y, Ashton L, Pedley SM*, et al.* (2013) Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding. *Ecology Letters* , **16** , 1245-1257.

Jo T, Murakami H, Masuda R*, et al.* (2017) Rapid degradation of longer DNA fragments enables the improved estimation of distribution and biomass using environmental DNA. *Molecular Ecology Resources* ,**17** , e25-e33.

Kupilas B, Friberg N, McKie BG*, et al.* (2016) River restoration and the trophic structure of benthic invertebrate communities across 16 European restoration projects. *Hydrobiologia* , **769** , 105-120.

Leray M, Knowlton N, Ho SL, Nguyen BN, Machida RJ (2019) GenBank is a reliable resource for 21st century biodiversity research.*Proceedings of the National Academy of Sciences of the United States of America* , **116** , 22651-22656.

Marquina D, Andersson AF, Ronquist F (2019) New mitochondrial primers for metabarcoding of insects, designed and evaluated using in silico methods. *Molecular Ecology Resources* , **19** , 90-104.

Marzin A, Archaimbault V, Belliard J*, et al.* (2012) Ecological assessment of running waters: Do macrophytes, macroinvertebrates, diatoms and fish show similar responses to human pressures? *Ecological Indicators* , **23** , 56-65.

Miler O, Porst G, McGoff E*, et al.* (2013) Morphological alterations of lake shores in Europe: A multimetric ecological assessment approach using benthic macroinvertebrates. *Ecological Indicators* , **34** , 398-410.

Mondy CP, Usseglio-Polatera P (2013) Using conditional tree forests and life history traits to assess specific risks of stream degradation under multiple pressure scenario. *Science of the Total Environment* ,**461** , 750-760.

Moriniere J, Balke M, Doczkal D*, et al.* (2019) A DNA barcode library for 5,200 German flies and midges (Insecta: Diptera) and its implications for metabarcoding-based biomonitoring. *Molecular Ecology Resources* , **19** , 900-928.

Moustakas A, Karakassis I (2005) How diverse is aquatic biodiversity research? *Aquatic Ecology* , **39** , 367-375.

Noges P, Argillier C, Borja A*, et al.* (2016) Quantified biotic and abiotic responses to multiple stress in freshwater, marine and ground waters. *Science of the Total Environment* , **540** , 43-52.

Outhwaite CL, Gregory RD, Chandler RE, Collen B, Isaac NJB (2020) Complex long-term biodiversity change among invertebrates, bryophytes and lichens. *Nature Ecology & Evolution* .

Pawlowski J, Kelly-Quinn M, Altermatt F*, et al.* (2018) The future of biotic indices in the ecogenomic era: Integrating (e)DNA metabarcoding in biological assessment of aquatic ecosystems. *Science of The Total Environment* , **637-638** , 1295-1310.

Pinol J, Mir G, Gomez-Polo P, Agusti N (2015) Universal and blocking primer mismatches limit the use of high-throughput DNA sequencing for the quantitative metabarcoding of arthropods. *Molecular Ecology Resources* , **15** , 819-830.

Pont D, Valentini A, Rocle M*, et al.* (2020) The future of fish-based ecological assessment of European rivers: from traditional EU Water Framework Directive compliant methods to eDNA metabarcoding-based approaches. *Journal of Fish Biology* , **in press** ,*https://doi.org/10.1111/jfb.14176*.

Ratnasingham S, Hebert PDN (2007) BOLD: The Barcode of Life Data System (*www.barcodinglife.org*). *Molecular Ecology Notes* , **7** , 355-364.

Rodrigues ASL, Gray CL, Crowter BJ*, et al.* (2010) A global assessment of amphibian taxonomic effort and expertise. *Bioscience* , **60** , 798-806.

Rubinoff D, Cameron S, Will K (2006) Are plant DNA barcodes a search for the Holy Grail? *Trends in Ecology & Evolution* , **21** , 1-2.

Schmidt-Kloiber A, Hering D (2015) *www.freshwaterecology.info* – An online tool that unifies, standardises and codifies more than 20,000 European freshwater organisms and their ecological preferences. *Ecological Indicators* , **53** , 271-282.

Serrana JM, 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.

Sigsgaard EE, Nielsen IB, Bach SS*, et al.* (2017) Population characteristics of a large whale shark aggregation inferred from seawater environmental DNA. *Nature Ecology & Evolution* ,**1** , 0004.

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

Tachet H, Richoux P, Bournaud M, Usseglio-Polatera P (2010)*Invertebres d'eau douce. Systematique, biologie, ecologie* CNRS Editions, Paris.

Tang CQ, Leasi F, Obertegger U*, et al.* (2012) The widely used small subunit 18S rDNA molecule greatly underestimates true diversity in biodiversity surveys of the meiofauna. *Proceedings of the National Academy of Sciences of the United States of America* , **109** , 16208-16212.

Theodoropoulos C, Karaouzas I, Vourka A, Skoulikidis N (2020) ELF - A benthic macroinvertebrate multimetric index for the assessment and classification of hydrological alteration in rivers. *Ecological Indicators* , **108** , 11.

Usseglio-Polatera P, Bournaud M, Richoux P, Tachet H (2000) Biological and ecological traits of benthic freshwater macroinvertebrates: relationships and definition of groups with similar traits.*Freshwater Biology* , **43** , 175-205.

Usseglio-Polatera P, Richoux P, Bournaud M, Tachet H (2001) A functional classification of benthic macroinvertebrates based on biological and ecological traits: application to river condition assessment and stream management. *Archiv fur Hydrobiologie* , **139 Suppl.1 Monogr. Stud.** , 53-83.

Yang JH, Zhang XW (2020) eDNA metabarcoding in zooplankton improves the ecological status assessment of aquatic ecosystems. *Environment International* , **134** , 9.

Yu DW, Ji Y, Emerson BC*, et al.* (2012) Biodiversity soup: metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. *Methods in Ecology and Evolution* , **3** , 613-623.

Data Accessibility: The complete database is available at fighshare:

10.6084/m9.figshare.12046242

(the link will become active upon manuscript acceptance)

The filtered sequences generated by the project will be uploaded on GenBank upon manuscript acceptance (accession numbers: XXXXXX-XXXXXX).

Author Contributions

Designed the research: Pierre Taberlet, Tony Dejean, Gentile Francesco Ficetola, Frederic Boyer, Alice Valentini, Aurelie Bonin, Albin Meyer, Philippe Usseglio-Polatera

Performed research: Alice Valentini, Aurelie Bonin, Albin Meyer, Coline Gaboriaud, Pierre Taberlet, Tony Dejean, Gentile Francesco Ficetola, Frederic Boyer, Philippe Usseglio-Polatera

Analyzed data: Gentile Francesco Ficetola, Frederic Boyer, Aurelie Bonin

Wrote the paper: Gentile Francesco Ficetola wrote the first draft of the paper, with subsequent contribution of all the authors

Figure legends

Figure 1. Amplification rate of the three markers across benthic macroinvertebrate taxa. A): all taxa; B): insects only.

Figure 2. Completeness of the combined database, combining the sequences produced in this study with sequences retrieved from public databases. For each taxon, the plots report the proportion of European genera of macroinvertebrates with at least one sequence in the database.

Figure 3. Resolution of the three markers at the species, genus and family levels. The resolution is measured as the proportion of metabarcodes that are associated with a) at least two species; b) at least two genera; c) at

least two families (non-identification), therefore low values of non-identification indicate a better performance of the markers.

Table 1. Inventory of macrobenthos specimens from which we extracted and amplified DNA.

Class / subphylum		*N* individuals	% identified at the genus level by taxonomists	*N* genera
Turbellaria		7	100%	2
Bivalvia		12	100%	5
Gastropoda		29	97%	11
Clitellata		35	69%	8
Arachnida			100%	1
Hydracnidia				
Crustacea		9	100%	4
Insecta				
	Coleoptera	117	97%	40
	Diptera	54	20%	6
	Ephemeroptera	338	100%	35
	Hemiptera	24	100%	14
	Lepidoptera	2	100%	2
	Megaloptera	4	100%	1
	Neuroptera	2	100%	1
	Odonata	9	78%	2
	Plecoptera	210	100%	20
	Trichoptera	651	100%	84

Table 2. Representativeness of specimens used for *in-vitro*analyses, relative to European and French genera of benthic macroinvertebrates.

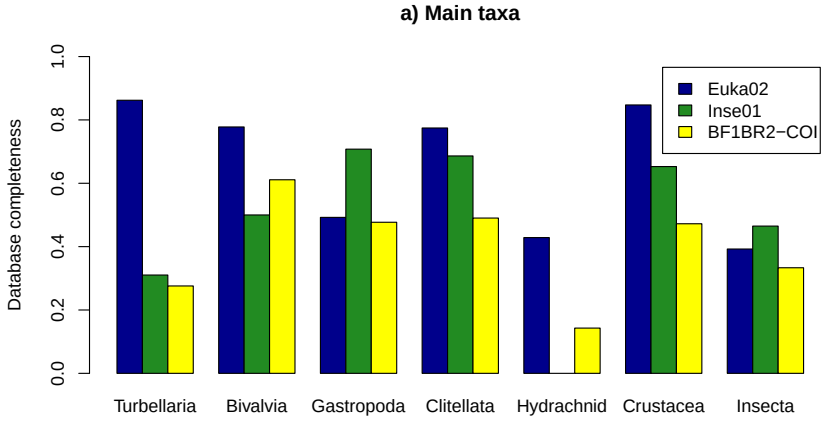
Class / subphylum	Order (insects only)	*N* genera Europe	*N* Genera France	*N* genera for *in vitro* analyses
Turbellaria		29	23	2
Bivalvia		18	11	5
Gastropoda		65	34	11
Clitellata		102	64	8
Arachnida		1	1	1
Hydracnidia		56	52	0
Crustacea		72	34	4
Insecta				
	Coleoptera	127	95	40
	Diptera	522	323	6
	Ephemeroptera	50	35	35
	Hemiptera	30	28	14
	Lepidoptera	8	5	2
	Megaloptera	1	1	1
	Neuroptera	3	3	1
	Odonata	43	36	2
	Plecoptera	38	27	20
	Trichoptera	130	108	84

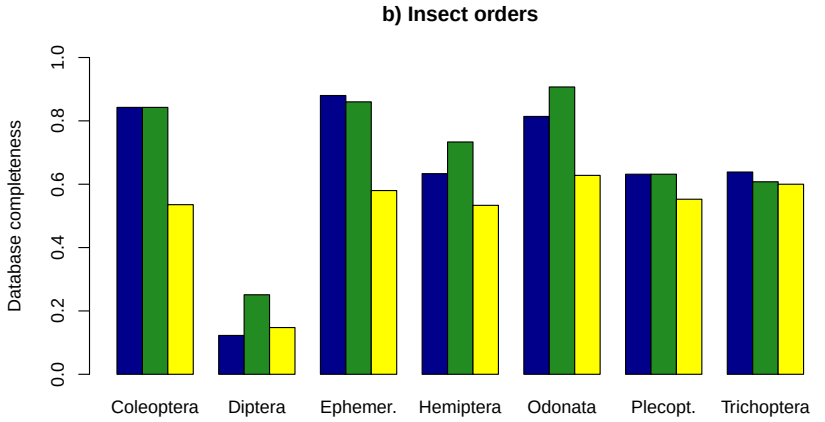
Table 3. Number of sequences and genera represented in the combined database, across taxa. Taxa for which >70% of European genera are represented in the database are highlighted in bold.

Class / subphylum	Order (insects only)	*N* sequences in the database	*N* sequences in the database	*N* sequences in the database	*N* genera Europe	*N* genera in the database	*N* genera in the database	*N* genera in the database	*N* in da
	|	Euka02	Inse01	COI		Euka02	Inse01	COI	To
Hydrozoa		31	134	21	6	**5**	**5**	3	**5**
Enopla (Nemertini)		4	1	-	1	**1**	**1**	0	**1**
Turbellaria		217	25	316	29	**25**	9	8	**26**
Gordioida (Nematomorpha)		9	-	-	1	**1**	-	-	**1**
Bivalvia		66	453	125	18	**14**	9	11	**16**
Gastropoda		81	1147	102	65	32	**46**	31	**51**
Clitellata		414	838	170	102	**79**	70	50	**86**
Polychaeta		33	74	64	11	**8**	4	2	**8**
Gymnolaemata (Bryozoa)		23	44	-	3	2	2	-	2
Phylactolaemata (Bryozoa)		17	36	1	6	**6**	**6**	1	**6**
Arachnida		2	2	4	1	**1**	**1**	**1**	**1**
Hydrachnidia		70	-	15	56	24	0	8	24
Crustacea		325	1980	303	72	**61**	47	34	**65**
Insecta		2149	4981	4557	981	385	456	327	53
	Coleoptera	450	1809	333	127	**107**	**107**	68	**11**
	Diptera	228	1078	2839	522	64	131	77	17
	Ephemeroptera	396	507	397	50	**44**	**43**	29	**44**
	Hemiptera	46	206	70	30	19	**22**	16	**24**
	Hymenoptera	4	21	33	29	3	4	5	9
	Lepidoptera	2	4	5	8	2	3	3	3
	Megaloptera	9	10	6	1	**1**	**1**	**1**	**1**
	Neuroptera	6	12	5	3	**3**	**3**	2	**3**
	Odonata	164	537	294	43	**35**	**39**	27	**41**
	Plecoptera	243	233	118	38	24	24	21	26
	Trichoptera	601	564	457	130	83	79	78	**94**

Hosted file

`Fig 1.pptx` available at https://authorea.com/users/308244/articles/439301-metabarcodes-for-the-monitoring-of-freshwater-benthic-biodiversity-through-environmental-dna





Hosted file

`Fig 3.pptx` available at https://authorea.com/users/308244/articles/439301-metabarcodes-for-the-monitoring-of-freshwater-benthic-biodiversity-through-environmental-dna