

RESEARCH ARTICLE

Reference databases, primer choice, and assay sensitivity for environmental metabarcoding: Lessons learnt from a re-evaluation of an eDNA fish assessment in the Volga headwaters

Tamara Schenekar¹  | Martin Schletterer^{2,3}  | Laurène A. Lecaudey^{1,4}  | Steven J. Weiss¹ 

¹Institute of Biology, University of Graz, Graz, Austria

²Institute of Hydrobiology and Aquatic Ecosystem Management, BOKU—University of Natural Resources and Life Sciences, Vienna, Austria

³Department of Hydropower Engineering (Section Ecology), TIWAG—Tiroler Wasserkraft AG, Innsbruck, Austria

⁴Department of Natural History, NTNU University Museum, Norwegian University of Science and Technology, Trondheim, Norway

Correspondence

Tamara Schenekar, Institute of Biology, University of Graz, Graz, Austria.
Email: tamara.schenekar@uni-graz.at

Funding information

University of Graz, Grant/Award Number: Funding for Open Access publication; VÖU - Verein für Ökologie und Umwelt, Grant/Award Number: Single species targeting via qPCRs in fish monitor; University of Graz is also acknowledged

Associate Editor: Paul Wood

Abstract

Biodiversity monitoring via environmental DNA, particularly metabarcoding, is evolving into a powerful assessment tool for riverine systems. However, for metabarcoding to be fully integrated into standardized monitoring programmes, some current challenges concerning sampling design, laboratory workflow, and data analysis need to be overcome. Here, we review some of these major challenges and potential solutions. We further illustrate three potential pitfalls, namely the choice of suitable metabarcoding primers, the necessity of complete reference databases, and varying assay sensitivities, by a reappraisal of our own recently carried out metabarcoding study in the Volga headwaters. TaqMan qPCRs had detected catfish (*Silurus glanis*) and European eel (*Anguilla anguilla*), whereas metabarcoding had not, in the same samples. Furthermore, after extending the genetic reference database by 12 additional species and re-analysing the metabarcoding data, we additionally detected the Siberian spiny loach (*Cobitis sibirica*) and Ukrainian brook lamprey (*Eudontomyzon mariae*) and reassigned the operational taxonomic units previously assigned to *Misgurnus fossilis* to *Cobitis sibirica*. In silico analysis of metabarcoding primer efficiencies revealed considerable variability among primer pairs and among target species, which could lead to strong primer bias and potential false-negatives in metabarcoding studies if not properly compensated for. These results highlight some of the pitfalls of eDNA-metabarcoding as a means of monitoring fish biodiversity in large rivers, which need to be considered in order to fully unleash the full potential of these approaches for freshwater biodiversity monitoring.

KEYWORDS

eDNA, false negatives, fish, inhibition, metabarcoding, primer bias, qPCR, real-time PCR

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2020 The Authors. *River Research and Applications* published by John Wiley & Sons Ltd

1 | eDNA-METABARCODING AS A TOOL FOR BIODIVERSITY ASSESSMENTS AND ITS ASSOCIATED CHALLENGES

Biodiversity assessments are vital to every biomonitoring effort and aim to record all resident taxa of specific bioindicator groups in a target area, like a river catchment. A potentially powerful addition to conventional biodiversity surveys is DNA-based assessment. The utilization of environmental DNA (or eDNA; Taberlet, Coissac, Pompanon, Brochmann, & Willerslev, 2012) for biomonitoring has been promoted by numerous researchers, particularly for freshwater ecosystems (e.g., Hering et al., 2018; Senapati et al., 2018). Bioassessments utilizing eDNA have various positive assets that can compensate for the drawbacks and limitations associated with conventional sampling and species identification methods. These include non-invasiveness, better cost and time effectiveness, and increased taxonomic identification (Deiner et al., 2017). For overall biodiversity assessment of a target species community, eDNA metabarcoding is the ideal method. This approach uses universal primers to bind to the eDNA left by members of the target group in an environmental sample and amplifies a genetic barcode for each of these species in parallel during polymerase chain reaction (PCR). The amplicons then get sequenced using high-throughput sequencing, and species composition is reconstructed using the retrieved DNA sequences.

However, eDNA and metabarcoding approaches are still under development, and multiple issues need to be resolved before they can be fully integrated into standard biomonitoring systems. These issues concern the influence of various factors during eDNA sampling, laboratory workflow, and data analysis on detection probabilities. Even before sampling, researchers need to consider DNA persistence in the environment, which has been shown to be influenced by factors such as pH, temperature, UV radiation, flow, and chemical properties of bedrock (e.g., Barnes & Turner, 2016; Seymour et al., 2018; Stoeckle et al., 2017). For eDNA sampling, storage, and extraction, a plurality of options exist, and the choice of method can significantly influence detection probabilities (e.g., Deiner, Walsler, Maechler, & Altermatt, 2015; Spens et al., 2017), making it increasingly difficult to pick the optimal protocol without preliminary information on the study system. A crucial step in any metabarcoding analysis is the PCR. Ideally, PCR amplifies all target DNA at an equal and exponential rate. However, this is rarely true due to various reasons. First, eDNA samples frequently contain PCR inhibitors, hampering polymerase function and therefore increasing the risk of false negatives, particularly when working with species occurring at low abundance. Inhibitors might stem from abiotic or biological sources such as bedrock, algae, or organic input (McKee, Spear, & Pierson, 2015; Stoeckle et al., 2017). Inhibitor-removal kits have been shown to efficiently remove inhibitors but add costs. While inhibition influences PCR function in general, PCR bias favours particular template DNA strands from the mixed DNA template. PCR bias can stem from (a) PCR stochasticity – random preference of individual DNA strands during PCR reaction (Kebschull & Zador, 2015); (b) preference of more abundant DNA templates, potentially masking the presence of low-abundant species (Bylemans, Gleeson, Duncan, Hardy, & Furlan, 2019); and (c) primer bias. Primer bias originates from sequence variation in the primer annealing sites among individual

species, leading to higher efficiency of the PCR reaction for certain species in a mixed DNA template, while other species might get amplified less or not at all (Bylemans, Gleeson, Hardy, & Furlan, 2018; Elbrecht & Leese, 2015). While the first two can – at least to some extent – be overcome by performing multiple PCR replicates (Ficetola et al., 2015), primer bias is best avoided by careful evaluation of the metabarcoding primers used, in order to avoid polymorphism in the DNA binding region. This goal might be complicated by additional primer criteria, such as small amplicon size, high taxonomic resolution of the amplified fragment, and target group specificity (Bylemans et al., 2018). Therefore, primer performance is best evaluated for the specific target community and geographic region under study (Elbrecht & Leese, 2017).

After data generation and raw data filtration, the retrieved DNA sequences are usually compared with a reference database of the expected species community in order to translate the obtained molecular operational taxonomic units into biological species for final data interpretation. However, such reference databases are still far from complete despite global initiatives like the international Barcode of Life (iBOL; Weigand et al., 2019). Additionally, the target fragment of iBOL, a part of the mitochondrial cytochrome oxidase subunit I (COI), has been shown to be less suitable for metabarcoding work for vertebrates (Deagle, Jarman, Coissac, Pompanon, & Taberlet, 2014). Therefore, other mitochondrial markers, particularly fragments of the mitochondrial 12S and 16S rRNA genes, have been the focus for metabarcoding primer design for this group (Evans et al., 2016; Miya et al., 2015). Databases for these gene fragments are less complete than for COI and have been shown to be heavily biased for certain geographic regions and taxonomic groups (Belle, Stoeckle, & Geist, 2019; Weigand et al., 2019), highlighting the need to expand databases for understudied areas and taxonomic groups.

An alternative to metabarcoding are species-specific approaches like real-time PCR (qPCR). Such assays are designed to specifically amplify only the target species from the bulk DNA in a sample. This approach does not require DNA sequencing and genetic reference databases. Instead, species presence is inferred via the amplification itself during the PCR step in real time. Whereas the overall informative value of qPCR is taxonomically narrow compared to metabarcoding, its application can circumvent some of metabarcoding's challenges. These include the elimination of primer bias, as the primers target one species only, easier detection of inhibition, as the amplification efficiency can be directly observed, and a higher potential to reliably correlate the results to species abundances (Doi et al., 2017; Lacoursière-Roussel, Côté, Leclerc, & Bernatchez, 2016). Additionally, such a targeted approach has been shown to be more sensitive than metabarcoding (Bylemans et al., 2019; Harper et al., 2018) and is more cost effective when the target taxa involve only a few species (Tsuji et al., 2018).

We tested for the influence of three potential problems – an incomplete reference database, PCR inhibition and primer bias – on the results of our own, recently carried out, metabarcoding study in the headwaters of the Volga River (Lecaudey, Schletterer, Kuzovlev, Hahn, & Weiss, 2019). Lecaudey et al. (2019), detected 23 fish species across the sampled stretches, but the number of detected species varied considerably between the three barcode markers used, namely 12S rRNA, 16S rRNA, and cytochrome B (CytB). The Ukrainian brook lamprey (*Eudontomyzon*

mariae), blue bream (*Ballerus ballerus*), white-eye bream (*Ballerus sapa*), asp (*Leuciscus aspius*), Siberian spiny loach (*Cobitis sibirica*), wels catfish (*Silurus glanis*), and European eel (*Anguilla anguilla*) were not detected by any of the three markers used. The first five species lacked reference sequences for the 12S and 16S rRNA, which had an overall better performance than CytB for species detection. For the latter two species (*S. glanis* and *A. anguilla*), reference sequences were available and recent presence at the sampled stretches has been documented (Schletterer, 2006; Schletterer et al., 2018). Additionally, potential presence of inhibitors was observed in the laboratory during PCR steps. Therefore, in this re-evaluation, we aimed to (a) verify the absence of *S. glanis* and *A. anguilla* DNA in the samples obtained from the Upper Volga drainage via TaqMan qPCRs, (b) complement the genetic reference databases of the 12S and 16S rRNA by generating new reference sequences for the remaining above-mentioned species, (c) check for potential inhibitors interfering with PCRs, and (d) test in silico the efficiencies of the three metabarcoding primer pairs used by Lecaudey et al. (2019) to investigate potential primer bias.

2 | RE-EVALUATION OF A METABARCODING STUDY IN THE VOLGA HEADWATERS

2.1 | Material and methods

2.1.1 | TaqMan qPCRs

For TaqMan qPCRs, we used the eDNA extracts produced by Lecaudey et al. (2019). These encompass 60 eDNA samples collected at 11 locations in the free-flowing section of the Volga River headwaters including three locations from the Volga River itself, one location each for the Tvertsa, Moksha, and Nochnaya Rivers, and five locations from the Tudovka River (Figure 1). Per site, either four or eight eDNA sampling replicates (250 mL each) were collected, depending on the size of the river channel. Details on the sampling and extraction procedures are given in Lecaudey et al. (2019) and in Data S1.

We performed TaqMan qPCRs for *S. glanis* and *A. anguilla* using previously published protocols (Jensen, Knudsen, Munk, Thomsen, & Møller, 2018; Roy, Belliveau, Mandrak, & Gagné, 2018). Sampling sites which either represent suitable habitat for target species or for which recent sightings were reported were identified (Figure 1). For each relevant site, the species-specific assay was applied to at least three randomly chosen sampling replicates and 10 qPCR replicates were carried out for each. An eDNA sample was recorded as positive when positive amplification was observed in at least two independent qPCR replicates. If a PCR replicate only yielded 1 positive amplification out of 10 amplifications trials, additional samples at that location were subjected to qPCR analysis until a clear result was obtained. We evaluated the potential presence of inhibitors by spiking positives controls with an aliquot of the eDNA extracts. Details on qPCR reaction set-up and inhibitor testing are given in Data S1.

2.1.2 | Expansion of genetic reference data and re-analysis of the metabarcoding data

We produced reference sequences for the targeted 12S and 16S rRNA fragments for *Lampetra planeri*, *E. mariae*, *B. ballerus*, *B. sapa*, *L. aspius*, and *C. sibirica*. The third marker, CytB, was excluded for genetic reference data generation because of its generally poor performance in the previous metabarcoding analysis. Most tissue samples of the target species were obtained from in-house collections of Austrian Barcode of Life. One sample of *C. sibirica*, along with two additional *Cobitis taenia* were obtained from sample collections of Mordovia State Nature Reserve and Smolny National Park, Mordovia. DNA was extracted from tissue samples using the DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany). Primer pairs were designed in the flanking regions of the targeted 12S and 16S rRNA fragments using congeneric or confamilial species sequence data. For details on all primers, PCR reaction set-up, and sequence generation, see Data S1.

For further reference database expansion, we added available sequence information for six additional non-native species for all three barcode markers. These species were listed as being present in the Upper Volga system by Litvinov et al. (2009) and include Amur sleeper (*Perccottus glenii*), guppy (*Poecilia reticulata*), ninespine stickleback (*Pungitius pungitius*), Southern ninespine stickleback (*Pungitius platygaster*), kilka (*Clupeonella cultriventris*), and Caspian bighead goby (*Neogobius gorlap*) (Data S1). For the Caspian bighead goby, only CytB sequences were available.

Metabarcoding re-analysis was conducted on the de-multiplexed raw data produced by Lecaudey et al. (2019), encompassing 49 eDNA samples + 11 negative controls from all the above-mentioned locations. The metabarcoding data also contained a mock community serving as a positive control consisting of equimolar amounts of tissue extracts of six Indo-Pacific fish species. For details on Illumina library preparation and bioinformatic analysis workflow, see Lecaudey et al. (2019).

2.1.3 | Testing for primer bias

We tested in silico for potential primer bias by calculating the primer efficiencies of the three primer pairs used for metabarcoding library preparation: 12S and 16S rRNA (Evans et al., 2016) and CytB (Burgener & Hübner, 1998; Meyer, Höfele, Lüthy, & Candrian, 1995). For this, we expanded the alignments of the reference databases at both the 5' and 3' ends, to include the primer binding regions of the respective primer sets. In total, 54, 56, and 43 species were included for this analysis for 12S and 16S rRNA and CytB, respectively, using a total of 153, 141, and 100 sequences (Data S1).

Primer efficiency analysis was carried out via the DECIPHER package v.2.10.2 (Wright, 2015) in R v.3.5.2 (R Core Team, 2018) with PCR amplification efficiency ranging from 0 (very low efficiency) to 1 (maximum efficiency possible) and calculated based on the hybridization and elongation efficiency of the primers to the target sites

(settings: temp = 58 °C for 12S and 16S rRNA, and 53 °C for CytB, $p = .0003$, ions = 0.1558, TaqEfficiency = True, maxDistance = 0.4, maxGaps = 2 and processors = Null). We then calculated the lower mean efficiency (LM_{Eff}) value between the forward and reverse primer for each species. Using the data of the positive control (mock community), we tested for each marker separately for correlations between (a) LM_{Eff} and the number of reads assigned to each species and (b) LM_{Eff} , and whether the species was scored as 'detected' in the metabarcoding data, using both no threshold for detection (i.e., detection was scored as positive as soon as one read was assigned to that species) and applying the threshold levels used by Lecaudey et al. (2019). Correlations were calculated for tests of (a) using Pearson correlation coefficient, whereas for (b) point-biserial correlations were applied. All statistical testing was carried out in SPSS 24.0 (IBM Corp, 2016).

2.2 | Results

2.2.1 | TaqMan qPCRs

Throughout both TaqMan qPCR assays, no amplification of negative controls was observed. A total of 16 eDNA samples stemming from the five relevant sampling sites were typed with the *S. glanis* TaqMan assay. Four samples showed positive amplification in two or more out of the 10 PCR replicates (Table 1). Three of these samples stem from Mel'nikovo at the Tvertsa River, whereas one is from the Volga River, at Rzhev. For *A. anguilla*, a total of 18 eDNA samples from four locations were tested (Table 2). Two samples from one location (Volga-Staritsa) amplified above our threshold. Out of the 35 sample-assay combinations used in TaqMan qPCR analysis (16 samples for *S. glanis*, 19 samples for *A. anguilla*), 4 showed a $\Delta C_T \geq 2$ after initial inhibition

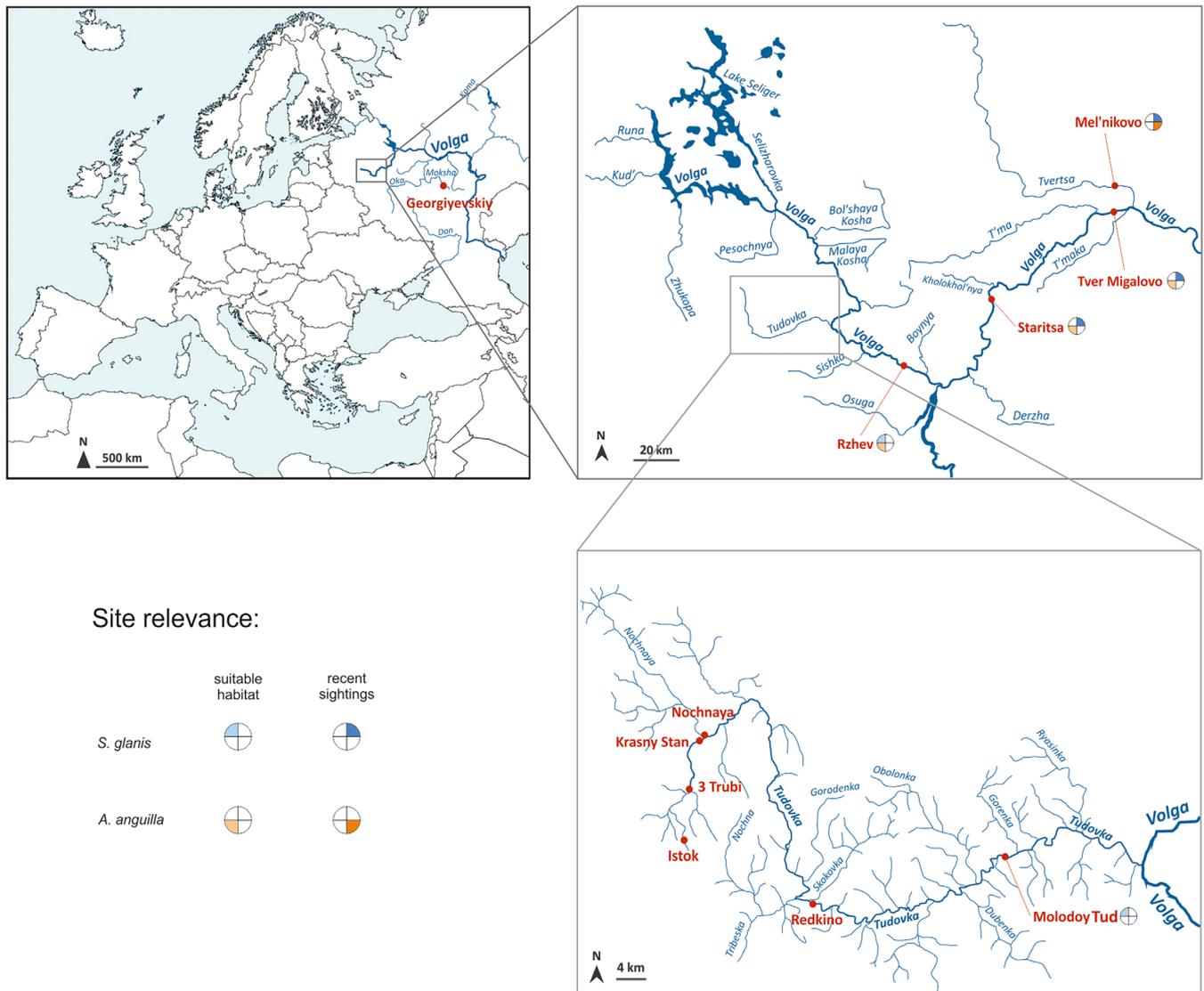


FIGURE 1 Locations of the sampling sites (red dots) included in this study. Relevance of individual sites for the two target species selected for TaqMan qPCRs is given by symbols on the right-hand side of site names (color key given on the bottom left). Modified from Lecaudey et al. (2019)

testing in duplicate, suggesting inhibition. After repeating these five samples with five additional PCR replicates, only one sample–assay combination (sample 2 at Volga-Staritsa for the *A. anguilla* assay) still yielded a $\Delta C_T \geq 2$ (2.58).

2.2.2 | Extension of genetic reference data and re-analysis of the metabarcoding data

We successfully sequenced target fragments of the 12S and 16S rRNA of two *L. planeri*, one *E. mariae*, two *B. sapa*, one *B. ballerus*, two *L. aspius*, and one *C. sibirica* (GenBank accession numbers: Data S1), providing the first reference sequences for the target fragments in these species. The extended genetic databases for metabarcoding consisted of a total of 235 reference sequences from 68 species for 12S rRNA, 193 sequences from 63 species for 16S rRNA, and 326 sequences from 66 species for CytB. After metabarcoding data re-analysis, we retrieved significant blast hits on two additional species, namely *C. sibirica* and *E. mariae*. The latter was detected in the Moksha River from the 12S rRNA dataset. The operational taxonomic units (OTUs) assigned to *E. mariae* had all been assigned to *L. planeri* in Lecaudey et al. (2019). *C. sibirica* was detected at two sampling sites in the Volga (Staritsa and Rzhhev) from both 12S and 16S rRNA data. The OTUs assigned to *C. sibirica* had been assigned to *C. taenia* or the family Cobitidae in Lecaudey et al. (2019). Additionally, all OTUs assigned to the *Misgurnus* genus in Lecaudey et al. (2019) were now

TABLE 1 Number of positive qPCRs (from 10 replicates) from the TaqMan qPCR assay for *Silurus glanis* on the selected samples

River	Location	Sampling replicate number			
		1	2	3	4
Tudovka	Molodoy Tud	0	0	0	–
Volga	Rzhhev	0	3	1	0
Volga	Staritsa	0	0	0	–
Volga	Tver Migalovo	0	0	0	–
Tvertsa	Mel'nikovo	2	3	4	–

Note: Shown are the river and location of each sample, as well as the respective sampling replicate. Number of amplifications above the detection threshold are given in bold.

TABLE 2 Number of positive qPCRs (from 10 replicates) from the TaqMan qPCR assay for *Anguilla anguilla* on the selected samples

River	Location	Sampling replicate number							
		1	2	3	4	5	6	7	8
Volga	Rzhhev	0	0	0	–	–	–	–	–
Volga	Staritsa	3	0	0	2	–	–	–	–
Volga	Tver Migalovo	1 & 0^a	0	0	–	–	–	–	–
Tvertsa	Mel'nikovo	1 & 0^a	0	1 & 0^a	0	0	0	0	0

Note: Shown are the river and location of each sample, as well as the respective sampling replicate. Number of amplifications above the detection threshold are given in bold.

^aSamples were analysed twice, each with 10 PCR replicates.

assigned to the newly added *C. taenia* sequences, eliminating *M. fossilis* from the final species list.

Considering the additions and one removal of species from the detected species list, the overall number of detected species in the studied free-flowing section of the headwaters of the Volga River now amounts to 26 (Table 3).

2.2.3 | Testing for primer bias

The observed PCR efficiencies of primers showed overall high mean efficiency values, but individual efficiencies of certain primer–species combinations revealed significant drops (Table 4, Figure 2), potentially leading to strong primer bias. A total of seven species–primer combinations revealed a primer efficiency of zero: one species for the 12S rRNA reverse primer, one species for the 16S rRNA forward and reverse primers, respectively, and four species for the CytB reverse primer.

Among the three markers, the 12S rRNA primer pair had the highest mean primer efficiencies for both forward and reverse primers. All but five species analysed (9.3%) showed almost optimal annealing conditions ($LM_{Eff} > 0.9$), for both 12S rRNA primers. Concerning the 16S rRNA primers, 10 species yielded an $LM_{Eff} < 0.9$ (17.9%). CytB had the lowest overall mean efficiency for both forward and reverse primers and the highest variability among species analysed. In total, 22 species (51.1%) revealed an $LM_{Eff} < 0.9$ for CytB. Overall, the mean primer pair efficiencies correlated well with the number of species detected by them in the original study (Table 4).

Among the six species of the mock community, LM_{Eff} varied between 0 and 0.999 (Data S1). The 16S rRNA primers had the lowest mean LM_{Eff} across all six species compared to 12S rRNA and CytB primers (0.518, 0.831, and 0.702, respectively). This went along with these primers being able to recover only three out of the six species of the mock community when no detection threshold was applied, whereas 12S rRNA and CytB both recovered five of the six species. However, when the respective detection threshold was applied, the CytB dataset yielded only one positive species detection. For 12S, LM_{Eff} was significantly correlated with detection with or without a detection threshold (both $p < .0001$, $r = 1.0$, $df = 4$). For CytB, a significant correlation between LM_{Eff} and detection without a threshold was observed ($p = .025$, $r = 0.87$, $df = 4$). For all three markers tested

TABLE 3 Updated overall list of the 26 detected fish and lamprey species in the Volga headwaters using eDNA

Species	Nochnaya			Tudovka			Volga					Tvertsa		Moksha Georgiyevskiy	Total
	Nochnaya	Istok	3 Trubi	Krasny Stan	Redkino	Molodoy Tud	Rzhev	Staritsa	Tver Migalovo	Mel'nikovo	Mel'nikovo	Mel'nikovo			
<i>Abramis brama</i>				x	x		x	x	x			x			6
<i>Alburnus alburnus</i>		x	x		x	x	x	x	x			x		x	9
<i>Anguilla anguilla</i> *								x							1
<i>Barbatula barbatula</i>	x	x	x	x	x	x	x	x	x			x		x	11
<i>Blicca bjoerkna</i>	x						x	x	x			x			5
<i>Chondrostoma variable</i>							x		x						2
<i>Cobitis taenia</i>							x	x				x			3
<i>Cobitis sibirica</i>							x	x							2
<i>Cottus gobio</i>	x				x	x	x	x	x			x			7
<i>Esox lucius</i>	x	x	x	x	x	x	x	x	x			x		x	11
<i>Eudontomyzon mariae</i>														x	1
<i>Gobio sp.</i> ^a	x				x	x	x	x	x			x		x	8
<i>Gymnocephalus cernua</i>	x			x	x	x		x	x						7
<i>Lampetra planeri</i>	x													x	2
<i>Leucaspis delineatus</i>	x							x							3
<i>Leuciscus idus</i>			x			x	x	x	x						5
<i>Lota lota</i>	x	x	x	x	x		x								6
<i>Neogobius melanostomus</i>									x						1
<i>Perca fluviatilis</i>	x			x				x	x			x		x	7
<i>Phoxinus phoxinus</i>	x				x	x									3
<i>Rutilus rutilus</i>	x	x	x	x	x	x	x	x	x			x		x	11
<i>Salmo trutta</i>	x	x	x	x	x	x	x	x	x			x		x	11
<i>Sander lucioperca</i>														x	1
<i>Silurus glanis</i> *												x			2
<i>Thymallus thymallus</i>									x						2
<i>Tinca tinca</i>								x	x			x			3
Total # of species	13	6	7	8	11	11	19	17	15	14	9	14	9	9	

Note: An 'x' indicates positive detection at a site. Data from both metabarcoding and TaqMan qPCR are combined. Species newly detected in this study are indicated in bold and those only detected by TaqMan qPCR with an asterisk (modified from Lecaudey et al., 2019). Last column 'Total' refer to the number of sites in which each species was detected. Last row refers to total number of species per site.

^aLecaudey et al. (2019) classified the detected species as *Gobio gobio*. However, according to Mendel et al. (2008), *Gobio volgensis* is the only member of *Gobio* inhabiting the Upper Volga basin.

TABLE 4 Summary statistics of the efficiencies for forward and reverse primers of the three markers used for metabarcoding

Marker	Reference	Number of species tested	Forward primer			Reverse primer			# species
			Min	Max	Mean	Min	Max	Mean	
12S rRNA	Evans et al. (2016)	53	0.757	1.000	0.991 (± 0.043)	0.000	1.000	0.959 (± 0.191)	21
16S rRNA	Evans et al. (2016)	55	0.000	0.999	0.978 (± 0.119)	0.000	1.000	0.935 (± 0.119)	18
Cytochrome b	Burgener and Hübner (1998)	43	0.011	1.000	0.954 (± 0.175)	0.000	1.000	0.741 (± 0.387)	16

Note: Min, minimum efficiency observed; Max, maximum efficiency observed; Mean, mean efficiency (\pm SD); # species, Number of species detected by Lecaudey et al. (2019).

combined, the correlation between LM_{Eff} and read count ($p_{one-tailed} = .04$; $r = 0.42$, $df = 16$) was significant, as was correlation between LM_{Eff} and detection with and without a threshold ($p = .009$, $r = 0.6$, $df = 16$ and $p < .0001$, $r = 0.83$, $df = 16$, respectively).

2.3 | Discussion

2.3.1 | TaqMan qPCRs

TaqMan qPCRs revealed unambiguous detection of *S. glanis* and *A. anguilla* at two and one sampling sites, respectively, which were missed with the metabarcoding data in Lecaudey et al. (2019). In Lecaudey et al. (2019), one OTU had been assigned to *S. glanis* at Mel'nikovo using 12S rRNA, but the assigned read count (6) had fallen below the set threshold for false-positive elimination and therefore was not counted as positive detection. These results are concordant with previous studies that showed a higher sensitivity of targeted qPCR compared to metabarcoding (Bylemans et al., 2019; Harper et al., 2018), indicating a better performance of targeted approaches particularly for species of low abundance. However, a recently developed TaqMan qPCR assay did not detect sterlet (*Acipenser ruthenus*) in the same system, despite recent reports of its presence at individual sites (Schenekar, Schletterer, & Weiss, 2020), potentially indicating the border of sensitivity of such assays for a species occurring at very low densities. The sensitivity of eDNA assays will always be influenced by the filtered water volume, storage and extraction methods, and the number of PCR replicates performed, as well as both metabarcoding thresholds or qPCR detection thresholds (as also discussed by Goldberg et al., 2016; Harper et al., 2019). Standardizations across different systems will be hard to achieve, as environmental conditions affecting the sensitivity of eDNA assays will very much differ. However, good scientific practice, such as meticulous reporting of the sampling and laboratory protocols and on experimentally assessed limits of detection or quantification, can nevertheless deliver comprehensible and interpretable results.

Overall, we did not detect strong signals of inhibition during our TaqMan qPCR assay. This is presumably due to both a spin-column treatment and the use of an inhibitor-tolerant enzyme. Therefore, we recommend using inhibitor-tolerant enzymes as well as the implementation of inhibition controls.

2.3.2 | Extension of reference databases and re-analysis of metabarcoding data

The extension of the genetic reference database by Sanger sequencing and using published sequences (12 species in total) resulted in the detection of two additional species, namely *C. sibirica* and *E. mariae*. *C. sibirica* was detected in the Volga River, and it is also known to occur in the Tudovka River but this could not be verified in the present study. *E. mariae* yielded detection signals in the Moksha River, where the occurrence of *E. mariae* has recently been reported near the sampling site (Artaev & Ruchin, 2017).

In addition to the newly detected taxa, we observed reassignment of OTUs based on the extended reference database, demonstrating importance of exhaustive reference sequences, particularly when a 'best match' approach like BLAST is used for taxonomic assignment. Not only does the lack of individual reference sequences result in potential false negatives, but they may also lead to erroneous detection of closely related sister-taxa.

2.3.3 | Primer bias and primer efficiencies of the metabarcoding primers

The observed primer efficiencies varied considerably among (a) the three metabarcoding primer pairs used, (b) among the target species, and (c) in some cases between forward and reverse primers, all leading to potential bias and associated problems in metabarcoding analyses. Mean primer pair efficiencies mirrored the numbers of species detected by Lecaudey et al. (2019), with 12S yielding the highest number of detected species and CytB the lowest. The consequences of the variability in priming efficiencies could also be well observed in vitro via the artificially created mock community, where primer efficiencies clearly influenced the read count and detection of individual species. This shows that even when DNA is highly concentrated and evenly distributed across a modest number of species, primer bias can have a decisive influence on species detection. Therefore, critical evaluation of the primer efficiencies of the barcode marker of choice in the target species community is imperative for every metabarcoding study (as suggested by Elbrecht & Leese, 2017). Additionally, a good way to overcome potential or unavoidable variability of primer efficiencies across species is the use of multiple markers. Thus, choice of

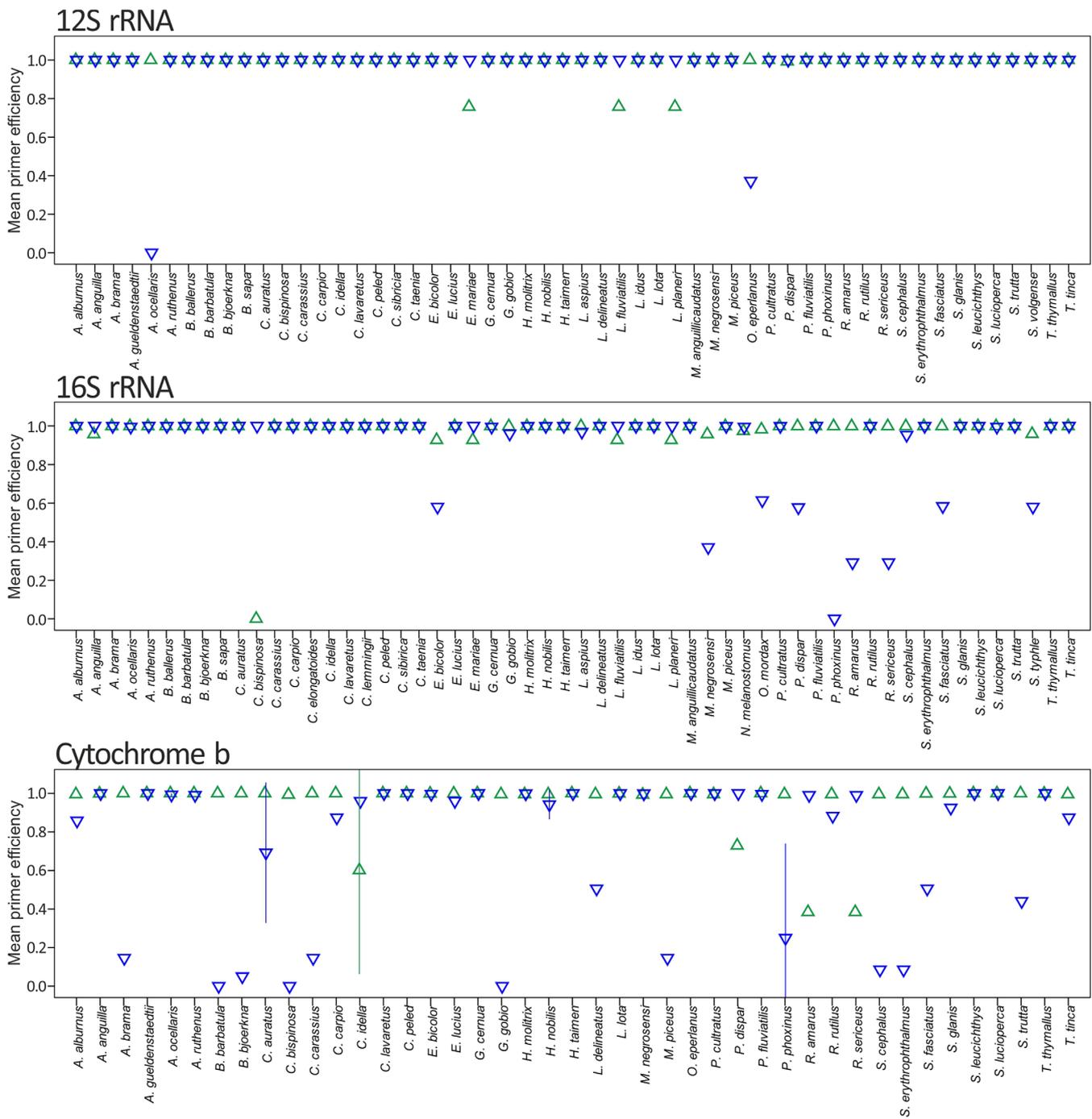


FIGURE 2 Mean primer efficiencies for the three metabarcoding primer sets analysed (12S rRNA, 16S rRNA, and CytB) for each species. Shown are the mean values for each species separated by forward (green triangle pointing upwards) and reverse (blue triangle pointing downwards) primers. Error bars represent standard deviations within species

multiple markers, in addition to taxonomic resolution, amplicon size, and group specificity, should achieve 'complementarity' with respect to primer efficiency, with the aim of each species having at least one primer pair with a high degree of primer efficiency. For the dataset of Lecaudey et al. (2019), we conclude that the combined usage of the 12S and 16S rRNA primers of Evans et al. (2016) yielded the best results if only two out of the three markers were to be used. Again, we emphasize that this cannot be generalized for all communities;

each researcher should carry out critical primer efficiency analysis on the target taxa before carrying out a metabarcoding study.

3 | CONCLUSIONS

The methodology and protocols around eDNA and metabarcoding surveys are still under development, and therefore it is imperative to

recognize the pitfalls and challenges to overcome before fully integrating them into standard biomonitoring efforts. Here we re-evaluated the results of our own metabarcoding assessment in the headwaters of the Volga River, as a case study. We showed that single-species TaqMan qPCR showed higher sensitivity than metabarcoding for selected rare species and highlighted the importance of complete reference databases to receive a full picture of the present species diversity. Finally, we showed that primer bias can affect the outcomes of metabarcoding analyses leading to potential false negatives. Thus, primer bias should be properly evaluated on each reference data set and primer pair.

ACKNOWLEDGEMENTS

We thank Y. N. Zhenikov, K. Y. Zhenikov, and V. V. Kuzovlev for their help during fieldwork as well as N. A. Potemkin and A. S. Zheltukhin for providing the sampling permission and infrastructure during the field survey along the Tudovka River. We would like also to thank L. Zangl, S. Koblmüller, and C. Ratschan for providing tissue samples. Special thanks go to A. B. Ruchin of the Mordovia State Nature Reserve for providing several reference species samples from the Volga catchment. Furthermore, we would like to thank C. Hahn for his support with the bioinformatics analysis. This study was financed by the VÖU (Verein für Ökologie und Umwelt), in the framework of the project 'Single species targeting via qPCRs in fish monitoring in the headwaters of the Volga River using eDNA-SSqeDNA'. Financial support of the University of Graz is also acknowledged.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

All generated sequences have been submitted to NCBI GenBank (accession numbers MN153552-MN153562 and MN153569-MN153579). Accession numbers for all additional genetic reference data used in the study is listed in the supplemental material.

ORCID

Tamara Schenekar  <https://orcid.org/0000-0002-4079-8550>

Martin Schletterer  <https://orcid.org/0000-0002-1896-3329>

Laurène A. Lecaudey  <https://orcid.org/0000-0002-6482-9251>

Steven J. Weiss  <https://orcid.org/0000-0002-4734-5984>

REFERENCES

- Artaev, O. N., & Ruchin, A. (2017). The ichthyofauna of the Moksha River, a tributary of the Volga River basin, Russia. *Check List*, 13, 185–202.
- Barnes, M. A., & Turner, C. R. (2016). The ecology of environmental DNA and implications for conservation genetics. *Conservation Genetics*, 17(1), 1–17. <https://doi.org/10.1007/s10592-015-0775-4>
- Belle, C. C., Stoeckle, B. C., & Geist, J. (2019). Taxonomic and geographical representation of freshwater environmental DNA research in aquatic conservation. *Aquatic Conservation: Marine and Freshwater Ecosystems*, 29, 1996–2009. <https://doi.org/10.1002/aqc.3208>
- Burgener, M., & Hübner, P. (1998). Mitochondrial DNA enrichment for species identification and evolutionary analysis. *Zeitschrift Fur Lebensmittel-Untersuchung Und -Forschung A*, 207, 261–263. <https://doi.org/10.1007/s002170050329>
- Bylemans, J., Gleeson, D. M., Duncan, R. P., Hardy, C. M., & Furlan, E. M. (2019). A performance evaluation of targeted eDNA and eDNA metabarcoding analyses for freshwater fishes. *Environmental DNA*, 1, 402–414. <https://doi.org/10.1002/edn3.41>
- Bylemans, J., Gleeson, D. M., Hardy, C. M., & Furlan, E. (2018). 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(17), 8697–8712. <https://doi.org/10.1002/ece3.4387>
- IBM Corp. (2016). *IBM SPSS Statistics for Windows*. Armonk, NY: IBM Corp.
- Deagle, B. E., Jarman, S. N., Coissac, E., Pompanon, F., & Taberlet, P. (2014). DNA metabarcoding and the cytochrome c oxidase subunit I marker: Not a perfect match. *Biology Letters*, 10(9), 2–5. <https://doi.org/10.1098/rsbl.2014.0562>
- Deiner, K., Bik, H. M., Mächler, E., Seymour, M., Lacoursière-Roussel, A., Altermatt, F., ... Bernatchez, L. (2017). Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. *Molecular Ecology*, 26(21), 5872–5895. <https://doi.org/10.1111/mec.14350>
- Deiner, K., Walsler, J.-C., Maechler, E., & Altermatt, F. (2015). Choice of capture and extraction methods affect detection of freshwater biodiversity from environmental DNA. *Biological Conservation*, 183(SI), 53–63. <https://doi.org/10.1016/j.biocon.2014.11.018>
- Doi, H., Inui, R., Akamatsu, Y., Kanno, K., Yamanaka, H., Takahara, T., & Minamoto, T. (2017). Environmental DNA analysis for estimating the abundance and biomass of stream fish. *Freshwater Biology*, 62(1), 30–39. <https://doi.org/10.1111/fwb.12846>
- Elbrecht, V., & Leese, F. (2015). Can DNA-based ecosystem assessments quantify species abundance? Testing primer bias and biomass–sequence relationships with an innovative metabarcoding protocol. *PLoS One*, 10(7), 1–16. <https://doi.org/10.1371/journal.pone.0130324>
- Elbrecht, V., & Leese, F. (2017). PrimerMiner: An R package for development and in silico validation of DNA metabarcoding primers. *Methods in Ecology and Evolution*, 8(5), 622–626. <https://doi.org/10.1111/2041-210X.12687>
- Evans, N. T., Olds, B. P., Renshaw, M. A., Turner, C. R., Li, Y., Jerde, C. L., ... Lodge, D. M. (2016). Quantification of mesocosm fish and amphibian species diversity via environmental DNA metabarcoding. *Molecular Ecology Resources*, 16(1), 29–41. <https://doi.org/10.1111/1755-0998.12433>
- Ficetola, G. F., Pansu, J., Bonin, A., Coissac, E., Giguët-Coxev, C., De Barba, M., ... Taberlet, P. (2015). Replication levels, false presences and the estimation of the presence/absence from eDNA metabarcoding data. *Molecular Ecology Resources*, 15(3), 543–556. <https://doi.org/10.1111/1755-0998.12338>
- Goldberg, C. S., Turner, C. R., Deiner, K., Klymus, K. E., Thomsen, P. F., Murphy, M. A., ... Taberlet, P. (2016). Critical considerations for the application of environmental DNA methods to detect aquatic species. *Methods in Ecology and Evolution*, 7(11), 1299–1307. <https://doi.org/10.1111/2041-210X.12595>
- Harper, L. R., Buxton, A. S., Rees, H. C., Bruce, K., Brys, R., Halfmaerten, D., ... Hänfling, B. (2019). Prospects and challenges of environmental DNA (eDNA) monitoring in freshwater ponds. *Hydrobiologia*, 826(1), 25–41. <https://doi.org/10.1007/s10750-018-3750-5>
- Harper, L. R., Lawson Handley, L., Hahn, C., Boonham, N., Rees, H. C., Gough, K. C., ... Hänfling, B. (2018). Needle in a haystack? A comparison of eDNA metabarcoding and targeted qPCR for detection of the great crested newt (*Triturus cristatus*). *Ecology and Evolution*, 8(12), 6330–6341. <https://doi.org/10.1002/ece3.4013>
- Hering, D., Borja, A., Jones, J. I., Pont, D., Boets, P., Bouchez, 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>
- Jensen, M. R., Knudsen, S. W., Munk, P., Thomsen, P. F., & Møller, P. R. (2018). Tracing European eel in the diet of mesopelagic fishes from the Sargasso Sea using DNA from fish stomachs. *Marine Biology*, 165(8), 1–11. <https://doi.org/10.1007/s00227-018-3390-3>
- Kebschull, J. M., & Zador, A. M. (2015). Sources of PCR-induced distortions in high-throughput sequencing data sets. *Nucleic Acids Research*, 43(21), e143. <https://doi.org/10.1093/nar/gkv717>
- Lacoursière-Roussel, A., Côté, G., Leclerc, V., & Bernatchez, L. (2016). Quantifying relative fish abundance with eDNA: A promising tool for fisheries management. *Journal of Applied Ecology*, 53(4), 1148–1157. <https://doi.org/10.1111/1365-2664.12598>
- Lecaudey, L. A., Schletterer, M., Kuzovlev, V. V., Hahn, C., & Weiss, S. J. (2019). Fish diversity assessment in the headwaters of the Volga River using environmental DNA metabarcoding. *Aquatic Conservation: Marine and Freshwater Ecosystems*, 29(10), 1785–1800. <https://doi.org/10.1002/aqc.3163>
- Litvinov, A. S., Mineeva, N., Papchenkov, V. G., Korneva, L. G., Lazareva, V., Shcherbina, G. K., ... Shurganova, G. V. (2009). Volga River Basin. In K. Tockner, U. Uehlinger, & C. T. Robinson (Eds.), *Rivers of Europe* (pp. 23–58). London, Oxford, Boston, New York, San Diego: Academic Press.
- McKee, A. M., Spear, S. F., & Pierson, T. W. (2015). The effect of dilution and the use of a post-extraction nucleic acid purification column on the accuracy, precision, and inhibition of environmental DNA samples. *Biological Conservation*, 183, 70–76. <https://doi.org/10.1016/j.biocon.2014.11.031>
- Mendel, J., Lusk, S., Vasil'eva, E. D., Vasil'ev, V. P., Lusková, V., Ekmekci, F. G., ... Reshetnikov, S. I. (2008). Molecular phylogeny of the genus *Gobio* Cuvier, 1816 (Teleostei: Cyprinidae) and its contribution to taxonomy. *Molecular Phylogenetics and Evolution*, 47(3), 1061–1075. <https://doi.org/10.1016/j.ympev.2008.03.005>
- Meyer, R., Höfele, C., Lüthy, J., & Candrian, U. (1995). Polymerase chain reaction-restriction fragment length polymorphism analysis: A simple method for species identification in food. *Journal of AOAC International*, 78(6), 1542–1551.
- Miya, M., Sato, Y., Fukunaga, T., Sado, T., Poulsen, J. Y., Sato, K., ... Iwasaki, W. (2015). MiFish, a set of universal PCR primers for metabarcoding environmental DNA from fishes: Detection of more than 230 subtropical marine species. *Royal Society Open Science*, 2(7), 150088. <https://doi.org/10.1098/rsos.150088>
- R Core Team (2018). *R: A language and environment for statistical computing*. Retrieved from <https://www.R-project.org>.
- Roy, M., Belliveau, V., Mandrak, N. E., & Gagné, N. (2018). Development of environmental DNA (eDNA) methods for detecting high-risk freshwater fishes in live trade in Canada. *Biological Invasions*, 20(2), 299–314. <https://doi.org/10.1007/s10530-017-1532-z>
- Schenekar, T., Schletterer, M., & Weiss, S. J. (2020). Development of a TaqMan qPCR protocol for detecting *Acipenser ruthenus* in the Volga headwaters from eDNA samples. *Conservation Genetics Resources*. <https://doi.org/10.1007/s12686-020-01128-w>
- Schletterer, M. (2006). Biological assessment on the Upper Volga River. In M. Schletterer & V. S. Kuzovlev (Eds.), *Proceedings of Freshwater Research: Upper Volga Expedition 2005: Technical report* (pp. 76–126). Tönning, Lübeck, Marburg: DAV (Der Andere Verlag).
- Schletterer, M., Kuzovlev, V. V., Zhenikov, Y. N., Tuhtan, J. A., Haidvogel, G., Friedrich, T., ... Füreder, L. (2018). Fish fauna and fisheries of large European rivers: Examples from the Volga and the Danube. *Hydrobiologia*, 814(1), 45–60. <https://doi.org/10.1007/s10750-017-3370-5>
- Senapati, D., Bhattacharya, M., Kar, A., Chini, D. S., Das, B. K., & Patra, B. C. (2018). Environmental DNA (eDNA): A promising biological survey tool for aquatic species detection. *Proceedings of the Zoological Society*, 72, 211–228. <https://doi.org/10.1007/s12595-018-0268-9>
- Seymour, M., Durance, I., Cosby, B. J., Ransom-Jones, E., Deiner, K., Ormerod, S. J., ... Creer, S. (2018). Acidity promotes degradation of multi-species environmental DNA in lotic mesocosms. *Communications Biology*, 1(1), 1–8. <https://doi.org/10.1038/s42003-017-0005-3>
- Spens, J., Evans, A. R., Halfmaerten, D., Knudsen, S. W., Sengupta, M. E., Mak, S. S. T., ... Hellström, M. (2017). Comparison of capture and storage methods for aqueous microbial eDNA using an optimized extraction protocol: Advantage of enclosed filter. *Methods in Ecology and Evolution*, 8(5), 635–645. <https://doi.org/10.1111/2041-210X.12683>
- Stoeckle, B. C., Beggel, S., Cerwenka, A. F., Motivans, E., Kuehn, R., & Geist, J. (2017). A systematic approach to evaluate the influence of environmental conditions on eDNA detection success in aquatic ecosystems. *PLoS One*, 12(12), e0189119. <https://doi.org/10.1371/journal.pone.0189119>
- Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C., & Willerslev, E. (2012). Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology*, 21(8), 2045–2050. <https://doi.org/10.1111/j.1365-294X.2012.05470.x>
- Tsuji, S., Iguchi, Y., Shibata, N., Teramura, I., Kitagawa, T., & Yamanaka, H. (2018). Real-time multiplex PCR for simultaneous detection of multiple species from environmental DNA: An application on two Japanese medaka species. *Scientific Reports*, 8(1), 1–5. <https://doi.org/10.1038/s41598-018-27434-w>
- Weigand, H., Beermann, A. J., Čiampor, F., Costa, F. O., Csabai, Z., Duarte, S., ... Ekrem, T. (2019). DNA barcode reference libraries for the monitoring of aquatic biota in Europe: Gap-analysis and recommendations for future work. *Science of the Total Environment*, 678, 499–524. <https://doi.org/10.1016/j.scitotenv.2019.04.247>
- Wright, E. S. (2015). DECIPHER: Harnessing local sequence context to improve protein multiple sequence alignment. *BMC Bioinformatics*, 16, 322. <https://doi.org/doi:10.1186/s12859-015-0749-z>

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Schenekar T, Schletterer M, Lecaudey LA, Weiss SJ. Reference databases, primer choice, and assay sensitivity for environmental metabarcoding: Lessons learnt from a re-evaluation of an eDNA fish assessment in the Volga headwaters. *River Res Applic*. 2020; 1–10. <https://doi.org/10.1002/rra.3610>