ARPHA Preprints

PENSOFT

PREPRINT

Author-formatted, not peer-reviewed document posted on 31/03/2023

DOI: https://doi.org/10.3897/arphapreprints.e104185

Evaluating five primer pairs for environmental DNA metabarcoding of Central European fish species based on mock communities

Till-Hendrik Macher, Robin Schütz, Atakan Yildiz, Arne Beermann, Florian Leese

ARPHA Preprints *Author-formatted, not peer-reviewed document posted on 31/03/2023* DOI: https://doi.org/10.3897/arphapreprints.e104185

Evaluating five primer pairs for environmental DNA metabarcoding of Central European fish species based on mock communities

Till-Hendrik Macher[1], Robin Schütz[1], Atakan Yildiz[2], Arne J. Beermann[1,3], Florian Leese[1,3]

[1] University of Duisburg-Essen, Aquatic Ecosystem Research, Universitaetsstr. 5, 45141 Essen, Germany
[2] Biotechnology Institute, Ankara University, Ankara 06135, Turkey
[3] University of Duisburg-Essen, Centre for Water and Environmental Research (ZWU), Universitaetsstr. 3, 45141 Essen, Germany

Running head: Fish mock community for eDNA metabarcoding

Keywords: biomonitoring, COI, 12S, 16S, eDNA

Abstract

Environmental DNA (eDNA) metabarcoding has become a powerful tool for examining fish communities. The demand for methodological standardization and the implementation of eDNA-based assessments into the regulatory monitoring (e.g., Water Framework Directive) are imminent. To ensure methodical accuracy and to meet regulatory standards, various sampling, laboratory and bioinformatic workflows have been established. However, a crucial prerequisite for a comprehensive fish monitoring is the choice of suitable primer pairs to accurately depict the present fish fauna. Various fish-specific primer pairs targeting different genetic marker regions were published over the past decade. However, a dedicated study to evaluate performance of frequently applied fish primer pairs to assess Central European fish species has not yet been conducted. Therefore, we created an artificial community composed of DNA from 45 Central European fish species and examined the discriminatory power and reproducibility of five fish primer pairs. Our study highlights the effect of the primer choice and bioinformatic filtering on the outcome of eDNA metabarcoding results. From the five primer pairs evaluated in our study the tele02 (12S gene) primer pair proved to be best choice for eDNA metabarcoding of Central European freshwater fish. Here, the MiFish-U (12S) and SeaDNA-mid (COI) primer pairs also displayed good discriminatory power and reproducibility. However, more general primer pairs (i.e., targeting vertebrates) were found to be less reliable and generated high numbers of false-positive and false-negative detections. Our study illustrates how careful selection of primer pairs and bioinformatic pipelines can make eDNA metabarcoding a more reliable tool for fish monitoring.

Introduction

Environmental DNA (eDNA) metabarcoding has become a valuable tool for monitoring fish species in different habitats (McDevitt et al. 2019, Wang et al. 2021, Miya 2022). Several studies have compared eDNA-based monitoring to traditional monitoring approaches, such as gillnetting or electrofishing, proving eDNA metabarcoding to be a reliable, fast, sensitive, non-invasive and cost-efficient method for fish detection (Pont et al. 2018, Fujii et al. 2019, Boivin-Delisle et al. 2021). However, applying eDNA metabarcoding comes with certain challenges such as the selection of appropriate sampling strategies and wet lab processing steps, completeness of reference databases, and choice of appropriate primers (Evans et al. 2017, Kumar et al. 2022). As a prerequisite for a comprehensive biodiversity monitoring, suitable primers are crucial to avoid false-negative detection and accurately depict the present fish fauna (Schenekar et al. 2020). Mock community metabarcoding is an efficient *in vitro* approach to test the performance of primers using an artificially composed DNA mixture representing the expected target community for biomonitoring (Hänfling et al. 2016, Elbrecht et al. 2019). While different metabarcoding fish primers have been evaluated on natural communities, larger systematic tests of primers with fish mock communities are missing (Bylemans et al. 2018a, Miya et al. 2020, Zhang et al. 2020, Shu et al. 2021). These studies focused on the detection of Asian and Australian fish species, which are genetically divergent and differing in species composition from the Central European fish fauna. Primer pairs for European fish communities have for now only been evaluated for estuarine and costal eDNA samples (Collins et al. 2019) and on smaller scale for UK lake fish (Hänfling et al. 2016). Thus, especially for the implementation of fish eDNA metabarcoding in routine monitoring programs such as the European Water Framework Directive (Hering et al. 2018, Pont et al. 2021), it is crucial to evaluate suitable primer pairs regarding their detection ability of the most common European freshwater fish species and investigate false-positive and false-negative detections.

In this study, we addressed this issue and evaluated five published fish eDNA metabarcoding primers (targeting fish and other vertebrates) by testing their performance on an artificial community composed of DNA from 45 Central European fish species. Here, we examined the discriminatory power and reproducibility of the five primer pairs, investigated their false-positive and false-negative detection rates, and investigated primer-specific biases. Finally, we conclude with a primer pair recommendation for eDNA metabarcoding approaches targeting fish in routine monitoring campaigns.

Methods

Fish swabs

Mucus samples of 66 specimens (45 species) were collected by fish bioassessment experts during electrofishing campaigns in autumn 2020 at five sites across Germany, covering both the Rhine and the Danube catchment. Each mucus sample was collected individually using sterile swabs (FLOQ Swab 80 mm, minitip, without medium, sterile sleeve; COPAN, Italy). All fish were handled as quickly as possible outside the water to keep the stress to a minimum, while a sterile swab was moved across the specimens' flank. Swabs were placed back into the sleeve and sealed. After field work, samples were stored at 4°C until delivery to the University of Duisburg-

Essen. Upon arrival the swabs were stored at -20°C overnight followed by DNA extraction.

DNA extraction

Swab tips were clipped off at the handle and placed in a sterile 1.5 mL Eppendorf tube before 1 mL TNES buffer and 15 µL Proteinase K (300 U/mL, 7BioScience, Neuenburg am Rhein, Germany) was added to the sample. Samples were incubated at 55°C and shaken at 1000 rpm for 3 h on an Eppendorf ThermoMixer C (Eppendorf AG, Hamburg, Germany). Subsequently, DNA was extracted using an adapted NucleoMag tissue kit (Macherey Nagel, Düren, Germany; Supplementary Material 1). In total, a volume of 400 µL per sample was extracted and DNA was eluted in a final volume of 50 µL elution buffer. DNA concentration of each sample was measured using a Qubit dsDNA HS Assay-Kit on a Qubit v2 fluorometer (Thermo Fisher Scientific).

Mock community composition

Two fish mock communities were created using the extracted fish swab DNA. In case of several collected specimens per species, only the sample with the highest DNA concentration was used for the composition of the mock community in order to represent each species only by a single individual. The first normalized mock community (MC1) was equimolarly pooled to 2 ng DNA per species. A second mock community (MC2) was pooled using 1 µL of each extract to generate a mock community with different DNA concentrations per species. MC2 was used to test for potential correlation between DNA concentration and number of reads. Both mock communities contained DNA of 45 fish species.

DNA amplification and sequencing

Both mock communities were assessed using five different published primer pairs for DNA amplification: tele02 (Taberlet et al. 2018), MiFish-U (Miya et al. 2015), 12Sv5 (Riaz et al. 2011), SeaDNA-mid (Collins et al. 2019) and L2513/H2714 (Kitano et al. 2007). A two-step PCR approach (Bohmann et al. 2022) was applied for amplifying the molecular marker genes and tagging of amplicons with barcodes and Illumina sequencing adaptors. In the 1st-step PCR step, tagged versions of the five fish primer pairs were used (Table 1).

In total, 60 1st-step PCR amplifications were conducted, including five replicates for each mock communities (MC1 and MC2) and two negative PCR controls for each of the 5 primer pairs. The reaction volume was 25 µL, consisting of 12.5 µL Multiplex Mastermix (Qiagen Multiplex PCR Plus Kit, Qiagen, Hilden, Germany), 7 µL PCR-grade water, 2.5 µL CoralLoad dye, 0.5 µL forward primer, 0.5 µL reverse primer (10 µM each), and 2 µL of DNA template. The 1st-step PCR included following steps: 5 min 95 °C initial denaturation, followed by 10 cycles of 30 s at 95 °C, 90 s at decreasing annealing temperature (starting from annealing temperature +10 °C), and 30 s at 72 °C, followed by 25 cycles of 30 s at 95 °C, 90 s at the respective annealing temperature (tele02: 52 °C, MiFish-U: 59 °C, SeaDNA-mid: 53 °C 12SV5: 52 °C, and LH16S: 55 °C), and 30 s at 72 °C. The final elongation was 10 min at 68 °C. Subsequently, PCR products were size selected using magnetic beads (ratio 0.7, dx.doi.org/10.17504/protocols.io.36wgqj45xvk5/v2) to remove excessive primers and reduce subsequent primer dimer formation.

In the 2nd-step PCR, Illumina sequencing adapters with a dual twin-indexing system were added (Buchner et al. 2021, Bohmann et al. 2022). For each sample, the 2nd-step

PCR mix contained, 7.5 µL Multiplex Mix, 1.8 µL PCR-grade water, 1.5 µL CoralLoad dye, 1.2 µL combined primer (5 µM), and 3 µL 1st-step PCR product. The 2nd-step PCR included the following steps: 5 min 95 °C initial denaturation, followed by 10 cycles of 30 s at 95 °C and 120 s at 72 °C. The final elongation was 10 min at 68 °C. The 2nd-step PCR products were visualized on a 1% agarose gel to evaluate amplification success. Negative controls did not produce bands on the gel. Then, PCR products were size selected using magnetic normalization beads (ratio 0.7, dx.doi.org/10.17504/protocols.io.q26g7y859gwz/v1) to normalize samples and remove excessive primers and primer dimers. Subsequently, all normalized PCR products were pooled into one library. The pooled library was concentrated using a NucleoSpin Gel and PCR Clean-up kit (Macherey Nagel, Düren, Germany) following the manufacturer's protocol. The final elution volume of the library was 40 µL. The library was then analysed using a Fragment Analyzer (High Sensitivity NGS Fragment Analysis Kit; Advanced Analytical, Ankeny, USA) to check for potential primer dimers and co-amplification, and to quantify the DNA concentration of the library. The final library was sequenced on a MiSeq 250bp PE V3 Illumina platform at CeGat (Tübingen, Germany).

Bioinformatics

Raw reads were received as demultiplexed fastq files. All samples were processed with the APSCALE-GUI pipeline v1.1.6 (Buchner et al. 2022), which is based on VSEARCH (Rognes et al. 2016) and cutadapt (Martin 2011). Each primer pair was processed separately. All settings were kept as default, and OTUs were clustered with a 97% percentage similarity threshold. Subsequently, taxonomy was assigned using the 'local BLAST' function in APSCALE with the Midori2 databases (v249 of CO1, lrRNA and srRNA; Leray, Knowlton, and Machida 2022) as reference.

The taxonomic assignment of each OTU was filtered using APSCALE-GUI (Supplementary Figure 1). Initially, taxonomic assignments were filtered by e-value (hits with the lowest e-value are kept) and hits with the same taxonomy were dereplicated. Subsequently, taxonomic assignments were adjusted according to similarity thresholds (species ≥ 97%, genus ≥ 95%, family ≥ 90%, order ≥ 85%). If at this point more than one taxon assigned to species level was remaining, additional filtering and flag raising steps were performed as follows: All ambiguous taxa were saved to a separate column in the taxonomy table. The number of occurrences per remaining taxon was counted. If a dominant species was present, it was selected as taxonomic assignment ("F1 - Dominant taxon"). Otherwise, if two species of the same genus remained, the genus was saved with the two possible species names separated by slash (e.g., *Leuciscus idus*/*leuciscus*; "F2 - Two species, one genus"). If more than two species belonging to different genera remained, the number of genera was counted. If one genus (and multiple species) was present the genus was saved (e.g., *Hucho* sp. with the ambiguous assignments *Hucho bleekeri*, *Hucho hucho*, and *Hucho taimen*; "F3 - Multiple species of one genus"). Lastly, if more than one genus remained and no dominant taxon was present, the taxonomic assignment was trimmed to the most recent common taxon ("F4 - Multiple genera"). Both the taxonomy and read tables were then converted to TaXon tables (Supplementary Material 2) for downstream analyses in TaxonTableTools v1.4.7 (Macher et al. 2021a). To account for potential contamination the sum of reads in the negative controls of each OTU was subtracted from the number of reads for the respective OTU of each sample ('Negative control subtraction' tool). Subsequently, all tables were filtered for fish and lamprey species (Supplementary Material 3). Here, all OTUs with a ≥97% similarity but without species

assignment were manually checked and adjusted if e.g., a hybrid or erroneous entry was preventing a species assignment (Supplementary Material 4). If the taxonomy was ambiguous due to the assignment to geographically clearly separated species with equal similarity values, the species which is reported from the area was selected. The distribution information was collected from the gbif database (www.gbif.org).
Analyses were performed using custom python scripts and results were visualized using the plotly package (https://plot.ly). For all primer pairs, the OTU and read proportions of target taxa (i.e., fish and lamprey) and bycatch taxa (i.e., all other taxa) were calculated. Additionally, the number of ambiguous species-level OTUs and the number of occurrences of each flag was calculated. For all subsequent analyses the manually adjusted TaXon tables were used.
First, the relative read abundances (%) for all species present in the mock community (i.e., true positive species) and all non-target species (i.e., false-positive species) were calculated. Here, for each species the number of positive detections and the standard deviation of the relative read abundances across the five primer pairs were calculated. Second, Venn diagrams comparing the detected species of each primer pair to the original fish mock community composition were created. Additional Venn diagrams were created to compare the pre-adjusted TaXon tables. Oversplitting rates (i.e., number of additional OTUs) were calculated for all species and each primer pair. PCR replicates were investigated by calculating the mean, minimum, and maximum Jaccard index of all five technical replicates per primer pair. The log transformed number of reads and the log transformed DNA concentration (ng/µL) were plotted and Spearman rho coefficients were calculated. Also, the log transformed number of reads per species of MC2 were plotted against the log transformed reads per species of MC1 and a Spearman rho coefficient was calculated. Lastly, the number of taxonomically assigned OTUs and unique species per family and the number of false-positive and false-negative assignments for each primer pair were calculated and plotted in two heatmaps.

Results

According to the fishbase database (fishbase.org), 123 fish species are reported from Germany (occurrence categories: “endemic”, “introduced”, “native”, “not established”, “questionable”, and “stray”). Here we manually added the round goby (*Neogobius melanostomus*) and the rainbow trout (*Oncorhynchus mykiss*), as they are both invasive species in Germany, but were not present in the fishbase list. Consequently, our fish mock community of 45 Central European freshwater fish species represents about 36.6% of fish reported from Germany (Supplementary Table 2). In detail, our mock community accounts for 50% of “native”, 26% of “introduced”, 22.2% of “questionable”, and 8% of “not established” fish species in Germany.
Sequencing yielded a total of 8,254,293 raw reads across all primer pairs. In total, 7,745,593 quality-filtered reads were clustered into 140 (tele02), 105 (MiFish-U), 120 (12SV5), 111 (SeaDNA-mid), and 142 (LH16S) OTUs, respectively. Nearly all primer pairs showed little amplification of non-fish OTUs (between 96 to 98% fish OTUs), except for the SeaDNA-mid primer pair, which exhibited 50% non-fish OTUs, (Figure 1A). However, only few reads were assigned to non-target OTUs for all primer pairs (between 98.2 and 100% fish OTUs; Figure 1B).
The proportions of flagged taxonomic assignments varied between the five different primer pairs. Here, both the MiFish-U and tele02 primer pairs had the highest proportion of supported species-level OTUs (both 60%), followed by the SeaDNA-mid (49%), 12SV5 (46%), and LH16S (41%) primer pairs (Figure 1C). The first flag (‘Two

species, one genus') was most prominent in the SeaDNA-mid (25%) and least prominent in the 12SV5 primer pair (10%). For the second flag ('Multiple species of one genus') again the SeaDNA-mid showed the highest proportions (16%), while both the tele02 and MiFish-U primer pairs had the fewest cases (5%). Furthermore, the 12SV5 primer pair showed the highest proportion of the third flag ('Dominant taxon') with 27% assigned species-level OTUs, while again the tele02 primer pair showed the fewest (9%). The SeaDNA-mid primer did not have any cases of flag four ('Multiple genera'), while the LH16S primer pair had the most (9%). Overall, the most abundant ambiguous assignment was *Leuciscus idus*/*leuciscus* (10 total occurrences), followed by *Sander canadensis*/*lucioperca* (8), *Blicca bjoerkna* (7), *Proterorhinus semilunaris*/*marmoratus* (6), and *Cyprinus carpio* and *Hucho* sp. with each 5 cases (Supplementary Table 3). Overall, the genera *Leuciscus* and *Sander* showed the highest number of ambiguous taxonomic assignments (14 and 13, respectively).

After removal of bycatch taxa and curation of ambiguous taxonomic assignments, the 12SV5 primer pair (45) included most species, followed by LH16S (40), tele02 (39 species), MiFish-U (37), and SeaDNA-mid (36). In comparison to the original mock community fish species composition, the tele02 dataset showed the highest congruence (2 false-positive species, 37 true positive, and 8 false-negative), followed by the MiFish-U (2, 35, 10) and SeaDNA-mid (3, 33, 12). Both the 12SV5 (18, 27, 18) and LH16S primer pair (17, 23, 22) were less congruent to the original mock community composition (Figure 2). The 12SV5 and LH16S primer pairs resulted in OTUs assigned to several marine fish taxa, which were not part of the mock community, including Acanthuridae (surgeon fishes), Kyphosidae (sea chubs), Ophidiidae (cusk-eel), Peristediidae (armoured sea robins), Pholidae, and Zoarcidae (eelpouts; Table 2). Regarding the number of false-positive and false-negative assignments per family, the LH16S primer pair showed high incongruencies to the mock community, particularly for the Leuciscidae (4 false-positive / 10 false-negative) and Percidae (2/2). Similarly, the 12SV5 primer pair had various false-positive and false-negative assignments for the Leuciscidae (6/5), Cyprinidae (4/0), or Gobionidae (3/1). The SeaDNA-mid primer showed only a moderate number of incorrect assignments in the Leuciscidae (2/6). Lastly, the tele02 and MiFish-U primer pairs were overall the least prone to false-positive assignments and only showed false-positive assignments in Leuciscidae (*Leucicus aspius*) and Salmonidae (*Parahucho perryi* and *Brachymstax lenok*).

As a measure of primer bias the standard deviation of relative read abundances was across primer pairs. Here the standard deviation varied between the primer pairs ranging from an average of < 0.01% (*Barbatula barbatula*, *Leucaspius delineatus*, *Neogobius melanostomus*, *Phoxinus phoxinus*, and *Romanogobio albipinnatus*) to a maximum of 7.5% (*Pungitius pungitius*; Table 3A). While most species were detected with at least four primer pairs (29 mock community species), 10 species were detected with three or less primer pairs. In total, six species were not detected by any of the primer pairs, namely *Cottus gobio*, *Gymnocephalus schraetser*, *Lampetra fluviatilis*, *Rutilus pigus*, *Umbra krameri*, and *Zingel zingel*. Most false-positive species were unique to one primer pair (34 of 37 species; Table 3B), while only three species were detected with two or more primer pairs, namely *Leuciscus aspius* (4 occurrences), *Pungitius platygaster* (2), and *Umbra pygmaea* (2).

In total, 48 cases of oversplitting (in our case species with more than one OTU assigned) were observed (Supplementary Table 4). Most over-split species-level assigned OTUs were found with the tele02 primer pair (12), while all other primer pairs showed 9 cases of oversplitting. The highest oversplitting rate was observed in *Gymnocephalus cernua* (7-fold OTU to species ratio, tele02 primer pair) and *Tinca tinca* (7-fold, 12SV5). While no over-split species was found in all five or even four of

the primer pairs, six species were over split in three primer pairs (i.e., *Abramis brama*, *Blicca bjoerkna*, *Ctenopharyngodon idella*, *Gymnocephalus cernua*, *Hucho hucho*, and *Sander lucioperca*).
PCR replicates were highly consistent for all investigated primer pairs. The 12SV5 primer pair showed the highest reproducibility (mean Jaccard similarity of 0.99), followed by LH16S (0.98), SeaDNA-mid (0.96), tele02 (0.96), and MiFish-U (0.95). No correlations between log transformed input DNA concentration (ng/µL) and log transformed reads of the second mock community (MC2) were found for most of the primer pairs (Spearman's rho between 0.21 and 0.34, p≥0.05) except for the MiFish-U primer pair (p≥0.05), which showed a moderate positive correlation (Spearman's rho = 0.41) (Supplementary Figure 2). However, when comparing the number of log transformed reads per species between MC1 and MC2, significant correlations for the tele02 (rho=0.79, p≤0.05), MiFish-U (0.78, p≤0.05), 12SV5 (0.85, p≤0.05), SeaDNA-mid (0.81, p≤0.05), and LH16S (0.6, p≤0.05) primer (Supplementary Figure 3) were found.

Discussion

Discriminatory power, and reproducibility

Our primer evaluation based on mock communities of 45 European freshwater fish species confirmed the previously reported high discrimination power for two primer pairs (MiFish-U and tele02) belonging to the MiFish primer group (Bylemans et al. 2018a, Taberlet et al. 2018, Collins et al. 2019, Polanco F. et al. 2021). The tele02 primer pair (a modified version of the MiFish-U primer pair) performed particularly well in our study and clearly showed the highest species specificity and discriminatory power for European freshwater species. Until now the tele02 primer pair was evaluated *in silico* (Taberlet et al. 2018, Collins et al. 2019) as well as for water samples from Beijing, where it exhibited outstanding detection success of fish diversity in comparison with other fish-specific primers tested (Zhang et al. 2020). Accordingly, our results show that the tele02 primer pair recovered most true-positive species while producing the lowest number of false-positive and negative detections. From all primer pairs tested in this and other studies, the tele02 primer pair is arguably the best currently available choice for fish eDNA metabarcoding of European freshwater fish. While the SeaDNA-mid primer pair, targeting the COI gene, showed comparable good discriminatory power (i.e., true-positive detections), the co-amplification of non-fish taxa with this primer pair might be of concern. The fish mucus likely accumulates eDNA molecules and thus also contains DNA from other organisms than the fish itself. Here, the SeaDNA-mid primer pair was the only primer pair that showed high numbers of non-target OTUs. While non-target OTUs were observed in low read abundances for the mock communities, co-amplification issues could be more pronounced when applying the SeaDNA-mid primer pair on environmental samples. Here, comparably deeper sequencing depths might be required to detect all present fish species in an environmental sample with more non-target DNA, which would reduce the cost-efficiency per sample. The remaining two primer pairs 12SV5 and LH16S were designed to generally amplify vertebrate DNA (Kitano et al. 2007, Riaz et al. 2011, Hänfling et al. 2016, Harper et al. 2019). We decided to include these primer pairs since they have the potential for more holistic monitoring approaches, e.g., targeting the whole vertebrate community associated to a freshwater habitat (Pertoldi et al. 2021, Dou et al. 2023). However, the broader target range resulted in a drastically lower detection rate of fish species for the vertebrate primer pairs.

Overall, all primer pairs generated highly reproducible taxa lists among the PCR replicates for the fish mock communities. However, this reproducibility might not be achieved for environmental samples. Here, a generally lower reproducibility is expected and it is recommended to consider sufficient field and laboratory replicates to maximise species detection and minimize stochastic sampling effects (Sato et al. 2017, Bylemans et al. 2018b, Macher et al. 2021b, Rojahn et al. 2021). Particularly the SeaDNA-mid primer pair might suffer from lower reproducibility for environmental samples due to the strong co-amplification.

Primer bias

Generally, several cases of potential primer amplification bias were observed, where certain species exhibited over-proportional read abundance in comparison to other primers. Despite our mucus samples most likely did not solely contain DNA of the target species, the amount of target input DNA was equally biased for all primer pairs. Thus, under optimal conditions without amplification biases, equal relative read abundances per species are to be expected. However, we observed distinct differences in the relative read abundances per primer pair. While no general amplification bias trend was observed, several species showed significantly higher read abundances for one of the primer pairs, such as *Perca fluviatilis* (12SV5: 12.65% to an average of other primer pairs of 1.04%), *Hucho hucho* (SeaDNA-mid: 17.36% to 2.5%), or *Pungitius pungitius* (SeaDNA-mid: 17.49 to 0.65%). Using reads as a proxy for fish biomass has been addressed in various studies (Takahara et al. 2012, Kelly et al. 2019, Muri et al. 2020). However, next to uncertainties about the fate and state of eDNA in the environment, primer-specific PCR amplification biases as observed in our results can drastically affect read counts, depending on the choice of primer pair. This can lead to ambiguous or false conclusions about biomass estimates for the investigated fish community. While trends might exist, the interpretation of reads as proxy for biomass should be taken with care.

Another issue was the detection of multiple OTUs for certain species. While this does not affect the analysis when working on species level (i.e., OTUs of the same species are merged), the OTU alpha diversity is artificially inflated. Here we observed that particularly the Leuciscidae showed drastically higher numbers of OTUs than species. If the analysis of OTUs is of particular interest, this issue can be tackled by e.g., using a post-clustering curation algorithm, such as LULU filtering (Frøslev et al. 2017), which should give more reliable biodiversity estimates e.g., when taxonomic references are lacking.

False-negative assignments

The here used Midori2 database is a curated version of the larger GenBank database and can be used as a reliable source for taxonomic assignment of fish OTUs. All species present in the mock community have reference sequences available for at least one genetic marker. However, seven species were not detected at all.

Amongst these was *Cottus gobio*, a common fish species in Central Europe for which 117 reference sequences comprising all three investigated markers are deposited in GenBank. Although a taxonomic assignment was possible, no primer pair detected *C. gobio* in the mock communities. Since this species is frequently detected with eDNA metabarcoding from various sites and samples (Macher et al., unpublished data; tele02 primer pair), it is likely that the *C. gobio* sample itself was the reason for the false-negative detection, as it might not have contained *C. gobio* DNA in sufficient

concentration or due to sampling or laboratory errors (e.g., specimen misidentification, swab inaccurately taken or DNA degradation).
The stripped ruffe (*Gymnocephalus schraetser*) only has 13 reference sequences available in the Midori2 database, none of which is a 12S sequence. Consequently, the lack of reference for the 12S marker prevents a species level assignment for the tele02, MiFish-U, and 12SV5 primer pairs. However, all 12S primer pairs included OTUs assigned to *Gymnocephalus* that were trimmed to genus level due to low reference similarity threshold (< 97%). While no primer pair was able to detect *G. schraetser*, the SeaDNA-mid COI primer contained one ambiguous OTU assigned to *G. schraetser/cernua*. Thus, it remains unclear if the stripped ruffe can be distinguished from *G. cernua*, using eDNA primer pairs.
Furthermore, various species are known to be indistinguishable with the short target fragment lengths used for eDNA metabarcoding. Particularly the two common lamprey species *Lampetra fluviatilis* and *L. planeri* could not be distinguished with any of the used primer pairs. The species status of these two 'sister species' has puzzled scientists for decades and while a genome-wide divergence can be observed (Mateus et al. 2013), they are known to share mitochondrial haplotypes (Espanhol et al. 2007). Considering that most eDNA primer pairs target short mitochondrial fragments of approximately 180 bp, a distinction of these species with eDNA metabarcoding will most likely not be possible in the foreseeable future.
The zingel (*Zingel zingel*) was not detected by any primer pair despite the availability of various whole genome shotgun, COI, and sRNA reference sequences in the Midori2 database. The closely related Danube zingel (*Zingel streber*) has various COI and 16S reference sequence available and was detected by the SeaDNA-mid and LH16S primer pair. Thus, the most likely explanation for the absence of *Zingel zingel* is errors in sampling or laboratory handling that led to the sample failure.

Ambiguous assignments

In several instances, the distinction between true-positive, false-positive, and false-negative assignments was very narrow. For several species, we observed misidentification with closely related species, which resulted in false-positive and false-negative assignments in single cases. For example, a species that was not detected by any primer pair is *Rutilus pigus*, the Danube roach. This species is closely related to the cactus roach (*R. virgo*) which was once considered a subspecies (*Rutilus pigus* subsp. *virgo* (Heckel, 1852)) and occurs in the same habitats. However, since molecular data showed that *R. pigus* and *R. virgo* are separate species (Pourshabanan et al. 2022), either the reference taxonomy is incorrect, which can occur in a non-curated database such as Genbank, or the specimen that was sampled for the mock community was actually *R. virgo*. For both species COI reference sequences are available in the Midori2 database, however, no 12S or 16S reference sequences are present. Here, the tele02 (1 OTU, 96.5%) and MiFish-U (1 OTU, 96.0%) both detected OTUs assigned to the genus *Rutilus*, besides *Rutilus rutilus* (which was present in the mock community), rendering these false-negative assignments as result of missing 12S reference sequences. Furthermore, the false-positive *Rutilus virgo* assignment by the SeaDNA primer pair was most likely not a false-positive detection due to primer bias or lack of reference sequences but rather a lack of species name harmonisation or misidentification.
For the European mudminnow (*Umbra krameri*), only 10 reference sequences (for 12S, COI, 16S or whole genome) are available in GenBank and it was not detected by any primer pair in our study. However, the SeaDNA-mid and 12SV5 primer pairs false-

positively detected the closely related species *Umbra pygmaea* and the teleo2, MiFish-U, and LH16S detected *Umbra limi*/*pygmaea*. Both *U*. *limi* (Central mudminnow) and *U*. *pygmaea* (Eastern mudminnow) are native to North America, and particularly the latter has been introduced to Western and Central Europe. One explanation for the incorrect assignments could be a misidentification of the specimen from which the mucus sample was taken. If so, the specimen identified as European mudminnow was truly an invasive Eastern mudminnow. This case should be further investigated since the European mudminnow is listed as 'vulnerable' (IUCN Red List of Threatened Species in 2010) and should ideally be distinguishable from the invasive Eastern mudminnow with eDNA metabarcoding.

Furthermore, we observed several cases of "difficult" taxonomic assignments. Here, particularly OTUs assigned to the genera *Hucho*, *Sander* and *Leuciscus* caused ambiguities. The Danube salmon (*Hucho hucho*) was initially only detected by the SeaDNA-mid and LH16S primer pairs. The three 12S primer pairs faced ambiguities caused by hits to the Sichuan taimen (*Hucho bleekeri*) and the Siberian taimen (*Hucho taimen*), which all share identical 12S sequences. However, since the Danube salmon is the only present species of the genus *Hucho* in Central Europe, *H*. *bleekeri* and *H*. *taimen* were ruled out for the tele02, MiFish-U and 12SV5 primer pairs. Similarly, the pikeperch (*Sander lucioperca*) is geographically clearly separated from the sauger (*S*. *canadensis*), but the two species are not genetically distinguishable with the investigated markers, leading to flag 1 ambiguities ("Two species, one genus"). In this case, however, based on the current distribution ranges, one can account for this ambiguity, similarly to the Danube salmon. Nevertheless, if one of the *Hucho* or *Sander* species were to be introduced to Central Europe, not all primer pairs could distinguish the native species, which could be of concern for invasive species monitoring. The common dace (*Leuciscus leuciscus*) and ide (*L*. *idus*), however, are highly prone to causing flag 1 ambiguities. This can be caused by several reasons: for instance, species of the family Leuciscidae are known to commonly hybridize, such as the bleak (*Alburnus alburnus*) and chub (*Leuciscus cephalus*) (Wheeler 1978) or chub and roach (*Rutilus rutilus*) (Wheeler and Easton 1978). This can lead to mitochondrial introgression, causing reference sequences of different species to be identical. Another reason is the wide distribution of common dace across Europe and its habitus typical for the family Leuciscidae. This can result in false species identification, that is propagated to incorrect database entries, which ultimately can lead to ambiguous assignments. Here, a sophisticated curation of the Midori2 database, or the usage of a custom reference database, including reference sequences from known source, might help to reliably distinguish *L*. *leuciscus* and *L*. *idus*. Another reason for false-negative assignments may occur in the automated taxonomic assignment of OTUs due to unclear species status or the use of synonyms. For example, we were aware from previous eDNA metabarcoding datasets that *Rhodeus amarus* and *R*. *sericeus* are used synonymously and we corrected our dataset for this issue (*Rhodeus amarus*/*sericeus*).

While in this study we used the Midori2 database, which is a curated version of the Genbank database, another widely used reference library for mitochondrial sequences is the MitoFish database (Sato et al. 2018). While reference sequences for most fish are available in the MitoFish database, some species cannot be assigned due to the absence of e.g., whole genome sequences (e.g., *Romanogobio albipinnatus*). Additionally, the comparably lower overall number of reference sequences might be of concern in light of intraspecific variation and could lead to false-negative assignments.

False-positive assignments

The detection of false-positives is of particular concern since it drastically reduces the robustness of taxa lists. Particularly the more general vertebrate primer pairs were prone to produce comparably high numbers of false-positive assignments. Here, 12SV5 and LH16S were the only datasets that included marine fish taxa, which were not present in the mock community of Central European freshwater fish. Since no marine samples have been processed in this laboratory, cross-contaminations can be ruled out. The most likely explanation for these false-positive assignments is the placement of target fragments in conserved regions to amplify a broader taxonomic range (e.g., vertebrates). However, this will ultimately decrease the taxonomic resolution for specific taxa within that group (e.g., fish species). For the here investigated primer pairs most likely the short fragment length (12SV5 primer pair; 106 bp) or the fragment location for the LH16S primer pair the number of substitutes is too low for reliable fish identification.

Furthermore, incorrect assignments of closely related species were observed for the more general vertebrate primer pairs 12SV5 and LH16S. These included the Asian *Chondrostoma prespense* instead of *C. nasus*, the North American *Thymallus arcticus* instead of *T. thymallus*, or *Pungitius platygaster* instead of *P. pungitius*. Again, the conserved regions amplified by the 12SV5 and LH16S primer pairs could have led to these false-positive assignments. Particularly phylogenetically 'young' species that have not been separated long and e.g., share mitochondrial haplotypes (Espanhol et al. 2007) or closely related species that exhibit hybridisation and introgression (Hata et al. 2019, De Santis et al. 2021) are potentially not distinguishable with short and conserved target fragments.

However, also the tele02, MiFish-U and SeaDNA-mid primer pairs showed false-positive assignments. Even though the asp (*Leuciscus aspius*) was not included in the mock community, it was detected by all three primer pairs. Since it was consistently detected by the tele02 (2 OTUs, 98% similarity to reference sequence, 8578 reads, 10/10 samples), MiFish-U (2 OTUs, 98%, 7246 reads, 10/10 samples), and the SeaDNA primer pair (1 OTU, 100%, 156 reads, 9/10 samples), the most likely explanation for the detection of *L. aspius* is be a misidentification during sampling (e.g., another closely related cyprinid species). Another explanation is that the DNA of one species can be found in the mucus of another species' mucus, which could potentially also contain eDNA traces from other fish that were present during sampling. Another case of false-positive detection is the Japanese huchen (*Parahucho perryi*), which was detected in low read abundances by the tele02 primer pair (1 OTU, 98%, 114 reads, 9/10 samples). The Japanese huchen is not recorded from Central Europe but is related to both the huchen (*Hucho hucho*) and brown trout (*Salmo trutta*), which were both present in the mock community. The most likely explanation is that this false-positive assignment originates from huchen or brown trout DNA that is amplified by the tele02 primer pair followed by misassignment. The low read abundance observed in this dataset and its occurrence in combination with the brown trout in other eDNA metabarcoding datasets using the tele02 primer pair (Macher et al., unpublished data) hints towards a systematically false-positive detection of the Japanese huchen in the presence of the brown trout. A similar case is the detection of the Asian sharp-snouted lenok (*Brachymystax lenok*) with the MiFish-U primer pair, which is a salmonoid species related to trouts.

While most false-positive assignments can be easily corrected, primer pairs that are not prone to false-positive assignments, such as the tele02, MiFish-U and the SeaDNA-mid primer pairs, are to be preferred over the more general 12SV5 and

LH16S primer pairs when investigating fish communities based on eDNA metabarcoding.

Conclusion

In conclusion, our study highlights how the choice of primer has a major effect on the outcome of eDNA metabarcoding analysis. The tele02 primer pair proved to be best choice for eDNA metabarcoding of Central European freshwater fish, showing the highest discriminatory power and good reproducibility with fewest false-positive and false-negative detections of the here tested primer pairs. We also observed that gaps in reference libraries can still lead to false-negative detections and thus should be addressed. Through careful selection of the primer pair, laboratory protocol, and bioinformatic pipeline, eDNA metabarcoding is becoming an increasingly reliable tool for fish monitoring.

References

Bohmann K, Elbrecht V, Carøe C, Bista I, Leese F, Bunce M, Yu DW, Seymour M, Dumbrell AJ, Creer S (2022) Strategies for sample labelling and library preparation in DNA metabarcoding studies. Molecular Ecology Resources 22: 1231–1246. https://doi.org/10.1111/1755-0998.13512

Boivin-Delisle D, Laporte M, Burton F, Dion R, Normandeau E, Bernatchez L (2021) Using environmental DNA for biomonitoring of freshwater fish communities: Comparison with established gillnet surveys in a boreal hydroelectric impoundment. Environmental DNA 3: 105–120. https://doi.org/10.1002/edn3.135

Buchner D, Macher T-H, Leese F (2022) APSCALE: advanced pipeline for simple yet comprehensive analyses of DNA Meta-barcoding data. Bioinformatics: btac588. https://doi.org/10.1093/bioinformatics/btac588

Buchner D, Macher T-H, Beermann AJ, Werner M-T, Leese F (2021) Standardized high-throughput biomonitoring using DNA metabarcoding: Strategies for the adoption of automated liquid handlers. Environmental Science and Ecotechnology 8: 100122. https://doi.org/10.1016/j.ese.2021.100122

Bylemans J, Gleeson DM, Hardy CM, Furlan E (2018a) Toward an ecoregion scale evaluation of eDNA metabarcoding primers: A case study for the freshwater fish biodiversity of the Murray–Darling Basin (Australia). Ecology and Evolution 8: 8697–8712. https://doi.org/10.1002/ece3.4387

Bylemans J, Gleeson DM, Lintermans M, Hardy CM, Beitzel M, Gilligan DM, Furlan EM (2018b) Monitoring riverine fish communities through eDNA metabarcoding: Determining optimal sampling strategies along an altitudinal and biodiversity gradient. Metabarcoding and Metagenomics 2: 1–12. https://doi.org/10.3897/mbmg.2.30457

Collins RA, Bakker J, Wangensteen OS, Soto AZ, Corrigan L, Sims DW, Genner MJ, Mariani S (2019) Non-specific amplification compromises environmental DNA

metabarcoding with COI. Methods in Ecology and Evolution 10: 1985–2001. https://doi.org/10.1111/2041-210X.13276

De Santis V, Quadroni S, Britton RJ, Carosi A, Gutmann Roberts C, Lorenzoni M, Crosa G, Zaccara S (2021) Biological and trophic consequences of genetic introgression between endemic and invasive Barbus fishes. Biological Invasions 23: 3351–3368. https://doi.org/10.1007/s10530-021-02577-6

Dou H, Wang M, Yin X, Feng L, Yang H (2023) Can the Eurasian otter (Lutra lutra) be used as an effective sampler of fish diversity? Using molecular assessment of otter diet to survey fish communities. Metabarcoding and Metagenomics 7: e96733. https://doi.org/10.3897/mbmg.7.96733

Elbrecht V, Braukmann TWA, Ivanova NV, Prosser SWJ, Hajibabaei M, Wright M, Zakharov EV, Hebert PDN, Steinke D (2019) Validation of COI metabarcoding primers for terrestrial arthropods. PeerJ 7: e7745. https://doi.org/10.7717/peerj.7745

Espanhol R, Almeida PR, Alves MJ (2007) Evolutionary history of lamprey paired species Lampetra fluviatilis (L.) and Lampetra planeri (Bloch) as inferred from mitochondrial DNA variation. Molecular Ecology 16: 1909–1924. https://doi.org/10.1111/j.1365-294X.2007.03279.x

Evans NT, Li Y, Renshaw MA, Olds BP, Deiner K, Turner CR, Jerde CL, Lodge DM, Lamberti GA, Pfrender ME (2017) Fish community assessment with eDNA metabarcoding: effects of sampling design and bioinformatic filtering. Canadian Journal of Fisheries and Aquatic Sciences. https://doi.org/10.1139/cjfas-2016-0306

Frøslev TG, Kjøller R, Bruun HH, Ejrnæs R, Brunbjerg AK, Pietroni C, Hansen AJ (2017) Algorithm for post-clustering curation of DNA amplicon data yields reliable biodiversity estimates. Nature Communications 8: 1188. https://doi.org/10.1038/s41467-017-01312-x

Fujii K, Doi H, Matsuoka S, Nagano M, Sato H, Yamanaka H (2019) Environmental DNA metabarcoding for fish community analysis in backwater lakes: A comparison of capture methods. PLOS ONE 14: e0210357. https://doi.org/10.1371/journal.pone.0210357

Hänfling B, Handley LL, Read DS, Hahn C, Li J, Nichols P, Blackman RC, Oliver A, Winfield IJ (2016) Environmental DNA metabarcoding of lake fish communities reflects long-term data from established survey methods. Molecular Ecology 25: 3101–3119. https://doi.org/10.1111/mec.13660

Harper LR, Lawson Handley L, Carpenter AI, Ghazali M, Di Muri C, Macgregor CJ, Logan TW, Law A, Breithaupt T, Read DS, McDevitt AD, Hänfling B (2019) Environmental DNA (eDNA) metabarcoding of pond water as a tool to survey conservation and management priority mammals. Biological Conservation 238: 108225. https://doi.org/10.1016/j.biocon.2019.108225

Hata H, Uemura Y, Ouchi K, Matsuba H (2019) Hybridization between an endangered freshwater fish and an introduced congeneric species and

consequent genetic introgression. PLOS ONE 14: e0212452. https://doi.org/10.1371/journal.pone.0212452

Hering D, Borja A, Jones JI, 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

Kelly RP, Shelton AO, Gallego R (2019) Understanding PCR Processes to Draw Meaningful Conclusions from Environmental DNA Studies. Scientific Reports 9: 12133. https://doi.org/10.1038/s41598-019-48546-x

Kitano T, Umetsu K, Tian W, Osawa M (2007) Two universal primer sets for species identification among vertebrates. International Journal of Legal Medicine 121: 423–427. https://doi.org/10.1007/s00414-006-0113-y

Kumar G, Reaume AM, Farrell E, Gaither MR (2022) Comparing eDNA metabarcoding primers for assessing fish communities in a biodiverse estuary. PLOS ONE 17: e0266720. https://doi.org/10.1371/journal.pone.0266720

Leray M, Knowlton N, Machida RJ (2022) MIDORI2: A collection of quality controlled, preformatted, and regularly updated reference databases for taxonomic assignment of eukaryotic mitochondrial sequences. Environmental DNA 4: 894–907. https://doi.org/10.1002/edn3.303

Macher T-H, Beermann AJ, Leese F (2021a) TaxonTableTools: A comprehensive, platform-independent graphical user interface software to explore and visualise DNA metabarcoding data. Molecular Ecology Resources 21: 1705–1714. https://doi.org/10.1111/1755-0998.13358

Macher T-H, Schütz R, Arle J, Beermann AJ, Koschorreck J, Leese F (2021b) Beyond fish eDNA metabarcoding: Field replicates disproportionately improve the detection of stream associated vertebrate species. Metabarcoding and Metagenomics 5: e66557. https://doi.org/10.3897/mbmg.5.66557

Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet. journal 17: 10–12.

Mateus CS, Stange M, Berner D, Roesti M, Quintella BR, Alves MJ, Almeida PR, Salzburger W (2013) Strong genome-wide divergence between sympatric European river and brook lampreys. Current Biology 23: R649–R650. https://doi.org/10.1016/j.cub.2013.06.026

McDevitt AD, Sales NG, Browett SS, Sparnenn AO, Mariani S, Wangensteen OS, Coscia I, Benvenuto C (2019) Environmental DNA metabarcoding as an effective and rapid tool for fish monitoring in canals. Journal of Fish Biology 95: 679–682. https://doi.org/10.1111/jfb.14053

683 Miya M (2022) Environmental DNA Metabarcoding: A Novel Method for Biodiversity
684 Monitoring of Marine Fish Communities. Annual Review of Marine Science 14:
685 161–185. https://doi.org/10.1146/annurev-marine-041421-082251

686 Miya M, Gotoh RO, Sado T (2020) MiFish metabarcoding: a high-throughput
687 approach for simultaneous detection of multiple fish species from
688 environmental DNA and other samples. Fisheries Science 86: 939–970.
689 https://doi.org/10.1007/s12562-020-01461-x

690 Miya M, Sato Y, Fukunaga T, Sado T, Poulsen JY, Sato K, Minamoto T, Yamamoto
691 S, Yamanaka H, Araki H, Kondoh M, Iwasaki W (2015) MiFish, a set of
692 universal PCR primers for metabarcoding environmental DNA from fishes:
693 detection of more than 230 subtropical marine species. Royal Society Open
694 Science 2: 150088. https://doi.org/10.1098/rsos.150088

695 Muri CD, Handley LL, Bean CW, Li J, Peirson G, Sellers GS, Walsh K, Watson HV,
696 Winfield IJ, Hänfling B (2020) Read counts from environmental DNA (eDNA)
697 metabarcoding reflect fish abundance and biomass in drained ponds.
698 Metabarcoding and Metagenomics 4: e56959.
699 https://doi.org/10.3897/mbmg.4.56959

700 Pertoldi C, Schmidt JB, Thomsen PM, Nielsen LB, de Jonge N, Iacolina L, Muro F,
701 Nielsen KT, Pagh S, Lauridsen TL, Andersen LH, Yashiro E, Lukassen MB,
702 Nielsen JL, Elmeros M, Bruhn D (2021) Comparing DNA metabarcoding with
703 faecal analysis for diet determination of the Eurasian otter (Lutra lutra) in
704 Vejlerne, Denmark. Mammal Research 66: 115–122.
705 https://doi.org/10.1007/s13364-020-00552-5

706 Polanco F. A, Richards E, Flück B, Valentini A, Altermatt F, Brosse S, Walser J-C,
707 Eme D, Marques V, Manel S, Albouy C, Dejean T, Pellissier L (2021)
708 Comparing the performance of 12S mitochondrial primers for fish
709 environmental DNA across ecosystems. Environmental DNA 3: 1113–1127.
710 https://doi.org/10.1002/edn3.232

711 Pont D, Valentini A, Rocle M, Maire A, Delaigue O, Jean P, Dejean T (2021) The
712 future of fish-based ecological assessment of European rivers: from traditional
713 EU Water Framework Directive compliant methods to eDNA metabarcoding-
714 based approaches. Journal of Fish Biology 98: 354–366.
715 https://doi.org/10.1111/jfb.14176

716 Pont D, Rocle M, Valentini A, Civade R, Jean P, Maire A, Roset N, Schabuss M,
717 Zornig H, Dejean T (2018) Environmental DNA reveals quantitative patterns of
718 fish biodiversity in large rivers despite its downstream transportation. Scientific
719 Reports 8: 10361. https://doi.org/10.1038/s41598-018-28424-8

720 Pourshabanan A, Moghaddam FY, Aliabadian M, Rossi G, Mousavi-Sabet H,
721 Vasil'eva E (2022) Molecular phylogeny and taxonomy of roaches (Rutilus,
722 Leuciscidae) in the southern part of the Caspian Sea.

723 Riaz T, Shehzad W, Viari A, Pompanon F, Taberlet P, Coissac E (2011) ecoPrimers:
724 inference of new DNA barcode markers from whole genome sequence

analysis. Nucleic Acids Research 39: e145–e145. https://doi.org/10.1093/nar/gkr732

Rognes T, Flouri T, Nichols B, Quince C, Mahé F (2016) VSEARCH: a versatile open source tool for metagenomics. PeerJ 4: e2584.

Rojahn J, Gleeson DM, Furlan E, Haeusler T, Bylemans J (2021) Improving the detection of rare native fish species in environmental DNA metabarcoding surveys. Aquatic Conservation: Marine and Freshwater Ecosystems 31: 990–997. https://doi.org/10.1002/aqc.3514

Sato H, Sogo Y, Doi H, Yamanaka H (2017) Usefulness and limitations of sample pooling for environmental DNA metabarcoding of freshwater fish communities. Scientific Reports 7: 14860. https://doi.org/10.1038/s41598-017-14978-6

Sato Y, Miya M, Fukunaga T, Sado T, Iwasaki W (2018) MitoFish and MiFish Pipeline: A Mitochondrial Genome Database of Fish with an Analysis Pipeline for Environmental DNA Metabarcoding. Molecular Biology and Evolution 35: 1553–1555. https://doi.org/10.1093/molbev/msy074

Schenekar T, Schletterer M, Weiss S (2020) eDNA als neues Werkzeug für das Gewässermonitoring – Potenzial und Rahmenbedingungen anhand ausgewählter Anwendungsbeispiele aus Österreich. Österreichische Wasser- und Abfallwirtschaft 72: 155–164. https://doi.org/10.1007/s00506-020-00656-x

Shu L, Ludwig A, Peng Z (2021) Environmental DNA metabarcoding primers for freshwater fish detection and quantification: In silico and in tanks. Ecology and Evolution 11: 8281–8294. https://doi.org/10.1002/ece3.7658

Taberlet P, Bonin A, Zinger L, Coissac E (2018) Environmental DNA: For Biodiversity Research and Monitoring. Oxford University Press, 268 pp.

Takahara T, Minamoto T, Yamanaka H, Doi H, Kawabata Z (2012) Estimation of Fish Biomass Using Environmental DNA. PLOS ONE 7: e35868. https://doi.org/10.1371/journal.pone.0035868

Wang S, Yan Z, Hänfling B, Zheng X, Wang P, Fan J, Li J (2021) Methodology of fish eDNA and its applications in ecology and environment. Science of The Total Environment 755: 142622. https://doi.org/10.1016/j.scitotenv.2020.142622

Wheeler A (1978) Hybrids of bleak, Alburnus alburnus, and chub, Leuciscus cephalus in English rivers. Journal of Fish Biology 13: 467–473. https://doi.org/10.1111/j.1095-8649.1978.tb03456.x

Wheeler A, Easton K (1978) Hybrids of chub and roach (Leuciscus cephalus and Rutilus rutilus) in English rivers. Journal of Fish Biology 12: 167–171. https://doi.org/10.1111/j.1095-8649.1978.tb04161.x

Zhang S, Zhao J, Yao M (2020) A comprehensive and comparative evaluation of primers for metabarcoding eDNA from fish. Methods in Ecology and Evolution 11: 1609–1625. https://doi.org/10.1111/2041-210X.13485

Acknowledgements

We thank Christoph Feick (LFU Bayern), Falko Wagner (IGF Jena), Christine Mosch (LAVES Niedersachsen), Franziska Neumann (LUNG MV), Gunnar Jacobs (EGLV) and all the collectors involved in the collection of the fish mucus samples. We thank all leeselab members who participated in the journal club and provided valuable feedback on the manuscript. This study was conducted as part of the GeDNA project, funded by the German Federal Environment Agency (Umweltbundesamt, FKZ 3719 24 2040).

Disclaimer

The collection of mucus samples is not categorized as animal experiment and did not require further authorisation. All sampling events were coordinated with local authorities. Fish specimens were solely caught during sampling events for monitoring campaigns and were handled by experts.

Data accessibility

The raw data were deposited at the European Nucleotide Archive (https://www.ebi.ac.uk/ena/browser/home) under the accession number PRJEB60937.

Author contributions

Till-Hendrik Macher: Conceptualization, Methodology, Formal analysis, Investigation, Visualization, Writing - original draft, Writing - review & editing; **Robin Schütz:** Conceptualization, Methodology, Formal analysis, Investigation, Writing - original draft, Writing - review & editing; **Atakan Yildiz:** Methodology, Writing - review & editing; **Arne J. Beermann:** Conceptualization, Validation, Supervision, Writing - review & editing; **Florian Leese:** Conceptualization, Resources, Supervision, Project administration, Funding acquisition, Writing - review & editing.

Figures and tables

Table 1: Primer pairs used for PCR amplification of the fish mock community.

Name	Gene	Primer pair	Forward sequence (5'-3')	Reverse sequence (5'-3')	Annealing temp.	Target length	Publication
tele02	12S	tele02_fw/tele02_rv	AAACTCGTGCCAGCCACC	GGGTATCTAATCCCAGTTTG	52 °C	~ 167 bp	Taberlet et al. 2018
MiFish-U	12S	MiFish-U_fw/MiFish-U_rv	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG	59 °C	~ 170 bp	Miya et al. 2015
SeaDNA-mid	COI	coi.175f/coi.345r	GGAGGCTTTGGMAAYTGRYT	TAGAGGRGGGTARACWGTYCA	53 °C	~ 130 bp	Collins et al. 2019
12SV5	12S	12S-V5f/12S-V5r	ACTGGGATTAGATACCCC	TAGAACAGGCTCCTCTAG	52 °C	~ 106 bp	Riaz et al. 2011
LH16S	16S	L2513/H2714	GCCTGTTTACCAAAAACATCAC	CTCCATAGGGTCTTCTCGTCTT	55 °C	~ 220 bp	Kitano et al. 2007

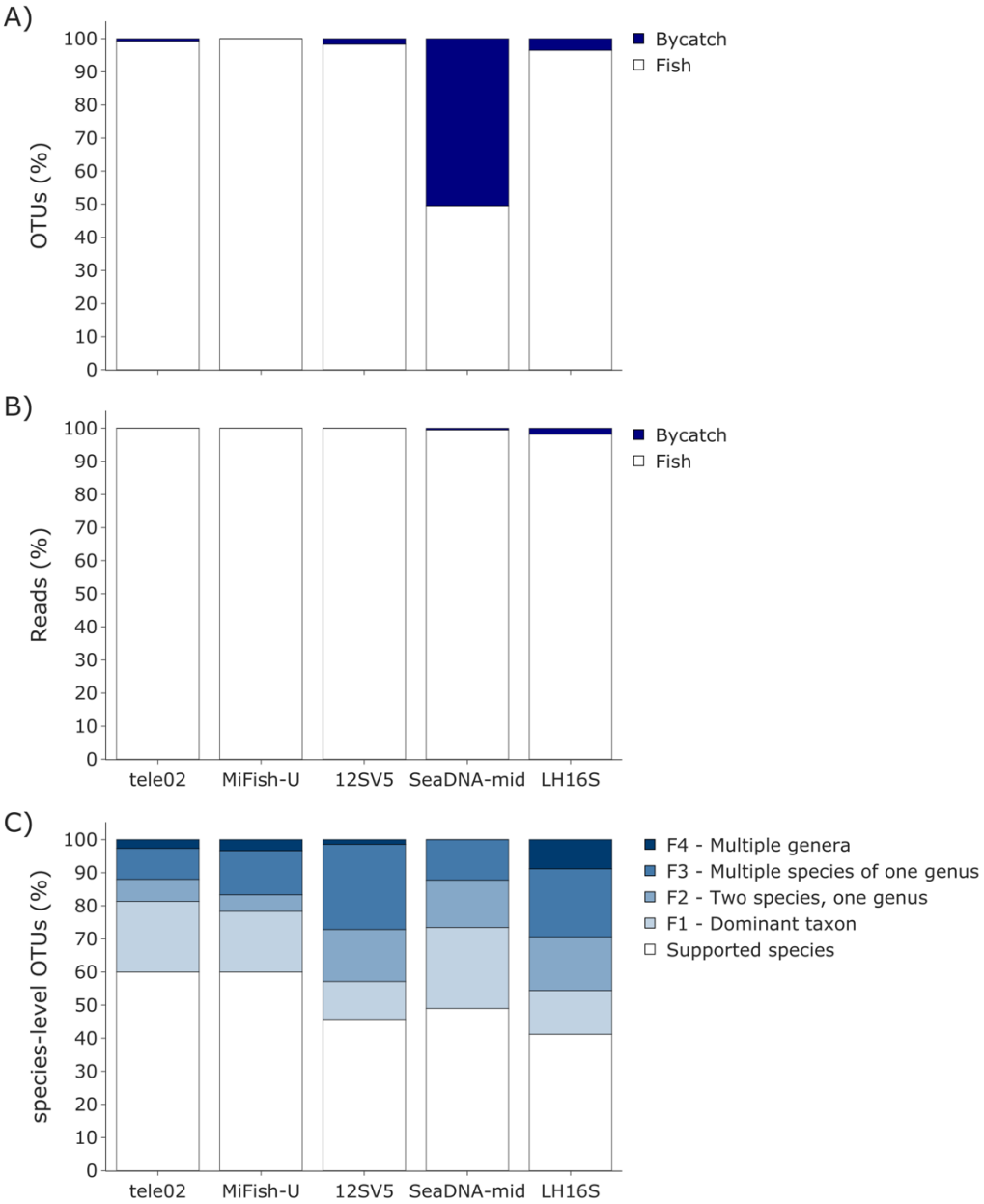


Figure 1: Proportions of fish and non-fish OTUs (A) and read proportions (B) detected with the five different primer pairs (A), and the proportions of ambiguous taxonomic assignments (flags 1-4) for all species-level OTUs (C), based on the pre-adjusted datasets.

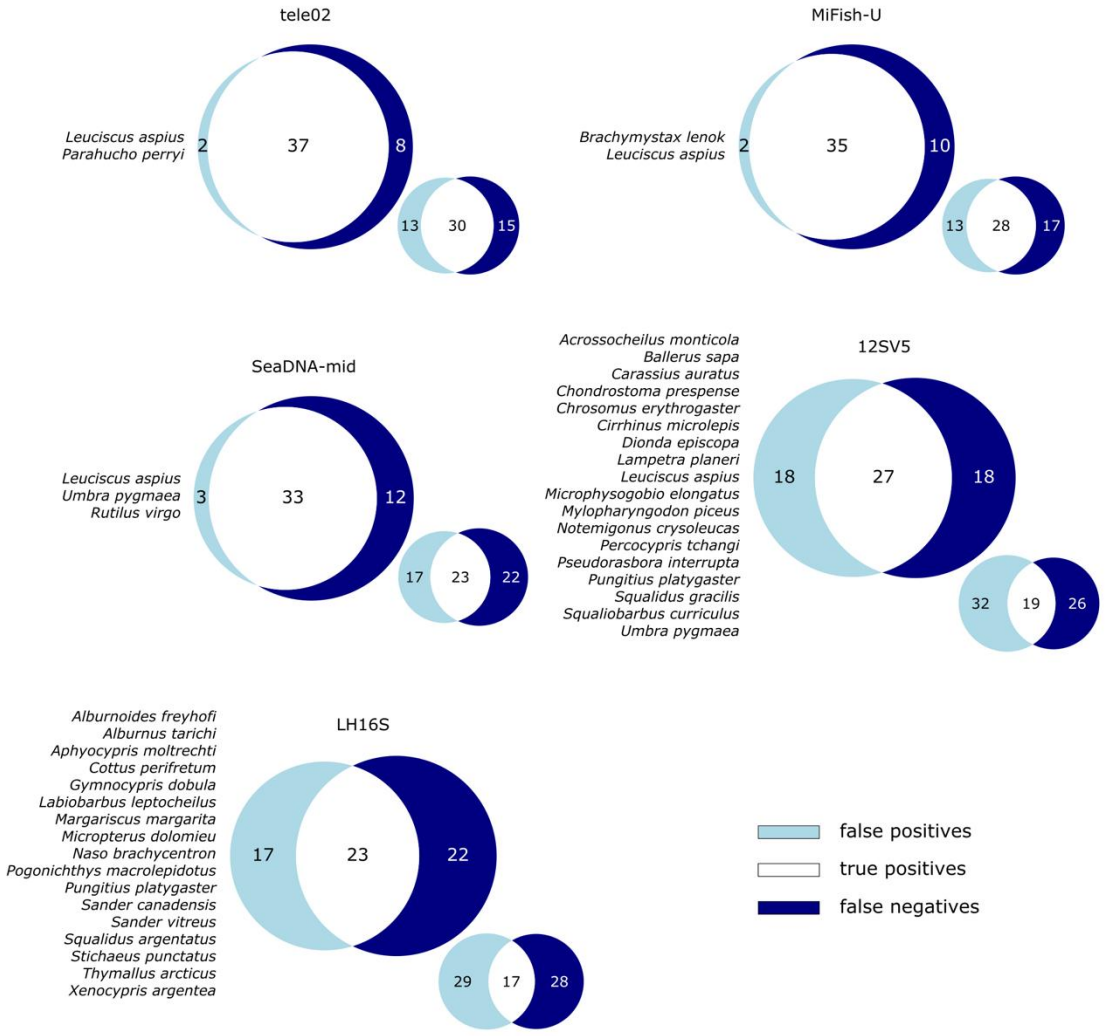


Figure 2: Comparison of the fish mock community species composition to the detected species with each primer pair for both the adjusted (large Venn diagrams) and the pre-adjusted datasets (small Venn diagrams). All species declared as false-positive detections are listed on the left-hand side of the respective Venn diagram.

Table 2: The overall number of fish species and the respective number of OTUs (in brackets) per family is shown in subplot A) for each primer pair. The number of false-positive (n/) and false-negative (/n) fish species detections compared to the original fish mock community composition is presented in subplot B).

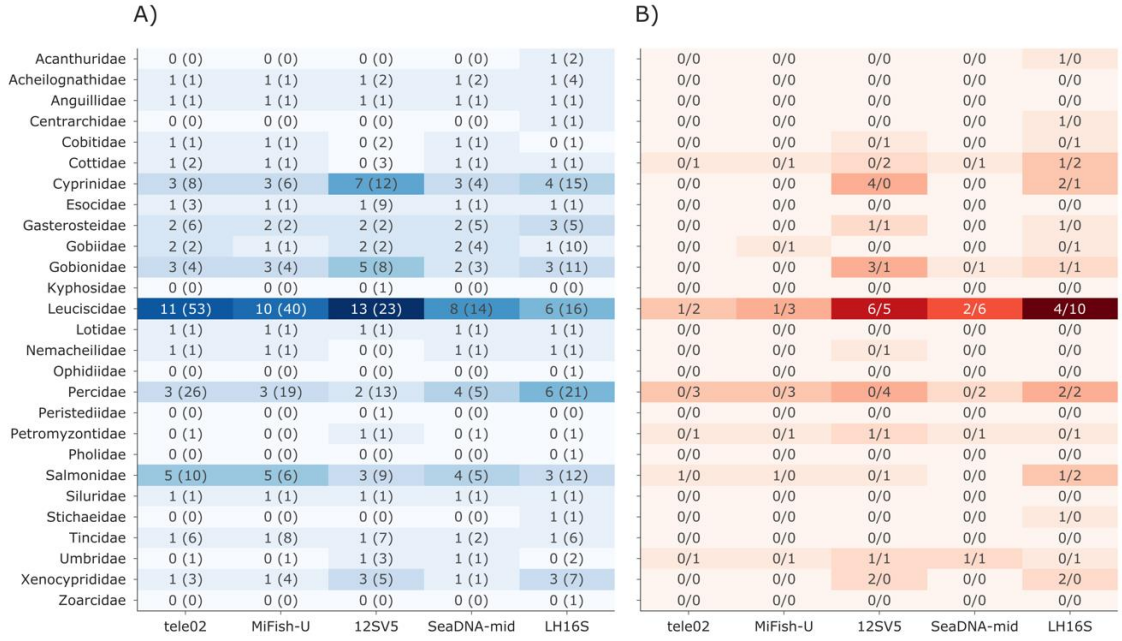


Table 3: Relative read abundances (%) for all detected fish and lamprey species of all five primer pairs, including all species present in the mock community (i.e., true positive species, A) and all non-target species (i.e., false-positive species, B). For each species the number of positive detections (occurrences) and the standard deviation (STDEV) were calculated.

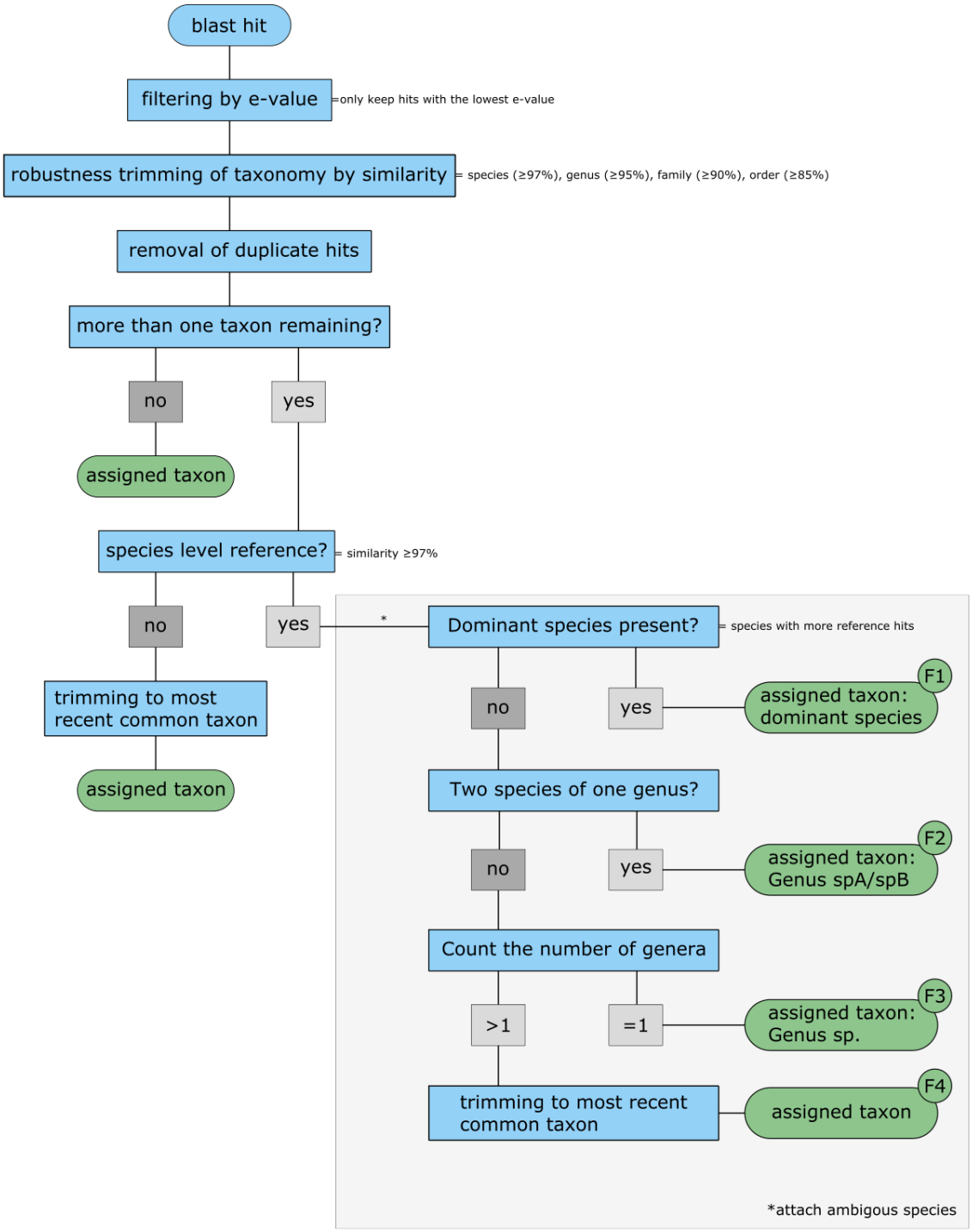
A)

True positive species	tele02	MiFish-U	12SV5	SeaDNA-mid	LH16S	Occurrences	STDEV
Anguilla anguilla	0.103	0.095	0.119	0.002	0.147	5	0.1
Silurus glanis	0.129	0.02	0.164	0.15	0.064	5	0.1
Barbus barbus	0.306	0.341	0.544	0.775	0.635	5	0.2
Thymallus thymallus	0.466	0.507	0.731	0.033	1.208	5	0.4
Tinca tinca	3.682	4.099	4.393	4.852	4.523	5	0.4
Gymnocephalus cernua	2.243	2.241	2.37	0.357	1.108	5	0.9
Ctenopharyngodon idella	1.15	0.153	1.39	2.576	0.667	5	0.9
Lota lota	0.264	0.253	0.329	3.746	0.191	5	1.6
Gobio gobio	0.179	0.192	0.24	3.816	0.141	5	1.6
Rhodeus sericeus/amarus	1.46	1.412	1.544	5.838	0.73	5	2.1
Carassius carassius	2.366	2.539	3.167	7.149	2.121	5	2.1
Rutilus rutilus	2.502	2.558	0.353	1.746	6.3	5	2.2
Esox lucius	6.232	5.7	5.131	0.049	0.432	5	3.0
Perca fluviatilis	1.485	1.321	12.65	0.18	1.174	5	5.2
Proterorhinus semilunaris	0.267	0.007	0.007	7.917	11.719	5	5.5
Hucho hucho	2.067	2.288	2.494	17.357	3.042	5	6.7
Pungitius pungitius	0.491	0.58	0.677	17.491	0.839	5	7.5
Phoxinus phoxinus	0.003	0.002	0.003	0	0.005	4	0.0
Barbatula barbatula	0.003	0.004	0	0.006	0.002	4	0.0
Oncorhynchus mykiss	0.149	0.192	0.294	0.54	0	4	0.2
Cyprinus carpio	0.372	0.363	0.506	0.019	0	4	0.2
Gasterosteus aculeatus	0.197	0.218	0	0.732	0.25	4	0.3
Squalius cephalus	1.072	1.09	0.09	0.007	0	4	0.6
Pseudorasbora parva	0.214	0.212	0	1.972	0.206	4	0.9
Chondrostoma nasus	2.333	2.508	0.047	0.408	0	4	1.3
Blicca bjoerkna	5.043	5.176	0.161	1.269	0	4	2.6
Abramis brama	19.839	19.776	25.78	17.748	0	4	3.5
Alburnus alburnus	6.532	6.837	10.588	1.189	0	4	3.9
Sander lucioperca	8.895	8.623	0	0.054	3.856	4	4.2
Romanogobio albipinnatus	0.003	0.005	0.006	0	0	3	0.0
Neogobius melanostomus	0.003	0	0.004	0.026	0	3	0.0
Salmo trutta	0.01	0.01	0	0.2	0	3	0.1
Misgurnus fossilis	0.039	0.038	0	0.821	0	3	0.5
Cottus rhenanus	1.065	1.029	0	0.005	0	3	0.6
Leucaspius delineatus	0.002	0.001	0	0	0	2	0.0
Zingel streber	0	0	0	0.158	0.333	2	0.1
Leuciscus idus	0	0.131	0	0	0	1	
Leuciscus leuciscus	0.053	0	0	0	0	1	
Scardinius erythrophthalmus	0.07	0	0	0	0	1	
Cottus gobio	0	0	0	0	0	0	
Gymnocephalus schraetser	0	0	0	0	0	0	
Lampetra fluviatilis	0	0	0	0	0	0	
Rutilus pigus	0	0	0	0	0	0	
Umbra krameri	0	0	0	0	0	0	
Zingel zingel	0	0	0	0	0	0	

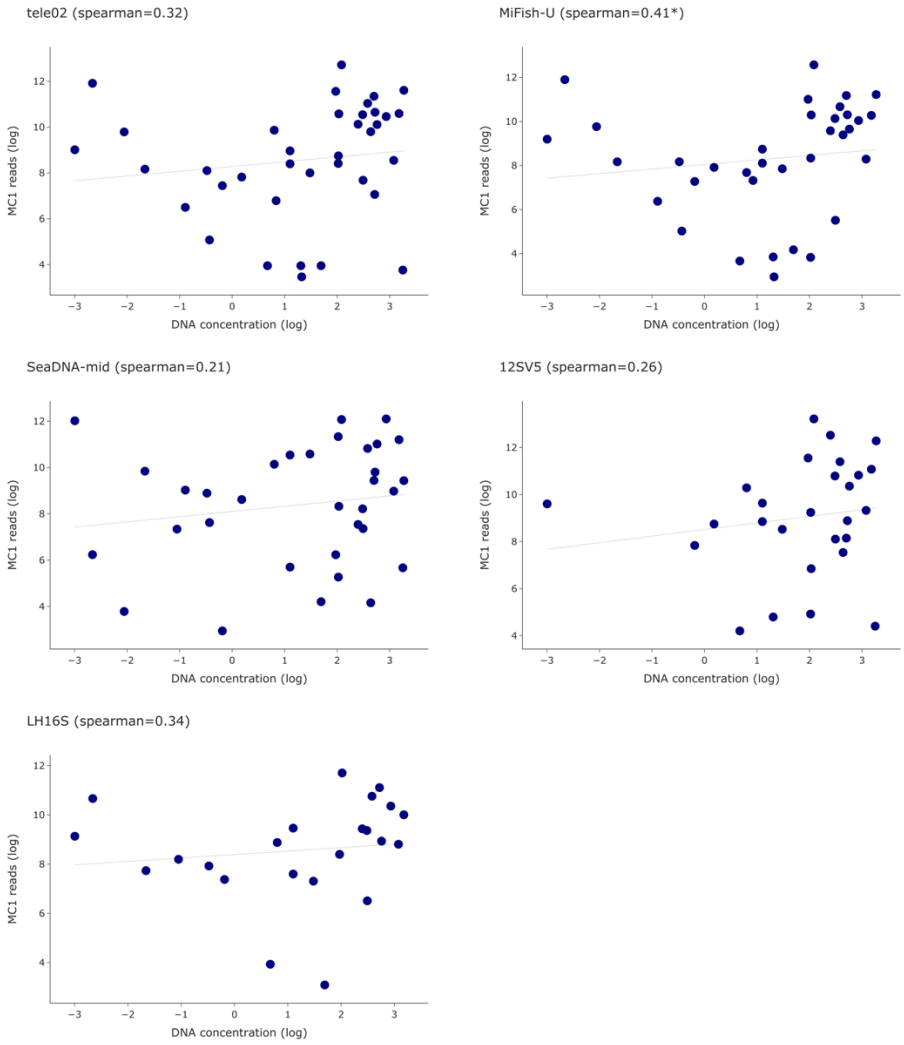
B)

False positive species	tele02	MiFish-U	12SV5	SeaDNA-mid	LH16S	Occurrences	STDEV
Leuciscus aspius	0.512	0.556	0.091	0.016	0	4	0.3
Pungitius platygaster	0	0	0.008	0	0.002	2	0.0
Umbra pygmaea	0	0	1.002	0.005	0	2	0.7
Acrossocheilus monticola	0	0	0.013	0	0	1	
Alburnoides freyhofi	0	0	0	0	37.686	1	
Alburnus tarichi	0	0	0	0	0.146	1	
Aphyocypris moltrechti	0	0	0	0	0.35	1	
Ballerus sapa	0	0	0.922	0	0	1	
Brachymystax lenok	0	0.005	0	0	0	1	
Carassius auratus	0	0	0.031	0	0	1	
Chondrostoma prespense	0	0	2.083	0	0	1	
Chrosomus erythrogaster	0	0	0.028	0	0	1	
Cirrhinus microlepis	0	0	0.017	0	0	1	
Cottus perifretum	0	0	0	0	0.964	1	
Dionda episcopa	0	0	0.118	0	0	1	
Gymnocypris dobula	0	0	0	0	0.05	1	
Labiobarbus leptocheilus	0	0	0	0	0.02	1	
Lampetra planeri	0	0	0.004	0	0	1	
Margariscus margarita	0	0	0	0	0.24	1	
Microphysogobio elongatus	0	0	0.007	0	0	1	
Micropterus dolomieu	0	0	0	0	0.284	1	
Mylopharyngodon piceus	0	0	0.057	0	0	1	
Naso brachycentron	0	0	0	0	0.687	1	
Notemigonus crysoleucas	0	0	0.159	0	0	1	
Parahucho perryi	0.007	0	0	0	0	1	
Percocypris tchangi	0	0	0.046	0	0	1	
Pogonichthys macrolepidotus	0	0	0	0	0.226	1	
Pseudorasbora interrupta	0	0	0.393	0	0	1	
Rutilus virgo	0	0	0	0.522	0	1	
Sander canadensis	0	0	0	0	0.01	1	
Sander vitreus	0	0	0	0	0.059	1	
Squalidus argentatus	0	0	0	0	0.034	1	
Squalidus gracilis	0	0	0.091	0	0	1	
Squaliobarbus curriculus	0	0	0.071	0	0	1	
Stichaeus punctatus	0	0	0	0	14.238	1	
Thymallus arcticus	0	0	0	0	0.157	1	
Xenocypris argentea	0	0	0	0	0.047	1	

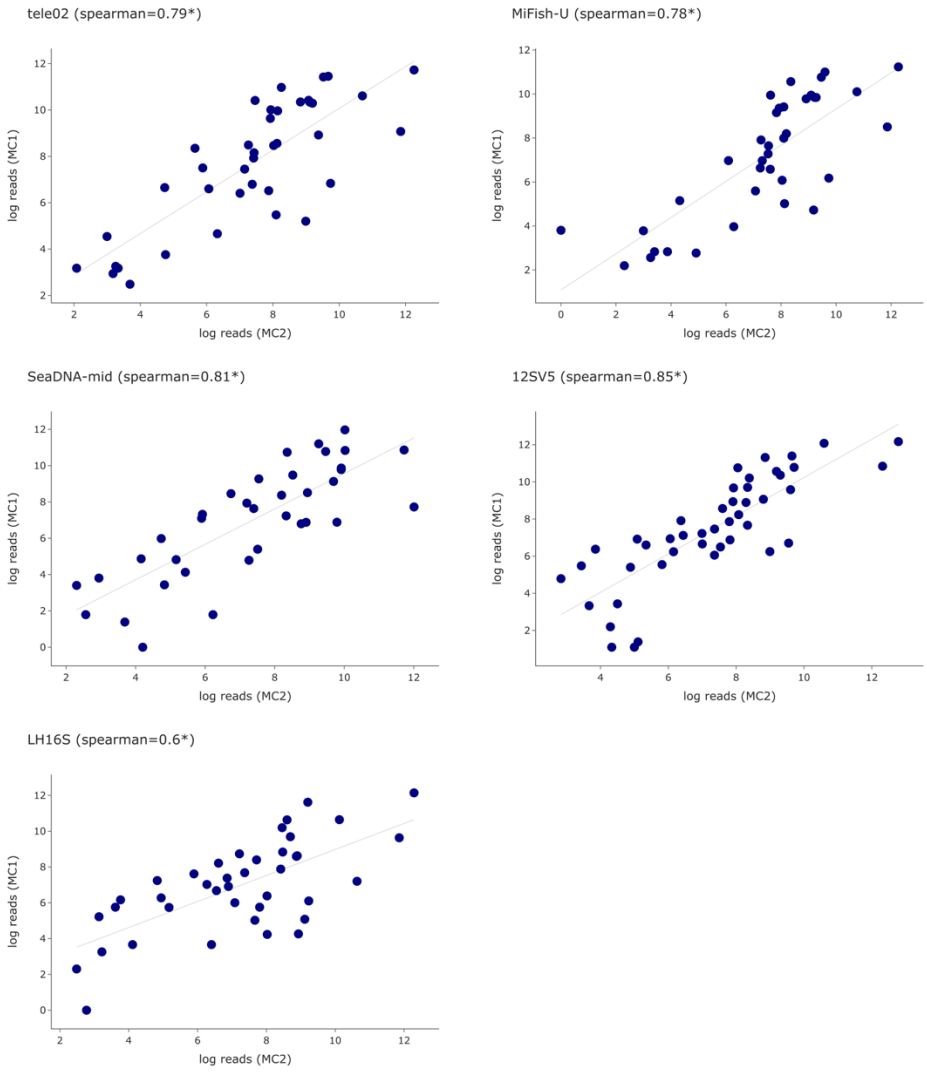
Supplementary figures and tables



Supplementary figure 1: Decision tree for taxonomic assignment implemented in APSCALE v1.2.0.



Supplementary figure 2: Pairwise comparison of the log-transformed reads of the non-normalized mock community (MC1) compared to the DNA concentration of each species.



Supplementary figure 3: Pairwise comparison of the log-transformed reads of the non-normalized mock community (MC1) compared to log-transformed reads of the normalized mock community (MC2) of each species.

Supplementary table 1: Sampled specimens and their respective species assignment collected for the fish mock community, extraction date, collection site, and concentration after DNA extraction.

Supplementary table 2: List of all species reported from Germany, their occurrence status, and their presence in the mock community (data from fishbase.org).

Supplementary table 3: List of all ambiguous assignments.

Supplementary table 4: List of over splitting rates per primer pair for each detected species.

Supplementary material

Supplementary material 1: Protocol for the adapted NucleoMag Tissue Kit.

Supplementary material 2: Unmodified TaXon tables of each primer pair.

867 **Supplementary material 3:** Processed TaXon tables of each primer pair (subtracted negative
868 controls and filtered for fish and lamprey taxa OTUs).
869
870 **Supplementary material 4:** Processed and manually curated TaXon tables of each primer pair.
871