

---

Article

# Universal Set of Primers for DNA Identification of Alien Fish Species in Central Russia (Volga-Kama region)

Dmitry P. Karabanov <sup>1,\*</sup>, Eugenia I. Bekker <sup>2</sup>, Dmitry D. Pavlov <sup>1</sup>, Elena A. Borovikova <sup>1</sup>, Yulia V. Kodukhova <sup>1</sup>, and Alexey A. Kotov <sup>2,\*</sup>

<sup>1</sup> I. D. Papanin Institute for Biology of Inland Waters of Russian Academy of Sciences, Borok Yaroslavl Area, 152742, Russia; dk@ibiw.ru

<sup>2</sup> A. N. Severtsov Institute of Ecology and Evolution of Russian Academy of Sciences, Leninsky Prospect 33, Moscow 119071, Russia; alexey-a-kotov@yandex.ru

\* Correspondence: [dk@ibiw.ru](mailto:dk@ibiw.ru) (DPK), [alexey-a-kotov@yandex.ru](mailto:alexey-a-kotov@yandex.ru) (AAK)

**Abstract:** Reliable species identification is critical for detection and monitoring of biological invasions. In this study, we propose four sets of primers for efficient amplification of several loci, including the mitochondrial cytochrome oxidase-*c* (COI) subunit I gene which is a basis for DNA barcoding. This set of primers gives a shorter product which can be used in high-throughput sequencing systems for metabarcoding purposes. Another mitochondrial locus encoding the large ribosomal subunit (16S) may be useful to study the population structure and as an additional source of information in the metabarcoding of communities. We propose to use a set of primers for the nuclear locus of the small ribosomal subunit (18S) as a positive control and to verify the results of the barcoding. Our proposed sets of primers demonstrate a high amplification efficiency and a high specificity both for freshwater alien and indigenous fishes. The proposed research design makes it possible to carry out extremely cheap studies on the assessment of biological diversity using genetic analysis without expensive equipment, and with the technique for conducting laboratory work and processing of the results available to any researcher. The paper also presents original data on the genetic polymorphism of all mass alien fish species in the Volga-Kama region. High efficiency of DNA identification based on our primers is shown as compared to traditional monitoring of biological invasions.

**Keywords:** biological invasions, fish, DNA, barcoding, primers, identification

---

## 1. Introduction

An in-depth study of species' expansions outside their historical ranges was formulated as a special task in the monograph of Charles Elton [1]. The growing anthropogenic impact, transforming significantly the natural environment, is now complicated by global climatic shifts. They caused an accelerated transformation of the habitats of many plant and animal species. The creation of a system of artificial reservoirs on the large rivers of Europe led to significant changes in indigenous communities whilst creating favorable conditions for a range expansion in a number of aquatic organisms. In addition to directly affecting local communities, some invasive species are able to transform native ecosystems strongly [2], to hybridize with local species [3], and also to act as vectors of non-indigenous parasitic diseases [4]. Also, biological invasions introduce a significant uncertainty in phylogeographic reconstructions [5], which makes it difficult (and often even leads to meaningless conclusions) to study the fauna evolution.

Identification of non-indigenous species outside their historical ranges, as well as corridors and vectors of biological invasions, is essential for rational environmental management and the implementation of sustainable development goals [6]. An important task of biodiversity studies is to determine the species composition quickly and qualitatively with an accurate isolation of any non-indigenous component. At the same time, early identification of invasive species is difficult because of their initial paucity,

and objective difficulties of the morphological identification of species that are new for a particular region. Therefore, the first reports of alien species often appear when the invaders have already successfully naturalized in new ecosystems. This fact makes complicated the management of biological resources in order to reduce the harmful impact of the invaders on indigenous fauna. Thus, the effectiveness of biological invasion monitoring and developing methods for controlling (and combating with) unwanted species is closely related to their fast and accurate identification [7].

It should be noted that rapid and high-quality identification of species of living organisms using a certain universal approach, available to all researchers without restrictions, could be regarded as one of the main goals of taxonomy. Currently, DNA barcoding is regarded sometimes as the *ultima ratio* for animal identification [8]. Sequences of the first subunit of mitochondrial cytochrome oxidase *c* (COI, COX) gene has a sufficient resolution for taxon identification [9]. The number of publications on DNA-based species barcoding has skyrocketed recently. For example, a Google Scholar search for "fishes AND barcoding AND identification AND [COI OR COX]" returns over 11,000 papers published during the past five years, and the number of such studies is increasing constantly.

The popularity of such a method is associated with the use of "universal" primers suitable for amplification of the 5'-region of COI, about 600-700 base pairs (bp) in length in many animal species. Traditionally, the 25/26 nucleotide primers LCO 1490 and HCO 2198 are used [10]. However, genetic differences between different groups of animals determine the low specificity of the primer for the hybridization site on the DNA template, which is reflected in a relatively low level of amplification success or complete absence of a PCR product even in case of a sufficient amount of DNA and its quality in template [11]. Therefore, the next logical step is the development of specific sets of primers for specific animal groups. For example, for fishes, such primers were developed originally for Australian species [12], then a "universal cocktail" was proposed [13] based on the mitochondrial genomes of economically important fish groups.

It should be noted that the primer sets mentioned above have both advantages and disadvantages. For example, use of the fish primers recorded above [12] requires up to 40 amplification cycles for each sample, as well as several amplification rounds, which complicates greatly their routine use. The use of degenerate "cocktail" sequences [13] significantly increases the amplification efficiency. However, the other face of this "omnivorousness" is the increased requirements for the quality of the matrix, the need to select the optimal PCR parameters and a high frequency of nonspecific annealing of the primers. Thus, for a specific fish group, the need for a balance between specificity and universality is obvious: if a researcher is aiming to work with freshwater European fishes only, there is no point in enriching the primers with degenerate regions specific to marine and tropical fishes.

Optimizing primers for the COI fragment for fast and high-quality diagnostics of fish dispersing in freshwater reservoirs of Europe, with an emphasis on alien fish of the largest aquatic system in Europe – the Volga-Kama, is a relevant direction of studies. In addition, the selection of primers for other regions of mitochondrial (mtDNA) and nuclear DNA (nuclDNA), which can be used successfully for diagnostics of fish species, also is necessary. Moreover, apart from the high-quality selectivity of primers, special attention should be paid to the optimization of the method costs.

In this study, we assess the efficiency of primer sets for the COI fragment (0.7 kb) used in "classical" DNA barcoding [12]; for the metabarcoding of a short 3'-fragment (0.3 kb), suitable for using in high-throughput sequencing systems [14]; primers for the 16S fragment of mitochondrial rRNA, including those designed to fill the library of reference sequences for work on the metabarcoding of communities and eDNA studies [15]; primers for a fragment of 18S nuclear rRNA including two hypervariable regions V1-V2, valuable for animal phylogeny [16]; as well as protocols for the cost-effective isolation of total DNA and purification of the PCR product without the use of commercial kits.

## 2. Materials and Methods

### *Sampling*

The sampling was carried out during coastal fish surveys in the water bodies and watercourses of Central Russia, mainly in the Volga-Kama basin [17] in accordance with the state permit issued to the IBIW RAS for catching biological resources. Some samples were provided by regional environmental inspectors. The dates and coordinates of the sampling sites are given in Supplement file 1. For molecular genetic analysis, a portion of the caudal fin blade or a piece of skin and muscles behind the dorsal fin was taken from each specimen and fixed in 95% ethanol cooled to -20°C. In this case, the marked voucher was fixed in 4% formalin for subsequent morphological analysis. Fish larvae, as well as fragmented samples, were fixed entirely in 95% ethanol, while the volume of ethanol was not less than 10 volumes of the sample. In toto, more than 700 specimens from 25 species were analyzed. At the same time, 188 samples were used for the analysis of COI polymorphism, and 94 samples were used for the analysis of other loci (calculated for standard 96-well plates, taking into account control wells). The samples were stored at + 4°C in the dark.

### *DNA extraction*

Genomic DNA was isolated using the Wizard Genomic DNA Purification Kit (Promega Corp., USA) and QuickExtract DNA Extraction Solution (Epicenter by part Illumina Inc., USA) according to the manufacturer's recommended protocols. After extraction, the concentration and purity of DNA were determined by measuring the optical density at  $\lambda$  260/280 nm on a microspectrophotometer N50 (Implen GmbH, Germany). Also, in order to minimize the cost of DNA extraction without the use of commercial kits, we can recommend a salt method for the isolation of nucleic acids without the use of expensive proteinase K, based on [18] with modifications. All manipulations, if not specified separately, are carried out at room temperature. The sample is taken from 300 mkg of alcoholized or frozen fish tissue (preferably muscles and skin), dried on filter paper and crushed into standard 1.5 ml microtubes. To each tube is added 500 mkl of sterile saline buffer preheated to + 60°C [0.5 M NaCl; 100 mM Tris-HCl pH 8.0; 5 mM EDTA pH 8.0] in which the fish tissues are ground until the finest homogenate is obtained. The presence of bone debris or scales in the sample does not affect the efficiency of DNA extraction. After grinding, 100 mkl of the mixture [10% SDS; 1%  $\beta$ -mercaptoethanol] is added. For more efficient lysis of especially valuable samples to improve the quality of recovery, an aqueous solution of proteinase K can also be added to a final concentration of 100 mkg/ml. The mixture is thoroughly mixed and incubated at + 60°C in a thermostat until the tissues are completely dissolved (depending on the sample, this process takes from 1 to 8 hours). During the lysis process, the samples are intensively mixed every 15 minutes on a vortex, after the drops are discarded by short-term centrifugation and the tubes are again placed in a thermostat (the best effect is achieved when using a heated shaker-incubator). Next, 300 mkl of an aqueous solution of 5 M NaCl is added to each sample, the mixture is intensively mixed on a vortex for 30 seconds, after which it is centrifuged for 20 minutes at 16000 g. The supernatant in a volume of 600 mkl is carefully (without stirring up the sediment) into individual clean microtubes, and an equal volume of 96% ethanol cooled to -20°C is added to it. The DNA precipitation takes place at -20°C for 1 hour. At the end of this stage, the samples are centrifuged for 20 minutes at 16000 g and the supernatant is carefully removed. The precipitate containing nucleic acids is washed with 600 mkl of 80% ethanol cooled to -20°C and the samples are finally centrifuged for 20 minutes at 16000 g. Depending on the amount of DNA, the washing step can be skipped, although this will somewhat reduce the purification quality, but will significantly increase the product yield. After removing the supernatant, the sediment is briefly dried (2 minutes at + 60 ° C) and then dissolved in 100 mkl of sterile water. The isolation quality by the  $\lambda$  260/280 ratio is 1.2-1.8, which allows the nucleic acid solution obtained in this way to be stored at -50°C for up to a year with practically no DNA degradation.

### Primers design

For the design of primers, we selected the complete sequences of the fragments of interest from the GenBank database (NCBI). For the analysis of the COI and 16S loci, 16 complete mitochondrial genomes of fish of the families Clupeidae, Cyprinidae, Gobiidae, Percidae, Salmonidae were selected. For the 18S locus, six complete sequences of the small subunit of nuclear ribosomal RNA of the families Clupeidae, Cyprinidae, Percidae, Salmonidae were taken. The sequences for each locus were aligned using the MAFFT v.7 algorithm [19] integrated into the Unipro UGENE v.38.1 package [20]. The target region was chosen for COI as a 0.7 Kb 5'-region corresponding to the standard fragment for DNA barcoding of fish [21]. A variable 3'-region of a standard fragment [14] with a length of about 0.3 Kb, which gives good prospects for use both in classical PCR and on NGS platforms was used for metabarcoding using COI. A region of 16S with a 5'-end of about 0.6 Kb was selected as potentially highly informative for the purposes of species identification of fish [15], while a large number of sequences for 16S are generated by metabarcoding of communities using high-throughput sequencing technologies [22]. The region of the 18S nuclear rRNA locus with a total length of about 0.5 kb, including the hypervariable regions V1-V2, although it shows less diversity than mitochondrial loci [16], but gives good results for the identification of almost all large groups of animals [23].

The primers used are represented in Table 1. For the primers, modified M13-tails were added to the COI gene fragment at the 5' end [24]. In the case of highly degenerate primers with inosine in the composition, this ensures a guaranteed high-quality sequence, preventing the formation of primer and product polymers and providing optimal parameters for the Sanger sequencing reaction [25].

**Table 1.** Genes, primers and annealing temperatures used in this study. M13 tails for sequencing are highlighted by lower case type.

Gene	Primer	Sequence 5'-3'	Annealing temperature (°C)	Full amplicon length (Kb)
COI (long)	MifCOI-F	tgtaaacgacggccagtt CACAAAGAVATTGGYACCCITITA	52	0.7
	MifCOI-R	caggaaacagctatgacta CIGGGTGICCRAARAAYCARAAIA		
COI (short)	ifCOImb-F	GGDACHGGNTGAACDGTHTAYCCBCC	56	0.4
	ifCOImb-R	TGDCCAAARAAYCAGAABARGTGATG		
16S	if16Sa	CTGCCCTGTGACCAAAAAGTT	52	0.6
	if16Sb	GGTCCTTTCGTACTAGGAAG		
18S	if18S1	TAACATATGCTTGTCTCAAAG		
	if18S2	CCTGTATTGTTATTTTTTCGTCAC		
M13s	M13seq-F	TGTAACGACGGCCAGTT	sequencing	
	M13seq-R	CAGAAACAGCTATGACTA		

### PCR amplification

Amplification was performed in individual 250 µL wells of a standard 96-well plate in a T-100 amplifier (Bio-Rad Laboratories Inc., USA), as well as in individual 600 µL microtubes in a BIS M111 amplifier (BIS Co., Russia). The polymerase chain reaction was carried out in the volume of 25 µL reaction mixture composed 1 µL DNA template (about 20 ng / µL), 1 unit standard Taq DNA polymerase, SE-buffer [60 mM Tris-HCl (pH 8.5); 2.5 mM MgCl<sub>2</sub>; 25 mM KCl; 10 mM β-mercaptoethanol; 0.1% Triton X-100 surfactant], 0.25 mM each of deoxy nucleotide triphosphate (dNTPs) (all the reagents were

produced by SibEnzyme Co., Russia) and 0.5 mM of each primer set. We used touch-down Polymerase Chain Reaction [26], which reduces the effect of non-specific primer binding and increases the yield of the target product. The PCR protocol consisted of the following steps: primary denaturation for 3 minutes at + 95 °C; 10 cycles of the "ladder" stage, consisting of 30 seconds of denaturation at + 94 °C, 45 seconds of primer annealing at a temperature of + 58 °C (+ 60°C for mb\_if COI primers) in increments of -1 °C per cycle, and a step elongation 80 sec (60 sec for mb\_ifCOI primers) at + 72 °C. This was followed by 30 cycles of basic PCR, consisting of 30 seconds of denaturation at + 94 °C, 30 seconds of primer annealing at the appropriate temperature (Table 1), and an elongation step of 60 seconds (40 seconds for mb\_ifCOI primers) at + 72 °C. At the end of the PCR, a final elongation step at + 72 °C for 5 min followed, followed by storage at + 12 °C.

The presence of the PCR product was checked by electrophoresis in 1.2% agarose gel in TBE buffer (pH 8.2) composed of THAM 89 mM, boric acid 89 mM and EDTA 2 mM. The approximate molecular weight of the product was determined by comparison with the standard 100 bp + 1.5 Kb + 3 Kb DNA Ladder (SibEnzyme Co., Russia). DNA was visualized in UV light on a Kvant-312 transilluminator (Helicon Co., Russia) with a UV wavelength of 312 nm with preliminary staining of the gel in a 0.1 mM aqueous solution of ethidium bromide for 10 min and subsequent washing of the gel in distilled water for 15 minutes.

#### *DNA sequencing*

The success of sequencing is largely determined by the quality of the PCR product purification from primers and deoxyribonucleotide triphosphates. There are many commercial kits available today that provide excellent product purification results. We used QIAquick PCR Purification Kit spin columns (Qiagen N.V., Netherlands) and ExoSAP-IT PCR Product Cleanup Reagent (Thermo Fisher Scientific Inc., USA). In addition, in the presence of a high-quality monomorphic PCR product, we used a simple method of alcohol reprecipitation under "mild conditions", which was successfully used earlier [27]. In this case, ethanol is added to the amplification mixture to a final concentration of 70% and ammonium acetate to a final concentration of 125 mM. The mixture is gently mixed and reprecipitation proceeds at room temperature for 20 minutes. At the end of this stage, the samples are centrifuged for 20 minutes at 16000 g and the supernatant is carefully removed. The precipitate containing nucleic acids is washed with 600 μl of 80% ethanol cooled to -20 °C and centrifuged again for 20 minutes at 16000 g. After removing the supernatant, the sediment is briefly dried (2 minutes at + 60 °C) and then dissolved in 20 μl of sterile water. The thus purified PCR product, after determining the DNA concentration, was prepared for sequencing using the BigDye Terminator v3.1 Cycle Sequencing Kit (Thermo Fisher Scientific Inc., USA). Sequencing of sense and anti-sense strands of PCR products was carried out on an Applied Biosystems 3730 DNA Analyzer automatic sequencer (Syntol Co., Russia). Bidirectional sequences were assembled in Sanger Reads Editor add-on of Unipro UGENE and manually edited. All obtained sequences were preliminarily verified to be fish in the GenBank database (NCBI) using the mBLAST query [28]. The sequences were deposited with GenBank under the numbers MT833701 – MT833840 (COI), MZ005797 – MZ005873 (16S) and MZ005706 – MZ005796 (18S).

#### *Alignment, nucleotide diversity and phylogenetic analysis*

Sequence alignment was performed using the MAFFT v.7 algorithm [29] on the Computational Biology Research Center server, Japan (<http://mafft.cbrc.jp>). The alignment of each locus occurred independently. For the sequences of the protein-coding COI locus, the "Translation Align" option with the FFT-NS-i strategy was used. To align the sequences of ribosome-coding loci during alignment, the secondary structure of the molecule was taken into account according to the Q-INS-i strategy.



Nucleotide diversity indices [30] and neutrality tests were calculated using the DnaSP v.6.12 software [31]. We conducted neutrality tests  $F_s$  [32] and  $D$  [33], which together provide sufficient information both to identify neutrality and to describe demographic processes [34,35].

ModelFinder v.1.6 [36] on the web-portal of Center for Integrative Bioinformatics Vienna, Austria (<http://www.iqtree.org>) [37] was used to search for the best model of nucleotide substitutions. For the COI locus, the substitution pattern was identified for each nucleotide position in the codon (1st, 2nd, 3rd). The selection of the most suitable model is based on the minimum values of the Bayesian information criterion, BIC [38]. It should be noted that the parameters of the BIC model were almost identical to those determined on the basis of corrected Akaike's information criterion, AICc [39], which may indicate a high agreement of the calculated models with the real best model.

According to the parameters of the selected model of nucleotide substitutions (Table 2), the phylogenetic tree was reconstructed for each locus. For maximum likelihood (ML) analysis, we used the IQ-TREE v.2.1 algorithm [40]. As a branch support test, we used 10k replicas of the UFboot2 bootstrap test [41], which takes up significantly less computational resources and is highly efficient compared to traditional tests. Topology estimation for ML trees is based on 1k replicates of SH-aLRT test [42], calculations performed on W-IQ-TREE server [37]. The ML tree constructed from the initial data is a realized (and not true) phylogenetic tree, and for such a case there is no unambiguous opinion about the correctness of using topological tests to check the monophyly of certain branches. However, in combination with a standard bootstrap, this procedure can be useful for assessing the group monophyly [30]. Reconstruction of the phylogeny using a stochastic approach (Bayesian inference, BI) was carried out using the BEAST2 v.2.6 software package [43]. Through the BEAUti tool [44], all parameters of the models of nucleotide substitutions identified by ModelFinder were recorded. Based on the ML-test for the presence of a molecular clock, implemented in MEGA-X [45], the null hypothesis of equal evolutionary rate throughout the tree was not rejected at a 5% significance level. The task of this work does not require establishing of the exact phylogenetic relationships between species; therefore, the strict molecular clock evolutionary model with the priority of speciation according to the Yule process was chosen as the most suitable for the datasets covering several species [46]. Each analysis includes six independent runs of MCMC, 50M generations each, and selecting of each 50k tree. The effectiveness of MCMC on the convergence of the results of all independent runs with the estimated effective sample size (ESS) for all parameters above 200 was carried out in the Tracer v.1.7 program [47]. After combining the results of all MCMC runs through the LogCombiner, a consensus tree was computed based on the maximum confidence clade (MCC) using the Tree-Annotator [44] with 25% burn-in. After finding consistency in the main clades between BI and ML, the illustrations show only the ultrametric BI trees.

**Table 2.** Models of nucleotide substitutions

Locus	Position	Best model (BIC)
COI (long)	1st	TIM3e {0.1698, 2.1566, 12.5089} + FQ + G4 {0.1709}
	2nd	F81 + F
	3th	TIM3 {0.7556, 9.7639, 4.9699} + F {0.2885, 0.3166, 0.1141, 0.2805} + R3 {0.0078, 0.0093, 0.6312, 0.5571, 0.3608, 1.7961}
16S		TPM2 {3.2961, 8.5094} + F {0.2888, 0.2442, 0.2371, 0.2298} + G4{0.2244}
18S		TIM2e {0.4296, 0.2174, 3.6492} + FQ + G4 {0.1817}

Base substitution rates: F81 – Felsenstein's model [74], variable base frequencies, all substitutions equally likely; TIM2e – transition model with equal base frequencies and AC=AT, CG=GT; TIM3 – transition model with unequal base frequencies and AC=CG, AT=GT; TIM3e – transition model with equal base frequencies and AC=CG, AT=GT; TPM2 – three parameter model with equal

base frequencies and AC=AT, AG=CT, CG=GT. Base frequencies: +F – empirical base frequencies; +FQ – equal base frequencies. Rate heterogeneity across sites: +G4 – discrete Gamma model [75] with four categories; +R – FreeRate model [76] that generalizes the +G model by relaxing the assumption of Gamma-distributed rates. Non-standard model parameters are indicated in {}.

#### *Species delimitation*

Initially, all sequences obtained in the course of this work were individually compared with the records in the NCBI Taxonomy Database [48] and BOLD BINs [49] (COI only). In case of insufficient data on reference sequences or insufficient resolution of the method, a delimitation procedure based on genetic data was used, which is more related to “integrative taxonomy”. Initially, the level of the “gap” between species was determined based on genetic differences. The implementation of this approach based on genetic distances was performed in the ASAP application on the web-server Atelier de BioInformatique, France (<https://bioinfo.mnhn.fr/abi/public/asap/>) [50]. The algorithm uses a number of preliminary intraspecific divergences to derive a one-way confidence interval from the data based on an interspecies divergence model. The set of species was determined for the COI locus based on the “simple” uncorrected p-distance, as more preferable for the purposes of DNA barcoding [51]. The delimitation scheme was determined by the best asap-score with the minimum P-val. In addition, another approach based on distances and implemented in the “divergence threshold optimizing and clustering approach”, locMin [52], was used for species delimitation. The calculations were performed on the COI gene tree using the algorithm [53] in the “Microsoft R-Open and MKL” 64-bit v.3.5 software (<http://mran.microsoft.com/>). This implementation is suitable for single-locus studies, correlates well with morphological data, and at the same time is not prone to excessive taxa fragmentation [54].

Since the methods based on the search for the “gaps” between species are well developed only for the COI locus, other delimitation methods were applied for rest of the loci that we analysed. In particular, probabilistic realizations based on Bayes' theorem and Poisson's distribution law were used, requiring the use of complex mathematical models. The first, the Generalized Mixed Yule Coalescent (GMYC) method, is based on a stochastic principle, which assumes that the branch points in the reconstructed ultrametric “gene tree” correspond to one of two events: divergence events between species level taxa (modeled by the Yule process), or stochastic events between intraspecific lines (using the coalescence model). Since the rate of coalescence within species is expected to be significantly higher than the rate of cladogenesis, GMYC seeks to find a demarcation between these types of branching [55]. For example, we analyzed Generalized Mixed Yule Coalescent (GMYC) in “Microsoft R-Open and MKL” software with ‘splits’ package on consensus ultrametric gene trees [56]. In addition, another model was used, the Poisson Tree Processes model, aimed at distinguishing between speciation processes among species and processes of diversification within species. This method is similar to GMYC, the latter analyzes the number of substitutions between branching events instead of time intervals [57]. The mPTP calculation was performed on individual ML gene trees on the Heidelberg Institute for Theoretical Studies web server (<http://mptp.h-its.org/>).

All original materials, namely DNA sequences, alignments, phylogenetic trees and images used in this study, are publicly available in the Open Science Framework repository [58] at the project address <https://osf.io/b8qfd/>.

### **3. Results**

#### *3.1. Comparison of the effectiveness of different methods of DNA extraction and purification of the PCR products*

The isolation on spin columns gave best DNA quality in terms of the  $\lambda$  260/280 1.8-2 ratio. However, this is the most expensive method for DNA isolation and purification. The use of QuickExtract gave the highest quantitative yield of nucleic acids, but in this

case, after isolation, the sample contains a lot of peptides, DNA and RNA fragments. Therefore, such an express method may be recommended for a direct PCR, without long-term storage of the extract. The salt method also showed good quality of the isolated DNA: the quality of the isolation according to the  $\lambda$  260/280 ratio is 1.2-1.8, which allows the nucleic acid solution to be stored at  $-50^{\circ}$  C for up to a year almost without a DNA degradation. All these methods provide a sufficient yield of the target product; however, the rather high cost and labor intensity of the process (when using spin columns) and the narrow temperature range of working with enzymes limit the usefulness of the commercial kits.

Purification of the PCR product and the sequencing reaction product is a key element for successful Sanger sequencing. Undoubtedly, the best homogeneous product may be obtained by gel purification with excision of specific bands and DNA extraction from the gel followed by isolation on spin columns or magnetic particles. However, this is an expensive and labor-intensive process that is not suitable for a mass analysis. An alternative method of PCR product purification, which is gaining popularity today due to its speed and ease of use, is based on the use of a set of endonucleases and alkaline phosphatase. However, this solution also has its limitations, primarily associated with the transportation and storage of enzymes for the reaction, as well as a low efficiency in relation to primer dimers and short DNA duplexes. The alcohol method we used for re-precipitation of DNA under "mild" conditions may not be ideal in terms of time and purity of isolation, but it gives a poor product at a minimum cost of reagents and fairly quickly. It is also necessary to note that this method removes completely the unbound dNTPs and primers, even in the form of duplexes up to 50 b. p., which is important for successful sequencing [59].

### 3.2. Comparison of the amplification efficiency of different primers sets

The success of PCR is ensured not only by the quality of the DNA-matrix, but also by the factors that determine the efficiency of primer hybridization and the conditions of the PCR reaction [60]. Unfortunately, for the high universality of primers, one often has to pay for their low specificity, or even the absence of a PCR product at all. To solve this problem, even in the presence of a high-quality DNA matrix, it is often necessary to use a set of several pairs of primers, their combinations, or special efforts on modeling of some unique specific primers [61]. Thus, it makes sense to develop less universal but more specific primers for the target group of organisms [13].

The results of our work using the MifCOI kit showed a high efficiency of PCR for freshwater fish in the European part of Russia. However, the sequencing success was slightly lower than expected (Table 3). It should be noted here that by "specific sequencings success" we specifically mean the exact match between DNA identification by COI and other loci, as well as by morphological characteristics of vouchers. It is likely that, in this case, the incomplete efficiency is explained by DNA degradation (some of the delivered samples were poorly preserved) and contamination from other fish specimens during a total sample preservation. Presumably, the problem of the contamination by DNA from other organisms is quite common, which has consequences even for the international databases [62,63]. Thus, to improve the accuracy of DNA identification, it is desirable to use several loci and, if possible, to study carefully the morphological characters of the vouchers. This is the only way to obtain a high-quality library of sequences that is unambiguously corresponded to a particular species [64]. Thus, having eliminated the unambiguously "bad" samples, we tested the ifCOI<sub>mb</sub> primer set to amplify a shorter product, similar to the meek reads used in the metabarcoding method [14]. Despite doubts about the universality of the COI locus, the latter, even when using incomplete fragments, provides reliable data for the animal species identification [65]. Our ifCOI<sub>mb</sub> primer set shows the highest level of efficiency for the DNA identification in fishes (Table 3). Using of "long" and "short" sets of sequences from our material led to similar topologies of the reconstructed trees (Figure 1). The amplicon length and design of these pri-

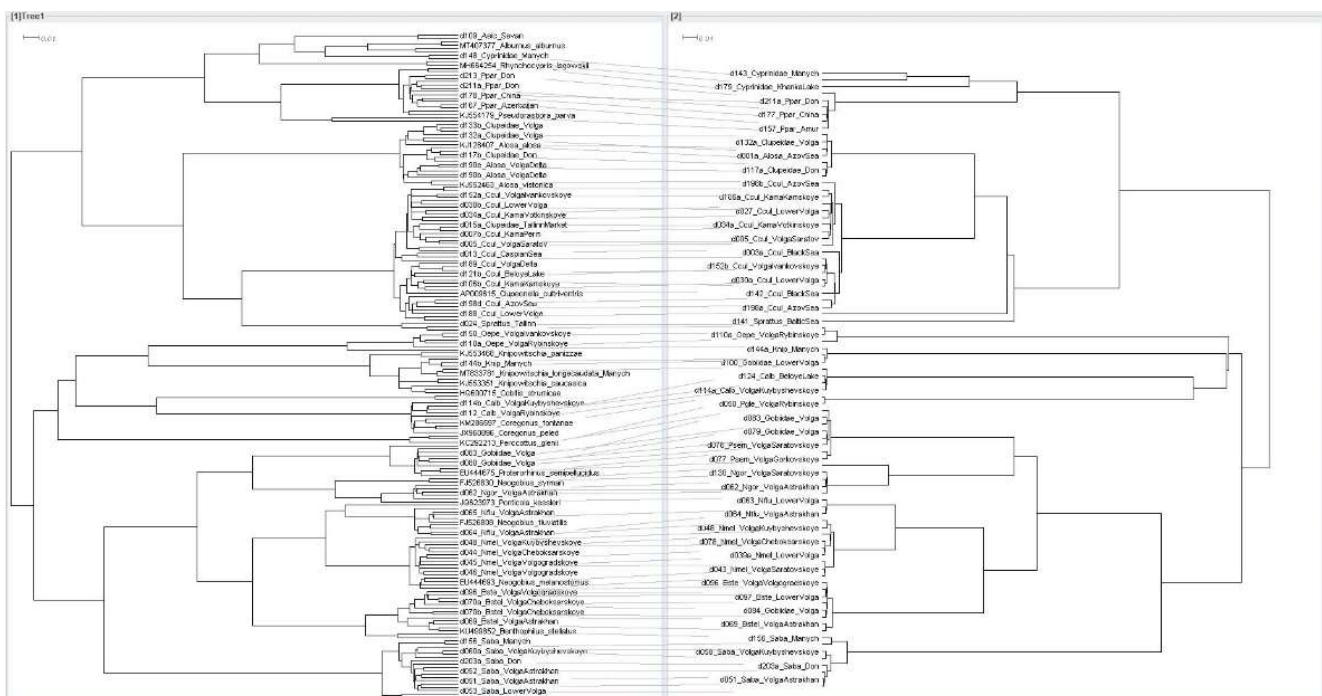


mers are suitable for modern high-throughput sequencing platforms [66] and may be used both for routine molecular research and work with community DNA or eDNA. Another advantage of these short fragments is less stringent requirements for the quality of the DNA matrix, which allows to analyse the samples with poorly preserved and fragmented DNA (that often arrives from the environmental services). Finally, the use of a fragment with a length of about 0.4 Kb may significantly reduce the amplification time and the spending of reagents for sequencing.

**Table 3.** Amplification and sequencing success with four primers sets

Primers set	Percent of PCR success	Percent of specific sequencings success
MifCOI	91	77
ifCOImb	98	96
if16S	98	81
if18S	97	95

The PCR with primers for the mitochondrial ribosomal large subunit (primers if16S) almost always gives a product. However, due to a high conservatism of the 16S gene, it demonstrates a very high rate of a nonspecific product, which is generally typical for studies conducted with this locus [67]. This locus does not provide fundamentally different information (in comparison with COI) about the species of animals, and its use may not be justified for mass analysis. At the same time, it seems expedient to accumulate data on species and, especially, intraspecific variability for this gene in order to form a comparison in course of the metabarcoding and eDNA studies.



**Figure 1.** Tanglegram for mitochondrial COI sequences "long" (left) and "short" (right) phylogenetic trees. All pictures in high resolution available at <https://osf.io/b8qfd/>.

The nuclear locus of the small ribosomal subunit used in our study (primers if18S) almost always gives a specific PCR product (Table 3). The only drawback of this locus is its rather low species variability [16] and poor representation of earlier obtained sequences in international databases. However, the presence of a conserved regions en-

sures efficient alignment of these sequences, and the hypervariable regions V1-V2 provide a fairly high level of variability. Also, the results of this locus analysis show a high similarity with results based on mitochondrial genes, hence, 18S can be successfully used in phylogeographic reconstructions [68]. An additional advantage of this locus is its multiple copies (in contrast to most protein-coding genes) and the absence of individual polymorphism (in contrast to internal spacers). Therefore, the 18S study may be a good addition to the "classical" DNA barcoding, including the using of high-throughput sequencing systems [23].

### 3.3. Polymorphism and nucleotide diversity of the studied loci

In terms of the level of genetic diversity, the alien fish species we studied in the European part of Russia differ significantly (Table 4). Mitochondrial loci exhibit a greater genetic variability than nuclear 18S locus, which is typical of other animals [69]. In general, the nucleotide diversity depends more on the number of species represented in the family than on a specific locus. This is especially noticeable for mtDNA, which is virtually all represents one linkage group and may even be considered as one locus [70]. This observation is confirmed by comparing the *F<sub>s</sub>* and *D* neutrality indices, which, in this case, reflect the level of differentiation between species within fish families. Both mitochondrial loci, in contrast to the nuclear 18S, are characterized by a relatively low proportion of "refractory" G + C nucleotides, which is generally characteristic of animals [71].

**Table 4.** Metrics of genetic diversity from mitochondrial and nuclear loci in the studied alien fish

Loci / Orders	N	n	G+C	S	Eta	h	Hd	Pi	k	Fs	D
COI (mitochondrial)											
Clupeiformes	55	669	0.486	173	206	36	0.97	0.087	58.2	1.81	1.04
Cypriniformes	17	669	0.451	172	220	16	0.99	0.079	53.2	-0.63	-0.78
Gobiiformes	49	669	0.478	233	342	24	0.94	0.129	86.7	15.9	0.47
Osmeriformes	4	669	0.486	8	8	3	0.83	0.006	4.00	1.16	-0.82
Perciformes	1	669	0.461	0	0	1	-	0	-	-	-
Salmoniformes	5	669	0.495	4	4	3	0.70	0.003	1.80	0.46	-0.41
Syngnathiformes	15	669	0.476	21	21	9	0.84	0.008	5.10	-1.04	-0.86
COI (total)	146	669	0.478	277	515	92	0.98	0.177	118.8	1.21	0.92
16S (mitochondrial)											
Clupeiformes	39	577	0.502	70	82	12	0.73	0.028	16.22	6.93	-0.59
Cypriniformes	19	576	0.450	32	33	11	0.78	0.016	9.43	0.21	-0.01
Gobiiformes	19	581	0.486	121	154	11	0.93	0.070	39.78	6.40	-0.40
16S (total)	77	584	0.483	176	261	32	0.91	0.119	66.24	14.8	0.85
18S (nuclear)											
Clupeiformes	43	458	0.559	15	16	7	0.70	0.010	4.70	3.75	0.86
Cypriniformes	7	486	0.573	11	11	4	0.71	0.012	5.81	2.07	1.58
Gobiiformes	27	445	0.546	11	12	7	0.63	0.004	1.80	-0.95	-1.41
Osmeriformes	4	463	0.555	0	0	1	-	0	-	-	-
Perciformes	1	442	0.545	0	0	1	-	0	-	-	-
Salmoniformes	4	441	0.540	2	2	3	0.83	0.002	1.00	-0.88	-0.71

Syngnathiformes	5	417	0.581	1	1	2	0.60	0.001	0.60	0.62	1.22
18S (total)	91	486	0.546	51	59	19	0.87	0.041	15.60	6.80	1.11

N – number of sequences; n – total number of sites (excluding sites with gaps or missing data); G+C – guanine and cytosine content; S – number of segregating (polymorphic) sites; Eta – total number of mutations; h – number of haplotypes; Hd – haplotype (gene) diversity; Pi – nucleotide diversity per site; k – average number of nucleotide differences; Fs – Fu’s neutrality statistic [32]; D – Tajima D neutrality test [33], all values are not statistically significant  $P < 0.05$ .

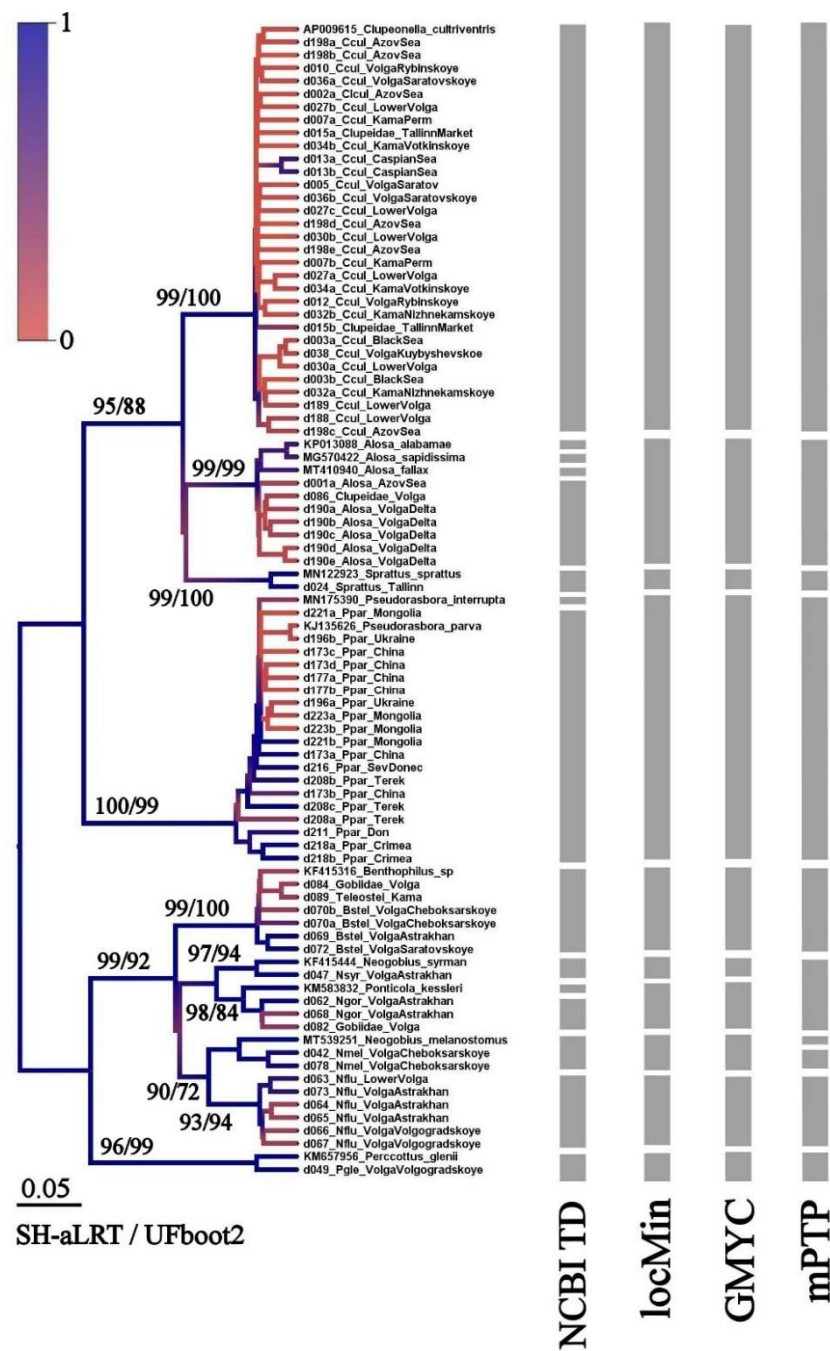
The study of the nucleotide substitution models (Table 2) showed a complicated pattern of the nucleotide substitutions. As a rule, for the protein-coding COI locus, the second codon position is more conservative, which is typical for other animals [72]. In general, ribosomal loci, despite their different localization, demonstrate a relatively similar pattern of the nucleotide substitutions. This may be explained by the peculiarities of the formation of ribosomal domains of duplexes and loops, which provides additional information for their use in phylogenetic reconstructions [73].

### 3.4. Results of species differentiation based on DNA analysis

As a result of the species delimitation based on sequences of all loci using different approaches, it is possible to identify successfully all alien fish species which have penetrated the European part of Russia. On the other hand, there is a problem of contamination of the vouchers by foreign DNA, and also of some misidentifications, especially of closely related and poorly distinguished species [77]. The identification is made more complicated by continuous “taxonomic innovations”: revisions of taxonomic status, global revisions of some species groups and, often unjustified, “splitting” of species [78]. For an ordinary routine user, the task is simplified in the case when it is necessary to identify a particular specimen up to level of a particular species, without penetrating deep into taxonomic nuances. In the simplest approach, the sequences of interest are compared with certain reference sequences. Currently, a huge amount of sequences has been accumulated in two international databases: GenBank National Center for Biotechnology Information ([www.ncbi.nlm.nih.gov/genbank/](http://www.ncbi.nlm.nih.gov/genbank/)) [79] and BOLD: The Barcode of Life Data System (<http://www.boldsystems.org/>) [80]. Complete openness and presence of the information on multiple genes are the advantages of the GenBank. BOLD’s attention is focused primarily on the COI sequences, with a significant portion of the data being private. However, in any case, the primary definition of a voucher belongs entirely to the author of the study, and automatic filters for initial screening are still under development [63] and are not yet very effective. As a result, the application of this “comparative” method for assessing biodiversity shows a significant variation between the NCBI GenBank and BOLD (Figure 2). Therefore, it becomes necessary to identify some alternative ways of estimation the species richness, rather than a simple comparison with some “reference” samples, as well as to highlight the “molecular operational taxonomic units”, which more or less correlate with real “biological” species [81].

The easiest way to isolate mOTUs is based on the search for a “barcoding gap”, the level of which for different species according to COI is 1-15% [9], but it is desirable that the sequence differences between two species exceed the intraspecific sequence differences by an order of magnitude [82]. The implementation of a similar approach, performed according to the ASAP and locMin algorithms (Figure 2, Figure 3, Figure 4), significantly reduces the species richness in the studied dataset, in comparison with the “comparative” approach.



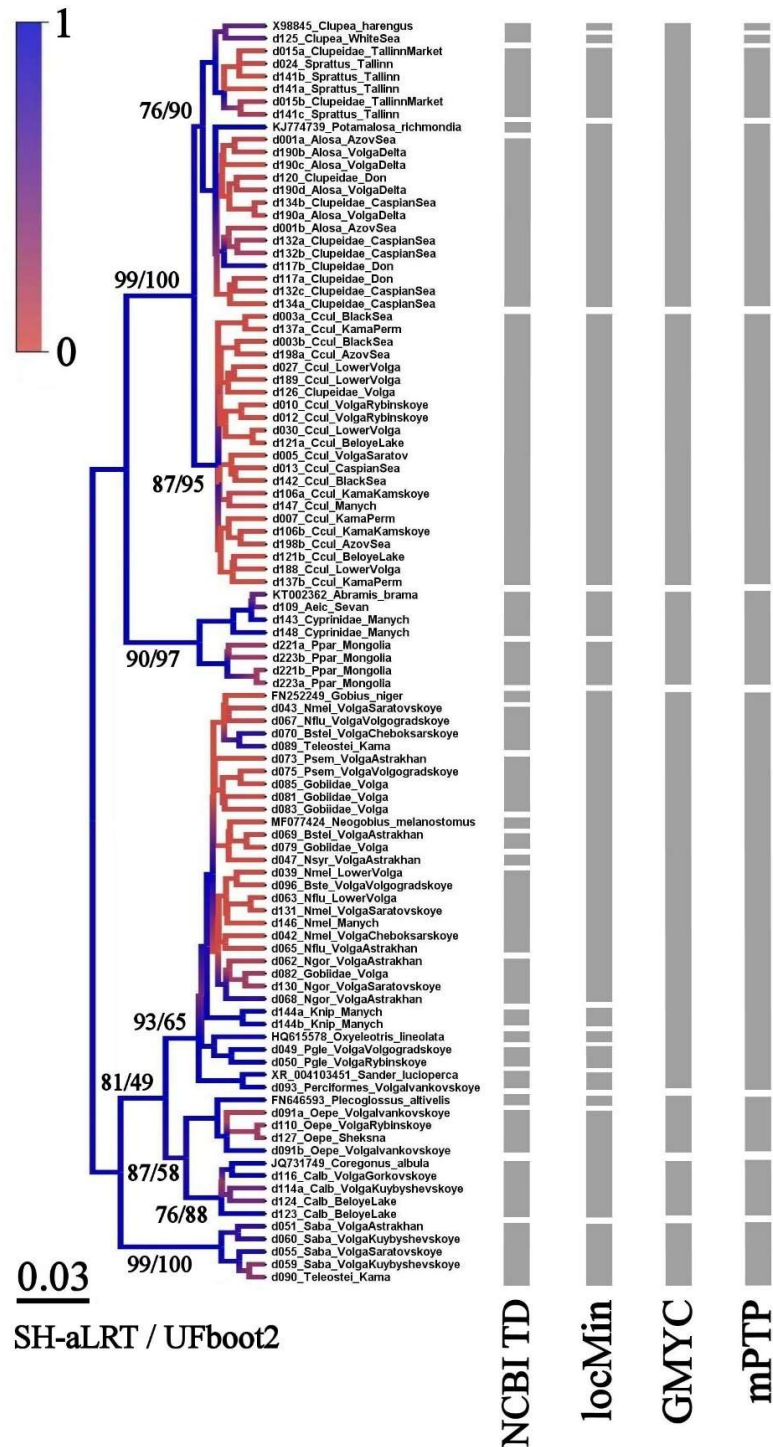


**Figure 3.** Molecular species delimitation via different methods. Shown the BI tree for mitochondrial 16S locus. Gray columns indicate group membership of specimens. Node supports are posterior probabilities indicated as coloration, SH-aLRT test and UFboot2 in percent.

At present, with the accumulation of knowledge about genetic variability in different animal species, we observe a gradual rejection of such an attractive and simple idea (which is served as the foundation for the “DNA barcoding” ideology). It should be noted that even Paul Hebert, the founder of this approach, noted too high variability of the regions of mitochondrial and nuclear genes used to differentiate species [9]. Therefore, recently, the delimitation methods based on genetic distances along a certain marker



(usually mitochondrial) gene are gradually being replaced by the analysis of several loci (not only mitochondrial, but also nuclear) and more complex mathematical algorithms of the species delimitation [83].



**Figure 4.** Molecular species delimitation via different methods. Shown the BI tree for nuclear 18S locus. Gray columns indicate group membership of specimens. Node supports are posterior probabilities indicated as coloration, SH-aLRT test and UFboot2 in percent.

In cases of other markers, such as, e. g., ribosome-coding genes, the implementation of this approach is still required the accumulation of a large data sets. Therefore, at present, methods that are not based on genetic distances, but on the use of complex mathematical models to describe a hierarchical biodiversity structure, are gaining more and more popularity. This includes probabilistic realizations based on Bayes' theorem or Poisson's distribution law. Usually, stochastic models tend to be too fragmented and strongly depend on the level of genetic variability of the studied samples [84,85]. However, for our dataset, such a tendency was not observed for these methods, and the number of isolated mOTUs is even less than the valid species taken for the study (Figure 2, Figure 3, Figure 4). This feature is probably associated with a relatively low nucleotide diversity in fish, where intraspecific differences usually do not exceed 3% [86,87], in contrast to invertebrates, where the intraspecific level may reach 20% [85,88,89].

#### 4. Discussion

##### *DNA manipulation, amplification and sequencing*

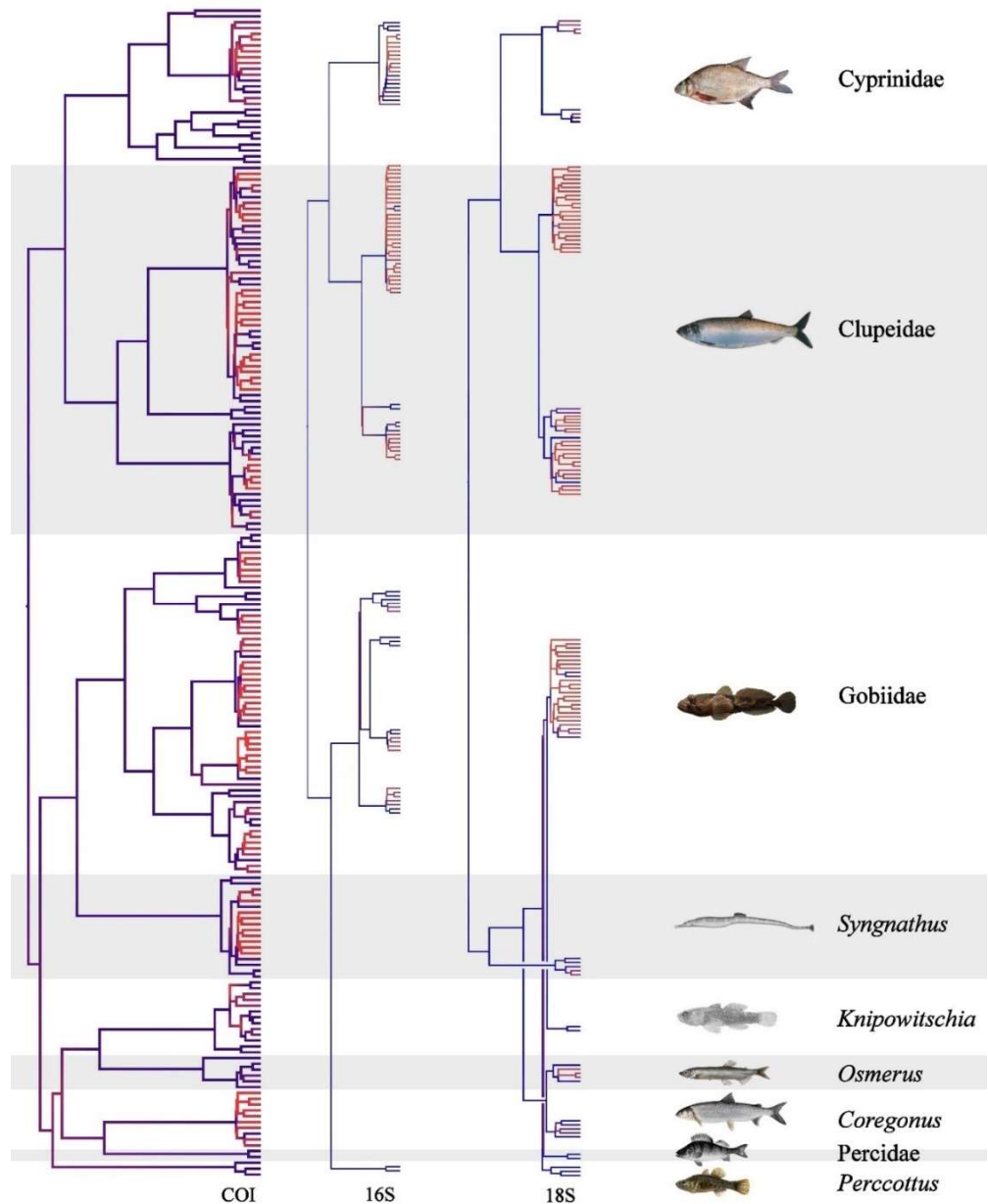
Simple and extremely cheap methods of manipulations with DNA (isolation, purification) that we propose allow to use more widely the possibilities of DNA analysis in the practice of routine hydrobiological researches, even in cases of a limited funding. Our experience has shown that even the use of the simplest reagents provides a sufficient level of efficiency and quality for specific DNA identification in fishes. Proposed research design allows us to estimate the full cost of a single sequence (with all associated costs from DNA extraction to obtaining the sequencing results) at about 2 USD (or about 3 USD for sequenced in both directions), which is comparable in its cost to the most modern high-throughput sequencing systems [90]. But our method does not require expensive equipment.

Undoubtedly, the 5'-region of the COI gene is a very effective marker for a DNA-based species barcoding system. However, the significant variability in the flanking regions of the "complete" fragment often requires the combination of several primer pairs or several cycles of PCR to obtain a final product. The transition from "universal" sets of primers to individual ones, more suitable for a relatively small number of invasive species, is a logical step in the monitoring of biological diversity.

The issue of a positive control when working with a degraded matrix (as the DNA can be damaged due to poor preservation, imperfect storage and transportation of samples) is a specific point, and the 16S locus is proposed for using in such cases [13]. However, as our study revealed, it frequently gives some false positive results. It is characteristic not only of fish, but also of human DNA, the contamination of which cannot be fully avoided. Therefore, it is more easy to reconcile the issue of positive control, for example, by routine studies on "short" COI sequences synthesized from the ifCOI<sub>mb</sub> primer pair and the 18S locus. Sequencing of the obtained products allows us to determine accurately the species of fish. Also, due to the requirement for a shorter matrix, the rate of successful sequences even exceeds that for 16S (Table 3). It may also be noted that the use of the proposed primers may be successful in ifCOI<sub>mb</sub> high-throughput sequencing systems for the purpose of DNA metabarcoding [91], and also to detect compliance with food products quality [92].

At the same time, for accurate identification of "short" COI sequences, we need to create a library of the reference sequences from precisely defined voucher types. It is advisable to carry out this effort for each region under consideration. Fortunately, a huge number of sequences has now been accumulated for many freshwater invasive species in European water bodies, and the missing data may become available when working with the products obtained from the MifCOI primers. For them, a good amplification success is provided by a high degeneracy and the presence of inosine in the 3'-region, which greatly increases the efficiency of hybridization of the primer with the matrix [93]. Sequencing problems for such primers are solved by using M13-tails, which makes it possible to neutralize an influence of the primer dimers, degeneracy, and the use of

non-canonical bases. In this case, it is possible to work without a single read from the forward or reverse primer, which significantly reduces the cost of sequencing. We should also note the success of the if18S primers. In our opinion, the 18S locus is strongly underestimated as a marker of species identification in fish. Sufficient variability in the V1-V2 loops and low variability in conservative domains make it possible to obtain a high-quality PCR product that can be used as a positive control, and also with aim to confirm the species identification by COI, as well as to search for possible hybrids between invaders and native taxa. The only limitation of the widespread use of this nuclear marker is still the low representation in the international databases of nucleotide sequences.



**Figure 5.** Topology of the BI genetic tree of COI, 16S and 18S loci.

*Application of DNA diagnostics for the detection of alien fish species in Central Russia (Volga-Kama region)*

Unfortunately, the objective difficulties in the identification of many fishes are complicated by still incessant phase of “splitting” in taxonomy resulted in separation of more and more new species from traditional subspecies, forms and races of the animals, creating problems for routine users of the taxonomy [78]. The lack of clear criteria for hiatuses between species has led to an exaggerated significance of the authority of the taxonomist, which can be expressed by words of M. Kotellat “... a number of species accounts include my best guess or educated guess of what is likely to appear once the discussed problems can be clarified” [94]. In contrast to this subjective approach, some attempt has been made to introduce some adequate mathematical criteria for identifying species based on the genetic data [81], which has been developed up to the purely mathematical concept of the “rank-free” PhyloCode system [95]. In the present paper, we do not want to touch upon such basic concepts, but only propose to be guided by the criterion of the sufficient need to apply certain approaches to solve specific problems of DNA identification in the case of biological invasions.

The majority of modern genetic methods for species differentiation operate by fitting some mathematical model of historical diversification to data collected from some natural systems. Nevertheless, the parameter space related to the differentiation of species is large, and all modern approaches suggest a number of certain assumptions and simplifications for each method. It should always be remembered that any mathematical models are imperfect imitations of a biological reality, which is best reflected by Box's aphorism “all models are wrong” [96]. Undoubtedly, an increase in the number of analyzed loci, the length of sequences, and the use of different approaches and techniques in the delimitation of species, should increase the reliability of conclusions about the taxonomic diversity of the sample [97]. Using the available data sets, we tried to analyse different loci with different levels of phylogenetic signal (loci of the mitochondrial and nuclear genomes, protein coding and ribosomal sequences) using both remote and stochastic methods, including those based on coalescence models. In our opinion, only a comprehensive integrated approach based on the genetic taxonomy and a “traditional” approach with morphological definition and construction of dichotomous identification keys may guarantee at least some objective assessment of the species richness [98]. In general, in the present, we adhered to a “cautious” taxonomic strategy in interpreting genosystematics data and follow the principle “in most contexts it is better to fail to delimit species than it is to falsely delimit entities that do not represent actual evolutionary lineages” [99].

Currently, 11 alien fish species are found in mass in the Volga-Kama region (the largest in Europe), and another ten are represented by single findings or local populations [17]. All common species are studied here, however, different efficiency of DNA identification may be noted for different taxa. Families Clupeidae and Cyprinidae demonstrate a high detection efficiency when using the genetic data. The former is represented by the most widespread and numerous alien species: common kilka, *Clupeonella cultriventris* (Nordmann, 1840). For comparison to this species, we used the sequences of a related species, the European sprat *Sprattus sprattus* (Linnaeus 1758). These two morphologically similar species are successfully separated based on all loci. But in the general population of *C. cultriventris*, some genetic differentiation is also observed [100].

A special task is the analysis of samples from the commercial networks. For instance, samples d015a, d015b of the European sprat, acquired through a large Russian commercial network with a governmental quality certificate in 2010, belong to the kilka according to our genetic identification. They clearly differed from the sequences of European sprat both from the samples studied by us and from sequences presented in international databases (Figure 2). Probably, the seller was misled by the supplier, who provided a less expensive kilka instead of a more expensive sprat. Such facts are common [92].

Another example of the successful application of DNA identification is the barcoding of the cyprinids (family Cyprinidae). The most widespread alien species in Europe from this family is the topmouth gudgeon, *Pseudorasbora parva* (Temminck et Schlegel 1846) [101]. Although this species has not yet been found in the Volga-Kama region, it is

common in the neighboring basin of the Don River, another large river in Europe [102]. Some samples received from our colleagues from scientific and environmental organizations of Russia were incorrectly identified and belonged to other species; this misidentification was revealed only by the results of DNA analysis (Figure 2). This is associated with the problem of identification of the cyprinid fry, where the morphometric characters of several species overlap strongly. An example of this is the almost complete coincidence of identification keys [103] in fry of the topmouth gudgeon and bleak *Leucaspis delineatus* (Heckel, 1843): the differences between them may only be traced analysing the pharyngeal teeth. We may also note the complexity of the taxonomic status of different species and groups of populations in gudgeons of the genus *Pseudorasbora* Bleeker, 1860 due to a high level of polymorphism and existence of several phylogenetic lineages [104], which significantly complicates the monitoring of the invasive process.

Other examples of the misidentifications are related to the shipment of a tissue fragment for DNA analysis (e. g., the fins preserved in ethanol). In this case, even having a high-quality photograph of the voucher, it is not possible to verify the primary identification of the species. Similar difficulties arise in the ichthyological studies in the lower reaches of large rivers, where freshwater, estuarine, and anadromous faunas are mixed. In particular, according to the results of DNA analysis, most juveniles of herring in the lower reaches of the Volga and Don are formed both by tulka and the representatives of the Caspian-Black Sea herring of the genus *Alosa* Linck, 1790. Thus, with traditional monitoring of biological invasions, some of the invaders may escape the attention of researchers. Also "the invader detection" could be a result of a false positive definition of a possibly invasive taxon. To mitigate such cases, it is proposed to introduce methods of DNA identification of alien species more widely [105].

Representatives of family Gobiidae are among the most numerous alien species in Europe [106]. Taxonomy of the gobiids is extremely complicated and the validity of a number of taxa requires additional studies [107]. However, if we ignore any phylogenetic reconstruction, DNA identification for most gobiids gives an unambiguous identification of the species [108]. DNA barcoding confirmed the findings of the syrman goby, *Ponticola syrman* (Nordmann, 1840) in the Astrakhan Region. Its distribution was previously regarded as limited to estuarine zones of the rivers [109]. In addition, in the upstream reaches of the undammed part of the Volga River near Volgograd, the long-tailed longtail dwarf goby *Knipowitschia longicaudata* (Kessler, 1877) was found [110]. Most likely, a donor region for this species entering the Volga basin is the Don River basin, and the Volga-Don navigable canal served as a transit corridor for this species. Another "invisible" species of the invading gobiids is the stellate tadpole goby, *Benthophilus stellatus* (Sauvage, 1874), a representative of a species group with inadequate taxonomy and the validity of a number of species questionable. The Kuibyshev Reservoir, where the Caspian and Azov-Black Sea phylogenetic lineages are mixed [111], serves as a secondary spread center of stellate tadpole goby. DNA identification of other gobiids is usually straightforward: for round goby, *Neogobius melanostomus* (Pallas, 1814), monkey goby, *N. fluviatilis* (Pallas, 1814), and Caspian bighead goby *N.* (= *Ponticola*) *gorlap* (Iljin, 1949), mOTUs coincide with "traditional" species. For those living in the Volga basin, i.e. tubenose goby g. s *Proterorhinus* Smitt, 1900, the only DNA analysis may adequately identify the species. It was previously believed that the Volga basin is inhabited by tubenose goby *P. semilunaris* (Heckel, 1837) [112]. However, a comparison of the sequences from the Volga with the reference ones from the Ponto-Caspian tubenose gobies [113] showed that the Volga populations are represented by *P. semipellucidus* (Kessler, 1877), which significantly differs from *P. semilunaris* of the Black Sea origin.

The family of freshwater sleepers, Odontobutidae Hoese et Gill, 1993, close to gobiid fishes, is represented in Europe by only a single species - Chinese (mud) sleeper, *Perccottus glenii* Dybowski, 1877. Genetic identification of this invader is simple, however, larvae of the mud sleeper may be confused with juvenile percids.



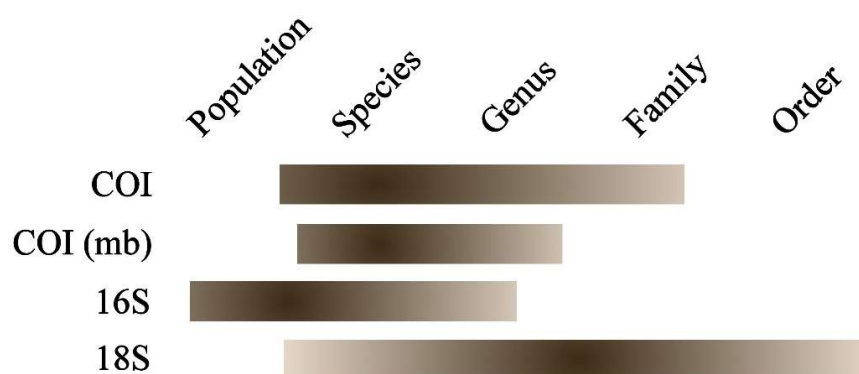
The identification of a sole representative of the pipefishes from the family Syngnathidae Rafinesque, 1810, namely black-striped pipefish *Syngnathus abaster* Risso, 1827, also does not cause any problems. Earlier, some genetically distinct groups of the pipefish populations of the Caspian and Black Seas were revealed [114]. DNA barcoding shows significant differences between the marine pipefish populations from the Caspian Sea and freshwater Eastern European populations, which possibly reflects their micro-phylogenesis and adaptation to the life in fresh waters (as it was shown for kilka, see above).

Finally, the last two alien fish species in the Volga basin belong to the “northern” invasive stream: salmonids of family Salmonidae Rafinesque, 1815. DNA identification of the European smelt *Osmerus eperlanus* (Linnaeus, 1758) does not cause problems. This species is reliably identified by both mtDNA COI and nuclDNA 18S. However, the DNA identification of another representative of the salmonids, vendace *Coregonus albula* (Linnaeus 1758), is difficult in frames of the “traditional” barcoding. For the European vendace from the Volga River basin, the haplotypes similar to those in four different species of the coregonids were found (Figure 2). This fact may be explained by an extremely low genetic variability at the COI locus, which is insufficient for an adequate discrimination of the species. It is also possible that the entire complex is represented by a single polymorphic species [115]. In addition, it could be a case of a relict preservation of mtDNA after numerous attempts to intentionally introduce different species and forms of the coregonids and their hybrids on the Upper Volga [116].

It should be emphasized that reconstructed phylogenies based on mtDNA at a level higher than the genus contradict the modern concepts of fish phylogeny [117], which is more consistent with phylogeny based on the 18S locus (Figure 5). A similar effect is well known for other gene trees [118]. It could be explained by a strongly varying rate of accumulation of nucleotide substitutions during the evolution of various genes, and may be accompanied by a long branch attraction [119]. But for utilitarian purposes of species identification based on the COI locus, the correctness of the reconstructed phylogeny of higher taxa is irrelevant. If required, data on some more conservative nuclear loci should be used for this purpose [120].

## 5. Conclusions

Our proposed sets of primers demonstrate a high efficiency in amplification and a high specificity for the DNA barcode region both for freshwater alien species and other fish taxa. We propose (Figure 6) to use the MifCOI primer set to accumulate sequence data for reference samples of invasive fishes from different locations. The sequences obtained from the if16S primers can be relevant in the study of local populations and for the accumulation of data for comparative analysis in the study of communities during metabarcoding by ribosomal mtDNA. For a mass routine analysis, we propose to use the “short” COI sequences obtained from the ifCOI<sub>mb</sub> primers. Such “mini”-barcodes have a number of advantages, namely: a higher rate of successful amplifications in comparison with a “long” fragment – a higher level of specificity and ensure the reliability of identification of both invasive and native freshwater fish species. The use of the if18S primer may be used simultaneously as a positive control of PCR success, as well as to verify the results of the barcoding by mtDNA. Reconstructed 18S gene tree may be used as a guide-tree to determine a general topology of the phylogenetic tree for higher taxa.



**Figure 6.** The approximate success of using various genetic loci for DNA identification of alien species of freshwater fish. The degree of efficiency is proportional to the gradient of the fill.

**Author Contributions:** Conceptualization, D.P.K. and A.A.K.; methodology, D.P.K.; software, D.P.K.; validation, D.P.K., E.I.B., E.A.B. and A.A.K.; formal analysis, D.P.K.; investigation, D.P.K., E.I.B. and A.A.K.; resources, D.P.K. and D.D.P.; data curation, D.P.K. and Yu.V.K.; writing—original draft preparation, D.P.K., D.D.P. and A.A.K.; writing—review and editing, D.P.K. and A.A.K.; visualization, D.P.K. and Yu.V.K.; supervision, A.A.K.; project administration, Yu.V.K.; funding acquisition, E.I.B. All authors have read and agreed to the published version of the manuscript.

**Funding:** Genetic research was supported by the Russian Foundation for Basic Research grant no. 20-34-70020. The sampling of the materials was carried out as part of a State Assignment no. 121051100104-6.

**Data Availability Statement:** All material examined in this study are openly available at the facilities listed, and by the catalogue numbers in the Materials and Methods section above. Data available at Open Science Framework project: Karabanov, D.P, 2021. OSF. Dataset. <https://osf.io/b8qfd/>.

**Acknowledgments:** This study would have been impossible without a large number of colleagues from scientific and environmental organizations who helped in organizing field studies and sampling. We are unable to list everyone here by name, so as not to miss and not offend anyone. Separately, we would like to thank Dr. V. S. Artamonova and Dr. A. A. Makhrov for help and advice on DNA analysis, as well as Prof. Yu. Yu. Dgebuadze for general guidance on the part of the study related to biological invasions. Many thanks to Dr. R. J. Shiel for linguistic edition of earlier draft.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

## References

1. Elton, C.S. *The ecology of invasions by animals and plants: New edition*; University of Chicago Press: Chicago, London, 2000, ISBN 0226206386.
2. Smirnov, A.K.; Pavlov, D.D.; Kodukhova, Y.V.; Karabanov, D.P. Impact of Zebra mussel *Dreissena polymorpha* Pallas, 1771 (bivalvia) appearance on fish populations in Lake Pleshcheevo, European Russia. *Zool. Zh.* **2020**, *99*, 1363–1374, doi:10.31857/S0044513420110070.
3. Hata, H.; Uemura, Y.; Ouchi, K.; Matsuba, H. Hybridization between an endangered freshwater fish and an introduced congeneric species and consequent genetic introgression. *PLoS ONE* **2019**, *14*, e0212452, doi:10.1371/journal.pone.0212452.
4. Boussellaa, W.; Neifar, L.; Goedknecht, M.A.; Thieltges, D.W. Lessepsian migration and parasitism: richness, prevalence and intensity of parasites in the invasive fish *Sphyrna chrysotaenia* compared to its native congener *Sphyrna sphyraena* in Tunisian coastal waters. *PeerJ* **2018**, *6*, e5558, doi:10.7717/peerj.5558.
5. Karabanov, D.P.; Bekker, E.I.; Kotov, A.A. Underestimated consequences of biological invasions in phylogeographic reconstructions as seen in *Daphnia magna* (Crustacea, Cladocera). *Zool. Zh.* **2020**, *99*, 1232–1241, doi:10.31857/S0044513420110033.

6. Robertson, P.A.; Mill, A.; Novoa, A.; Jeschke, J.M.; Essl, F.; Gallardo, B.; Geist, J.; Jarić, I.; Lambin, X.; Musseau, C.; et al. A proposed unified framework to describe the management of biological invasions. *Biol. Invasions* **2020**, *22*, 2633–2645, doi:10.1007/s10530-020-02298-2.
7. Makhrov, A.A.; Karabanov, D.P.; Koduhova, Y.V. Genetic methods for the control of alien species. *Russ. J. Biol. Invasions* **2014**, *5*, 194–202, doi:10.1134/S2075111714030096.
8. Hebert, P.D.N.; Cywinska, A.; Ball, S.L.; deWaard, J.R. Biological identifications through DNA barcodes. *Proc. R. Soc. Lond., B, Biol. Sci.* **2003**, *270*, 313–321, doi:10.1098/rspb.2002.2218.
9. Hebert, P.D.N.; Ratnasingham, S.; deWaard, J.R. Barcoding animal life: Cytochrome *c* oxidase subunit 1 divergences among closely related species. *Proc. R. Soc. Lond., B, Biol. Sci.* **2003**, *270* Suppl 1, S96–9, doi:10.1098/rsbl.2003.0025.
10. Folmer, O.; Black, M.; Hoeh, W.; Lutz, R.; Vrijenhoek, R. DNA primers for amplification of mitochondrial cytochrome *c* oxidase subunit I from diverse metazoan invertebrates. *Mol. Marine Biol. Biotechnol.* **1994**, *3*, 294–299.
11. Sharma, P.; Kobayashi, T. Are “universal” DNA primers really universal? *J. Appl. Genet.* **2014**, *55*, 485–496, doi:10.1007/s13353-014-0218-9.
12. Ward, R.D.; Zemlak, T.S.; Innes, B.H.; Last, P.R.; Hebert, P.D.N. DNA barcoding Australia’s fish species. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* **2005**, *360*, 1847–1857, doi:10.1098/rstb.2005.1716.
13. Ivanova, N.V.; Zemlak, T.S.; Hanner, R.H.; Hebert, P.D.N. Universal primer cocktails for fish DNA barcoding. *Mol. Ecol. Notes* **2007**, *7*, 544–548, doi:10.1111/j.1471-8286.2007.01748.x.
14. Leray, M.; Yang, J.Y.; Meyer, C.P.; Mills, S.C.; Agudelo, N.; Ranwez, V.; Boehm, J.T.; Machida, R.J. A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: Application for characterizing coral reef fish gut contents. *Front. Zool.* **2013**, *10*, 34, doi:10.1186/1742-9994-10-34.
15. Cawthorn, D.-M.; Steinman, H.A.; Witthuhn, R.C. Evaluation of the 16S and 12S rRNA genes as universal markers for the identification of commercial fish species in South Africa. *Gene* **2012**, *491*, 40–48, doi:10.1016/j.gene.2011.09.009.
16. Tang, C.Q.; Leasi, F.; Obertegger, U.; Kieneke, A.; Barraclough, T.G.; Fontaneto, D. The widely used small subunit 18S rDNA molecule greatly underestimates true diversity in biodiversity surveys of the meiofauna. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 16208–16212, doi:10.1073/pnas.1209160109.
17. Karabanov, D.P.; Pavlov, D.D.; Bazarov, M.I.; Borovikova, E.A.; Gerasimov, Y.V.; Kodukhova, Y.V.; Smirnov, A.K.; Stolbunov, I.A. Alien species of fish in the littoral of Volga and Kama reservoirs (Results of complex expeditions of IBIW RAS in 2005–2017). *Trans. IBIW RAS* **2018**, *82(85)*, 67–80, doi:10.24411/0320-3557-2018-1-0013.
18. Douglas, A.M.; Georgalis, A.M.; Benton, L.R.; Canavan, K.L.; Atchison, B.A. Purification of human leucocyte DNA: Proteinase K is not necessary. *Anal. Biochem.* **1992**, *201*, 362–365, doi:10.1016/0003-2697(92)90352-8.
19. Katoh, K.; Standley, D.M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **2013**, *30*, 772–780, doi:10.1093/molbev/mst010.
20. Okonechnikov, K.; Golosova, O.; Fursov, M. Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics* **2012**, *28*, 1166–1167, doi:10.1093/bioinformatics/bts091.
21. Ward, R.D.; Hanner, R.; Hebert, P.D.N. The campaign to DNA barcode all fishes, FISH-BOL. *J. Fish Biol.* **2009**, *74*, 329–356, doi:10.1111/j.1095-8649.2008.02080.x.
22. Santos, A.; van Aerle, R.; Barrientos, L.; Martinez-Urtaza, J. Computational methods for 16S metabarcoding studies using Nanopore sequencing data. *Comput. Struct. Biotechnol. J.* **2020**, *18*, 296–305, doi:10.1016/j.csbj.2020.01.005.
23. Hadziavdic, K.; Lekang, K.; Lanzen, A.; Jonassen, I.; Thompson, E.M.; Troedsson, C. Characterization of the 18S rRNA gene for designing universal eukaryote specific primers. *PLoS ONE* **2014**, *9*, e87624, doi:10.1371/journal.pone.0087624.
24. Messing, J. New M13 vectors for cloning. *Methods Enzymol.* **1983**, *101*, 20–78, doi:10.1016/0076-6879(83)01005-8.
25. Jennings, W.B.; Ruschi, P.A.; Ferraro, G.; Quijada, C.C.; Silva-Malanski, A.C.G.; Prosdocimi, F.; Buckup, P.A. Barcoding the Neotropical freshwater fish fauna using a new pair of universal COI primers with a discussion of primer dimers and M13 primer tails. *Genome* **2019**, *62*, 77–83, doi:10.1139/gen-2018-0145.
26. Green, M.R.; Sambrook, J. Touchdown Polymerase Chain Reaction (PCR). *Cold Spring Harb. Protoc.* **2018**, *2018*, doi:10.1101/pdb.prot095133.
27. Makhrov, A.A.; Artamonova, V.S.; Karabanov, D.P. Finding of topmouth gudgeon *Pseudorasbora parva* (Temminck et Schlegel) (Actinopterygii: Cyprinidae) in the Brahmaputra River basin (Tibetan Plateau, China). *Russ. J. Biol. Invasions* **2013**, *4*, 174–179, doi:10.1134/S2075111713030089.
28. Boratyn, G.M.; Camacho, C.; Cooper, P.S.; Coulouris, G.; Fong, A.; Ma, N.; Madden, T.L.; Matten, W.T.; McGinnis, S.D.; Merezhuk, Y.; et al. BLAST: a more efficient report with usability improvements. *Nucleic Acids Res.* **2013**, *41*, W29–33, doi:10.1093/nar/gkt282.
29. Katoh, K.; Rozewicki, J.; Yamada, K.D. MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization. *Brief. Bioinformatics* **2019**, *20*, 1160–1166, doi:10.1093/bib/bbx108.
30. Nei, M.; Kumar, S. *Molecular evolution and phylogenetics*; Oxford University Press: New York, 2000, ISBN 0195135857.
31. Rozas, J.; Ferrer-Mata, A.; Sanchez-DelBarrio, J.C.; Guirao-Rico, S.; Librado, P.; Ramos-Onsins, S.E.; Sanchez-Gracia, A. DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol. Biol. Evol.* **2017**, *34*, 3299–3302, doi:10.1093/molbev/msx248.
32. Fu, Y.X. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **1997**, *147*, 915–925.
33. Tajima, F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **1989**, *123*, 585–595.

34. Ramirez-Soriano, A.; Ramos-Onsins, S.E.; Rozas, J.; Calafell, F.; Navarro, A. Statistical power analysis of neutrality tests under demographic expansions, contractions and bottlenecks with recombination. *Genetics* **2008**, *179*, 555–567, doi:10.1534/genetics.107.083006.
35. Garrigan, D.; Lewontin, R.; Wakeley, J. Measuring the sensitivity of single-locus “neutrality tests” using a direct perturbation approach. *Mol. Biol. Evol.* **2010**, *27*, 73–89, doi:10.1093/molbev/msp209.
36. Kalyaanamoorthy, S.; Minh, B.Q.; Wong, T.K.F.; Haeseler, A. von; Jermiin, L.S. ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat. Methods* **2017**, *14*, 587–589, doi:10.1038/nmeth.4285.
37. Trifinopoulos, J.; Nguyen, L.-T.; Haeseler, A. von; Minh, B.Q. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res.* **2016**, *44*, W232–5, doi:10.1093/nar/gkw256.
38. Schwarz, G. Estimating the dimension of a model. *Ann. Stat.* **1978**, *6*, 461–464, doi:10.1214/aos/1176344136.
39. Hurvich, C.M.; Tsai, C.-L. Regression and time series model selection in small samples. *Biometrika* **1989**, *76*, 297–307, doi:10.1093/biomet/76.2.297.
40. Minh, B.Q.; Schmidt, H.A.; Chernomor, O.; Schrempf, D.; Woodhams, M.D.; Haeseler, A. von; Lanfear, R. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **2020**, *37*, 1530–1534, doi:10.1093/molbev/msaa015.
41. Hoang, D.T.; Chernomor, O.; Haeseler, A. von; Minh, B.Q.; Le Vinh, S. UFBoot2: Improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* **2018**, *35*, 518–522, doi:10.1093/molbev/msx281.
42. Guindon, S.; Dufayard, J.-F.; Lefort, V.; Anisimova, M.; Hordijk, W.; Gascuel, O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **2010**, *59*, 307–321, doi:10.1093/sysbio/syq010.
43. Bouckaert, R.; Vaughan, T.G.; Barido-Sottani, J.; Duchene, S.; Fourment, M.; Gavryushkina, A.; Heled, J.; Jones, G.; Kuhnert, D.; Maio, N. de; et al. BEAST 2.5 : An advanced software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* **2019**, *15*, e1006650, doi:10.1371/journal.pcbi.1006650.
44. Drummond, A.J.; Suchard, M.A.; Xie, D.; Rambaut, A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* **2012**, *29*, 1969–1973, doi:10.1093/molbev/mss075.
45. Kumar, S.; Stecher, G.; Li, M.; Nknyaz, C.; Tamura, K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol. Biol. Evol.* **2018**, *35*, 1547–1549, doi:10.1093/molbev/msy096.
46. Gernhard, T. The conditioned reconstructed process. *J. Theor. Biol.* **2008**, *253*, 769–778, doi:10.1016/j.jtbi.2008.04.005.
47. Rambaut, A.; Drummond, A.J.; Xie, D.; Baele, G.; Suchard, M.A. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Syst. Biol.* **2018**, *67*, 901–904, doi:10.1093/sysbio/syy032.
48. Schoch, C.L.; Ciufo, S.; Domrachev, M.; Hotton, C.L.; Kannan, S.; Khovanskaya, R.; Leipe, D.; Mcveigh, R.; O’Neill, K.; Robertse, B.; et al. NCBI Taxonomy: a comprehensive update on curation, resources and tools. *Database* **2020**, *2020*, baaa062, doi:10.1093/database/baaa062.
49. Ratnasingham, S.; Hebert, P.D.N. A DNA-based registry for all animal species: The barcode index number (BIN) system. *PLoS ONE* **2013**, *8*, e66213, doi:10.1371/journal.pone.0066213.
50. Puillandre, N.; Brouillet, S.; Achaz, G. ASAP: assemble species by automatic partitioning. *Mol. Ecol. Resour.* **2021**, *21*, 609–620, doi:10.1111/1755-0998.13281.
51. Collins, R.A.; Boykin, L.M.; Cruickshank, R.H.; Armstrong, K.F. Barcoding’s next top model: an evaluation of nucleotide substitution models for specimen identification. *Methods Ecol. Evol.* **2012**, *3*, 457–465, doi:10.1111/j.2041-210X.2011.00176.x.
52. Brown, S.D.J.; Collins, R.A.; Boyer, S.; Lefort, M.-C.; Malumbres-Olarte, J.; Vink, C.J.; Cruickshank, R.H. Spider: an R package for the analysis of species identity and evolution, with particular reference to DNA barcoding. *Mol. Ecol. Resour.* **2012**, *12*, 562–565, doi:10.1111/j.1755-0998.2011.03108.x.
53. Rangel-Medrano, J.D.; Ortega-Lara, A.; Marquez, E.J. Ancient genetic divergence in bumblebee catfish of the genus *Pseudopimelodus* (Pseudopimelodidae: Siluriformes) from northwestern South America. *PeerJ* **2020**, *8*, e9028, doi:10.7717/peerj.9028.
54. Ota, R.P.; Machado, V.N.; Andrade, M.C.; Collins, R.A.; Farias, I.P.; Hrbek, T. Integrative taxonomy reveals a new species of pacu (Characiformes: Serrasalminae: *Myloplus*) from the Brazilian Amazon. *Neotrop. ichthyol.* **2020**, *18*, doi:10.1590/1982-0224-20190112.
55. Pons, J.; Barraclough, T.G.; Gomez-Zurita, J.; Cardoso, A.; Duran, D.P.; Hazell, S.; Kamoun, S.; Sumlin, W.D.; Vogler, A.P. Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Syst. Biol.* **2006**, *55*, 595–609, doi:10.1080/10635150600852011.
56. Fujisawa, T.; Barraclough, T.G. Delimiting species using single-locus data and the Generalized Mixed Yule Coalescent approach: a revised method and evaluation on simulated data sets. *Syst. Biol.* **2013**, *62*, 707–724, doi:10.1093/sysbio/syt033.
57. Zhang, J.; Kapli, P.; Pavlidis, P.; Stamatakis, A. A general species delimitation method with applications to phylogenetic placements. *Bioinformatics* **2013**, *29*, 2869–2876, doi:10.1093/bioinformatics/btt499.
58. Foster, E.D.; Deardorff, A. Open Science Framework (OSF). *J. Med. Libr. Assoc.* **2017**, *105*, 203–206, doi:10.5195/JMLA.2017.88.
59. Crossley, B.M.; Bai, J.; Glaser, A.; Maes, R.; Porter, E.; Killian, M.L.; Clement, T.; Toohey-Kurth, K. Guidelines for Sanger sequencing and molecular assay monitoring. *J. Vet. Diagn. Invest.* **2020**, *32*, 767–775, doi:10.1177/1040638720905833.
60. Dieffenbach, C.W.; Lowe, T.M.; Dveksler, G.S. General concepts for PCR primer design. *PCR Methods Appl.* **1993**, *3*, S30–7, doi:10.1101/gr.3.3.s30.



61. Karabanov, D.P.; Bekker, E.I.; Shiel, R.J.; Kotov, A.A. Invasion of a Holarctic planktonic cladoceran *Daphnia galeata* Sars (Crustacea: Cladocera) in the Lower Lakes of South Australia. *Zootaxa* **2018**, *4402*, 136–148, doi:10.11646/zootaxa.4402.1.6.
62. Longo, M.S.; O'Neill, M.J.; O'Neill, R.J. Abundant human DNA contamination identified in non-primate genome databases. *PLoS ONE* **2011**, *6*, e16410, doi:10.1371/journal.pone.0016410.
63. Ciufu, S.; Kannan, S.; Sharma, S.; Badretdin, A.; Clark, K.; Turner, S.; Brover, S.; Schoch, C.L.; Kimchi, A.; DiCuccio, M. Using average nucleotide identity to improve taxonomic assignments in prokaryotic genomes at the NCBI. *Int. J. Syst. Evol. Microbiol.* **2018**, *68*, 2386–2392, doi:10.1099/ijsem.0.002809.
64. Palandacic, A.; Naseka, A.; Ramler, D.; Ahnelt, H. Contrasting morphology with molecular data: an approach to revision of species complexes based on the example of European *Phoxinus* (Cyprinidae). *BMC Evol. Biol.* **2017**, *17*, 184, doi:10.1186/s12862-017-1032-x.
65. Andujar, C.; Arribas, P.; Yu, D.W.; Vogler, A.P.; Emerson, B.C. Why the COI barcode should be the community DNA metabarcode for the metazoa. *Mol. Ecol.* **2018**, *27*, 3968–3975, doi:10.1111/mec.14844.
66. Piper, A.M.; Batovska, J.; Cogan, N.O.I.; Weiss, J.; Cunningham, J.P.; Rodoni, B.C.; Blacket, M.J. Prospects and challenges of implementing DNA metabarcoding for high-throughput insect surveillance. *Gigascience* **2019**, *8*, giz092, doi:10.1093/gigascience/giz092.
67. Walker, S.P.; Barrett, M.; Hogan, G.; Flores Bueso, Y.; Claesson, M.J.; Tangney, M. Non-specific amplification of human DNA is a major challenge for 16S rRNA gene sequence analysis. *Sci. Rep.* **2020**, *10*, 16356, doi:10.1038/s41598-020-73403-7.
68. Garibian, P.G.; Karabanov, D.P.; Neretina, A.N.; Taylor, D.J.; Kotov, A.A. *Bosminopsis deitersi* (Crustacea: Cladocera) as an ancient species group: a revision. *PeerJ* **2021**, *9*, e11310, doi:10.7717/peerj.11310.
69. Xu, S.; Hebert, P.D.N.; Kotov, A.A.; Cristescu, M.E. The noncosmopolitanism paradigm of freshwater zooplankton: Insights from the global phylogeography of the predatory cladoceran *Polyphemus pediculus* (Linnaeus, 1761) (Crustacea, Onychopoda). *Mol. Ecol.* **2009**, *18*, 5161–5179, doi:10.1111/j.1365-294X.2009.04422.x.
70. Rubinoff, D.; Holland, B.S. Between two extremes: mitochondrial DNA is neither the panacea nor the nemesis of phylogenetic and taxonomic inference. *Syst. Biol.* **2005**, *54*, 952–961, doi:10.1080/10635150500234674.
71. Wise, C.A.; Sraml, M.; Rubinsztein, D.C.; Easteal, S. Comparative nuclear and mitochondrial genome diversity in humans and chimpanzees. *Mol. Biol. Evol.* **1997**, *14*, 707–716, doi:10.1093/oxfordjournals.molbev.a025810.
72. Perlwitz, M.D.; Burks, C.; Waterman, M.S. Pattern analysis of the genetic code. *Adv. Appl. Math.* **1988**, *9*, 7–21, doi:10.1016/0196-8858(88)90003-6.
73. Crease, T.J.; Taylor, D.J. The origin and evolution of variable-region helices in V4 and V7 of the small-subunit ribosomal RNA of branchiopod crustaceans. *Mol. Biol. Evol.* **1998**, *15*, 1430–1446, doi:10.1093/oxfordjournals.molbev.a025871.
74. Felsenstein, J. Evolutionary trees from DNA sequences: a maximum likelihood approach. *J. Mol. Evol.* **1981**, *17*, 368–376.
75. Yang, Z. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: Approximate methods. *J. Mol. Evol.* **1994**, *39*, 306–314, doi:10.1007/BF00160154.
76. Soubrier, J.; Steel, M.; Lee, M.S.Y.; Der Sarkissian, C.; Guindon, S.; Ho, S.Y.W.; Cooper, A. The influence of rate heterogeneity among sites on the time dependence of molecular rates. *Mol. Biol. Evol.* **2012**, *29*, 3345–3358, doi:10.1093/molbev/mss140.
77. Pentinsaari, M.; Ratnasingham, S.; Miller, S.E.; Hebert, P.D.N. BOLD and GenBank revisited - Do identification errors arise in the lab or in the sequence libraries? *PLoS One* **2020**, *15*, e0231814, doi:10.1371/journal.pone.0231814.
78. Mina, M.V.; Reshetnikov, Y.S.; Dgebuadze, Y.Y. Taxonomic novelties and problems for users. *J. Ichthyol.* **2006**, *46*, 476–480, doi:10.1134/S0032945206060099.
79. Clark, K.; Karsch-Mizrachi, I.; Lipman, D.J.; Ostell, J.; Sayers, E.W. GenBank. *Nucleic Acids Res.* **2016**, *44*, D67-72, doi:10.1093/nar/gkv1276.
80. Ratnasingham, S.; Hebert, P.D.N. BOLD: the Barcode of Life Data System (<http://www.barcodinglife.org>). *Mol. Ecol. Notes* **2007**, *7*, 355–364, doi:10.1111/j.1471-8286.2007.01678.x.
81. Blaxter, M.; Mann, J.; Chapman, T.; Thomas, F.; Whitton, C.; Floyd, R.; Abebe, E. Defining operational taxonomic units using DNA barcode data. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* **2005**, *360*, 1935–1943, doi:10.1098/rstb.2005.1725.
82. Hebert, P.D.N.; Stoeckle, M.Y.; Zemplak, T.S.; Francis, C.M. Identification of Birds through DNA Barcodes. *PLoS Biol.* **2004**, *2*, e312, doi:10.1371/journal.pbio.0020312.
83. Pante, E.; Schoelincq, C.; Puillandre, N. From integrative taxonomy to species description: One step beyond. *Syst. Biol.* **2015**, *64*, 152–160, doi:10.1093/sysbio/syu083.
84. Vitecek, S.; Kucinic, M.; Previsic, A.; Zivic, I.; Stojanovic, K.; Keresztes, L.; Balint, M.; Hoppeler, F.; Waringer, J.; Graf, W.; et al. Integrative taxonomy by molecular species delimitation: Multi-locus data corroborate a new species of Balkan Drusinae micro-endemics. *BMC Evol. Biol.* **2017**, *17*, 129, doi:10.1186/s12862-017-0972-5.
85. Neretina, A.N.; Karabanov, D.P.; Sacherova, V.; Kotov, A.A. Unexpected mitochondrial lineage diversity within the genus *Alonella* Sars, 1862 (Crustacea: Cladocera) across the Northern Hemisphere. *PeerJ* **2021**, *9*, e10804, doi:10.7717/peerj.10804.
86. Kartavtsev, Y.P. Sequence divergence at mitochondrial genes in animals: Applicability of DNA data in genetics of speciation and molecular phylogenetics. *Mar. Genomics* **2011**, *4*, 71–81, doi:10.1016/j.margen.2011.02.002.
87. Ma, J.; Liu, J.; Shen, Y.; Fan, Z.; Yue, B.; Zhang, X. Population genetic structure and intraspecific genetic distance of *Periplaneta americana* (Blattodea: Blattellidae) based on mitochondrial and nuclear DNA markers. *Ecol. Evol.* **2019**, *9*, 12928–12939, doi:10.1002/ece3.5777.



88. Huemer, P.; Mutanen, M.; Sefc, K.M.; Hebert, P.D.N. Testing DNA barcode performance in 1000 species of European lepidoptera: Large geographic distances have small genetic impacts. *PLoS ONE* **2014**, *9*, e115774, doi:10.1371/journal.pone.0115774.
89. Bekker, E.I.; Karabanov, D.P.; Galimov, Y.R.; Haag, C.R.; Neretina, T.V.; Kotov, A.A. Phylogeography of *Daphnia magna* Straus (Crustacea: Cladocera) in Northern Eurasia: Evidence for a deep longitudinal split between mitochondrial lineages. *PLoS ONE* **2018**, *13*, e0194045, doi:10.1371/journal.pone.0194045.
90. Chang, J.J.M.; Ip, Y.C.A.; Ng, C.S.L.; Huang, D. Takeaways from Mobile DNA Barcoding with BentoLab and MinION. *Genes (Basel)* **2020**, *11*, doi:10.3390/genes11101121.
91. Miya, M.; Gotoh, R.O.; Sado, T. MiFish metabarcoding: a high-throughput approach for simultaneous detection of multiple fish species from environmental DNA and other samples. *Fish. Sci.* **2020**, *86*, 939–970, doi:10.1007/s12562-020-01461-x.
92. Sultana, S.; Ali, M.E.; Hossain, M.A.M.; Asing; Naquiah, N.; Zaidul, I.S.M. Universal mini COI barcode for the identification of fish species in processed products. *Food Res. Int.* **2018**, *105*, 19–28, doi:10.1016/j.foodres.2017.10.065.
93. Christopherson, C.; Sninsky, J.; Kwok, S. The effects of internal primer-template mismatches on RT-PCR: HIV-1 model studies. *Nucleic Acids Res.* **1997**, *25*, 654–658, doi:10.1093/nar/25.3.654.
94. Kottelat, M. *Fishes of Mongolia: A check-list of the fishes known to occur in Mongolia with comments on systematics and nomenclature*; Environment and Social Development East Asia and Pacific Region World Bank: Washington D.C., 2006.
95. Dayrat, B.; Cantino, P.D.; Clarke, J.A.; Queiroz, K. de. Species names in the PhyloCode. The approach adopted by the International Society for Phylogenetic Nomenclature. *Syst. Biol.* **2008**, *57*, 507–514, doi:10.1080/10635150802172176.
96. Box, G.E.P. Science and Statistics. *J. Am. Stat. Assoc.* **1976**, *71*, 791–799, doi:10.2307/2286841.
97. Coissac, E.; Hollingsworth, P.M.; Lavergne, S.; Taberlet, P. From barcodes to genomes: extending the concept of DNA barcoding. *Mol. Ecol.* **2016**, *25*, 1423–1428, doi:10.1111/mec.13549.
98. Kotov, A.A.; Garibian, P.G.; Bekker, E.I.; Taylor, D.J.; Karabanov, D.P. A new species group from the *Daphnia curvirostris* species complex (Cladocera: Anomopoda) from the eastern Palaearctic: taxonomy, phylogeny and phylogeography. *Zool. J. Linn. Soc.* **2021**, *191*, 772–822, doi:10.1093/zoolinnean/zlaa046.
99. Carstens, B.C.; Pelletier, T.A.; Reid, N.M.; Satler, J.D. How to fail at species delimitation. *Mol. Ecol.* **2013**, *22*, 4369–4383, doi:10.1111/mec.12413.
100. Karabanov, D.P.; Kodukhova, Y.V. Biochemical polymorphism and intraspecific structure in populations of Kilka *Clupeonella cultriventris* (Nordmann, 1840) from natural and invasive parts of its range. *Inland Water Biol.* **2018**, *11*, 496–500, doi:10.1134/S1995082918040107.
101. Karabanov, D.P.; Kodukhova, Y.V.; Kutsokon, Y.K. Expansion of stone moroko *Pseudorasbora parva* (Cypriniformes, Cyprinidae) into Eurasian reservoirs. *Vestn. Zool.* **2010**, *44*, 115–124.
102. Karabanov, D.P.; Kodukhova, Y.V.; Slyn'ko, Y.V. New finds of topmouth gudgeon *Pseudorasbora parva* (Temm. et Schl., 1846) in the European part of Russia. *Russ. J. Biol. Invasions* **2010**, *1*, 156–158, doi:10.1134/S2075111710030021.
103. Kottelat, M.; Freyhof, J. *Handbook of European freshwater fishes*; Publications Kottelat: Cornol, 2007, ISBN 2839902982.
104. Karabanov, D.P.; Kodukhova, Y.V.; Pashkov, A.N.; Reshetnikov, A.N.; Makhrov, A.A. "Journey to the West": Three phylogenetic lineages contributed to the Invasion of Stone Moroko, *Pseudorasbora parva* (Actinopterygii: Cyprinidae). *Russ. J. Biol. Invasions* **2021**, *12*, 67–78, doi:10.1134/S2075111721010070.
105. Comtet, T.; Sandionigi, A.; Viard, F.; Casiraghi, M. DNA (meta)barcoding of biological invasions: a powerful tool to elucidate invasion processes and help managing aliens. *Biol. Invasions* **2015**, *17*, 905–922, doi:10.1007/s10530-015-0854-y.
106. Hirsch, P.E.; N'Guyen, A.; Adrian-Kalchhauser, I.; Burkhardt-Holm, P. What do we really know about the impacts of one of the 100 worst invaders in Europe? A reality check. *Ambio* **2016**, *45*, 267–279, doi:10.1007/s13280-015-0718-9.
107. Medvedev, D.A.; Sorokin, P.A.; Vasil'ev, V.P.; Chernova, N.V.; Vasil'eva, E.D. Reconstruction of phylogenetic relations of Ponto-Caspian gobies (Gobiidae, Perciformes) based on mitochondrial genome variation and some problems of their taxonomy. *J. Ichthyol.* **2013**, *53*, 702–712, doi:10.1134/S0032945213060064.
108. Neilson, M.E.; Stepien, C.A. Escape from the Ponto-Caspian: evolution and biogeography of an endemic goby species flock (Benthophilinae: Gobiidae: Teleostei). *Mol. Phylogenet. Evol.* **2009**, *52*, 84–102, doi:10.1016/j.ympev.2008.12.023.
109. Freyhof, J. Diversity and distribution of freshwater Gobies from the Mediterranean, the Black and Caspian Seas. In *The Biology of Gobies*; Patzner, R., Van Tassell, J.L., Kovacic, M., Kapoor, B.G., Eds.; CRC Press: USA, Boca Raton, 2011; pp 279–288, ISBN 9780429062872.
110. Kodukhova, Y.V.; Karabanov, D.P. Finding of Longtail Dwarf Goby *Knipowitschia longicaudata* (Actinopterygii: Gobiidae) in the Upper Part of unregulated section of the Volga River. *Inland Water Biol.* **2021**, *14*, In press.
111. Kodukhova, Y.V.; Borovikova, E.A.; Karabanov, D.P. First record of stellate tadpole goby *Benthophilus stellatus* (Sauvage, 1874) (Actinopterygii: Gobiidae) in the Rybinsk Reservoir. *Inland Water Biol.* **2016**, *9*, 428–430, doi:10.1134/S1995082916030081.
112. Slynko, Y.V.; Borovikova, E.A.; Gurovskii, A.N. Phylogeography and origin of freshwater populations of tubenose gobies of genus *Proterorhinus* (Gobiidae: Pisces) in Ponto-Caspian Basin. *Russ. J. Genet.* **2013**, *49*, 1144–1154, doi:10.1134/S1022795413110057.
113. Neilson, M.E.; Stepien, C.A. Evolution and phylogeography of the tubenose goby genus *Proterorhinus* (Gobiidae: Teleostei): evidence for new cryptic species. *Biol. J. Linn. Soc.* **2009**, *96*, 664–684, doi:10.1111/j.1095-8312.2008.01135.x.
114. Kiryukhina, N.A. Molecular and genetic variability in populations of *Syngnathus nigrolineatus* Eichwald, 1831 and ways of expansion in the Volga River basins on the basis of mitochondrial DNA sequence analysis. *Russ. J. Biol. Invasions* **2013**, *4*, 249–254, doi:10.1134/S2075111713040073.

- 
115. Borovikova, E.A.; Alekseeva, Y.I.; Schreider, M.J.; Artamonova, V.S.; Makhrov, A.A. Morphology and genetics of the Ciscoes (Actinopterygii: Salmoniformes: Salmonidae: Coregoninae: *Coregonus*) from the Solovetsky archipelago (White Sea) as a key to determination of the taxonomic position of Ciscoes in Northeastern Europe. *Acta Icth. et Piscat.* **2013**, *43*, 183–194, doi:10.3750/AIP2013.43.3.02.
  116. Borovikova, E.A.; Gerasimov, Y.V.; Karabanov, D.P. Fishes. In *Structure and functioning of the ecosystem in the Rybinsk reservoir at the beginning of the 21<sup>st</sup> Century*; Lazareva, V.I., Ed.; Russian Academy of Sciences: Moscow, 2018; pp 329–339, ISBN 978-5-907036-18-5.
  117. Betancur-R, R.; Wiley, E.O.; Arratia, G.; Acero, A.; Bailly, N.; Miya, M.; Lecointre, G.; Ortí, G. Phylogenetic classification of bony fishes. *BMC Evol. Biol.* **2017**, *17*, 162, doi:10.1186/s12862-017-0958-3.
  118. Mohanty, A.; Swain, S.; Kar, S.K.; Hazra, R.K. Analysis of the phylogenetic relationship of Anopheles species, subgenus *Cellia* (Diptera: Culicidae) and using it to define the relationship of morphologically similar species. *Infect. Genet. Evol.* **2009**, *9*, 1204–1224, doi:10.1016/j.meegid.2009.06.021.
  119. Bergsten, J. A review of long-branch attraction. *Cladistics* **2005**, *21*, 163–193, doi:10.1111/j.1096-0031.2005.00059.x.
  120. Wright, J.J.; David, S.R.; Near, T.J. Gene trees, species trees, and morphology converge on a similar phylogeny of living gars (Actinopterygii: Holostei: Lepisosteidae), an ancient clade of ray-finned fishes. *Mol. Phylogenet. Evol.* **2012**, *63*, 848–856, doi:10.1016/j.ympev.2012.02.033.