Biochemical Systematics and Ecology 67 (2016) 29–36

Contents lists available at ScienceDirect

Biochemical Systematics and Ecology

journal homepage: www.elsevier.com/locate/biochemsyseco

Detection of rare and invasive freshwater fish species using eDNA pyrosequencing: Lake Iznik ichthyofauna revised

Emre Keskin*, Esra Mine Unal, Hasan Hüseyin Atar

Ankara University Faculty of Agriculture Department of Fisheries and Aquaculture Evolutionary Genetics Laboratory (eGL), 06110 Dışkapı Ankara, Turkey

ARTICLE INFO

Article history:
Received 21 April 2016
Received in revised form 16 May 2016
Accepted 22 May 2016
Available online 30 May 2016

Keywords:
eDNA
Pyrosequencing
Cytochrome *b*
Metabarcoding
Freshwater fish

ABSTRACT

Assessment of fish biodiversity in freshwater environments is challenging, especially when rare species or species with low population densities exist. Environmental DNA is becoming a common tool in molecular ecology to detect target species found in the environment. Moreover, eDNA metabarcoding is now used to determine a complete list of target organisms without any prior knowledge on the species inhabiting the environment. This study is the first environmental DNA study designed to assess complete ichthyofauna of the largest lake in Marmara Region of Turkey. For this purpose, an eDNA metabarcoding approach enhanced with tagged primers according to sampling stations for a station specific species listing was used to revise the ichthyofauna of Lake Iznik. Results of pyrosequencing data indicate the presence of 23 species in the lake, five of which are reported for the first time. Short fragment of cytochrome *b* gene sequences amplified in this study were able to make identifications at species level and the eDNA metabarcoding approach was more cost effective and precise compared to conventional surveys. More molecular data from further studies will enhance the reference databases and eDNA metabarcoding could be used more efficiently as an important molecular tool in biodiversity assessment studies.

1. Introduction

As a result of anthropogenic impacts and over exploitation, freshwater habitats become the most threatened ecosystems in terms of species extinction (Hambler et al., 2011). According to IUCN, more than 25% of freshwater animal species were under threat or already extinct worldwide. The most important element in monitoring of a healthy freshwater ecosystem, fish species constitute the main species among the vertebrates with the highest species diversity.

Deliberate or accidental introductions of non-native species are one of the major challenges in terms of anthropogenic impacts on the freshwater ecosystems. Conservation managements, such as early detection, rapid response and eradication have to be implemented at low population densities in order to challenge this problem (Simberloff, 2014). This problem was also a major concern for the European Community, as they have adopted a new regulation (CR 1143/2014) regarding to prevent, minimize and mitigate the adverse impacts of invasive alien species on biodiversity. The regulation specifically recognizes the importance of the rapid identification and detection of non-native species. This could be only achieved by

* Corresponding author.
E-mail address: keskin@ankara.edu.tr (E. Keskin).

http://dx.doi.org/10.1016/j.bse.2016.05.020
0305-1978/

adapting innovative tools to identify target species even at low population densities and at all developmental stages of the organism (Comtet et al., 2015). Freshwater fish species were traditionally monitored by methods based on visual detection, such as staining, tagging and electrofishing. These approaches were not standardized and dependant on the expertise on the field and taxonomic knowledge. Detection of rare or cryptic species is another difficulty in monitoring or conservation biology. It is not always possible to detect the species with low population densities, even with repeated sampling effort which are time consuming and costly.

One of the most widely used molecular tool in identification of target species during any time of its developmental stage is DNA barcoding. At first, DNA barcoding was mainly focussed on taxonomic studies. Today, DNA barcoding, or at least the main principle behind it could be used for a wide range of scientific studies. Many studies were started to use molecular based identification protocols using not only the intact DNA from tissue samples but also the DNA that is released into the environment by the target species. This kind of DNA is called environmental DNA (eDNA) (Pilliod et al., 2013; Keskin, 2014). It could be isolated from environmental samples like water and soil. The main disadvantage of eDNA is that the fragments are shorter and highly degraded and thus an adjusted protocol targeting shorter fragments of targeted gene should be used accordingly (Hajibabaei et al., 2006). Another development related to the use of environmental samples is the ability to identify all the species from a single PCR yield. This new approach is called metabarcoding and its use was enhanced by the progress of Next Generation Sequencing technologies (NGS). NGS allowed consecutive reading of nucleotide sequences from a single (or pooled) PCR yield without even necessity for vectoral cloning (Valentini et al., 2009; Thomsen et al., 2012a, 2012b). eDNA Metabarcoding could be a very useful molecular tool for ecological monitoring studies as it has the potential of identifying all target organisms from a single environmental sample.

This study was designed to test the success of pyrosequencing as eDNA metabarcoding approach for determining the fish fauna of a deep lake of tectonic origin in south-east Marmara region of Turkey. The potential of this approach for monitoring native, non-native and rare (low density) species will be tested for the first time as a freshwater fish fauna survey.

2. Materials and methods

2.1. Locality

Lake Iznik (Fig. 1) is in the Bursa province and the largest freshwater lake in the Marmara region. Lake Iznik is the fifth largest lake in Turkey with a surface area of 313 km^2 and maximum depth of 80 m (Ülgen et al., 2012). Lake Iznik has five main freshwater inputs (Nadir, Kara, Kuru, Kiran and Sölöz rivers) and also some groundwater input (Wester, 1989). The main outflow from the lake to the Marmara Sea occurs through Karsak River. The most detailed fauna study of Lake Iznik (Özuluğ et al., 2005) was resulted in the detection of 19 fish species in the lake.

2.2. Water sampling

Samplings were made through 18 stations from Lake Iznik. Water samples of 1.5 L were collected into sterile containers along the grids shown in Fig. 1. Water samplings were made from surface water at various depths between 1 and 5 m. Four

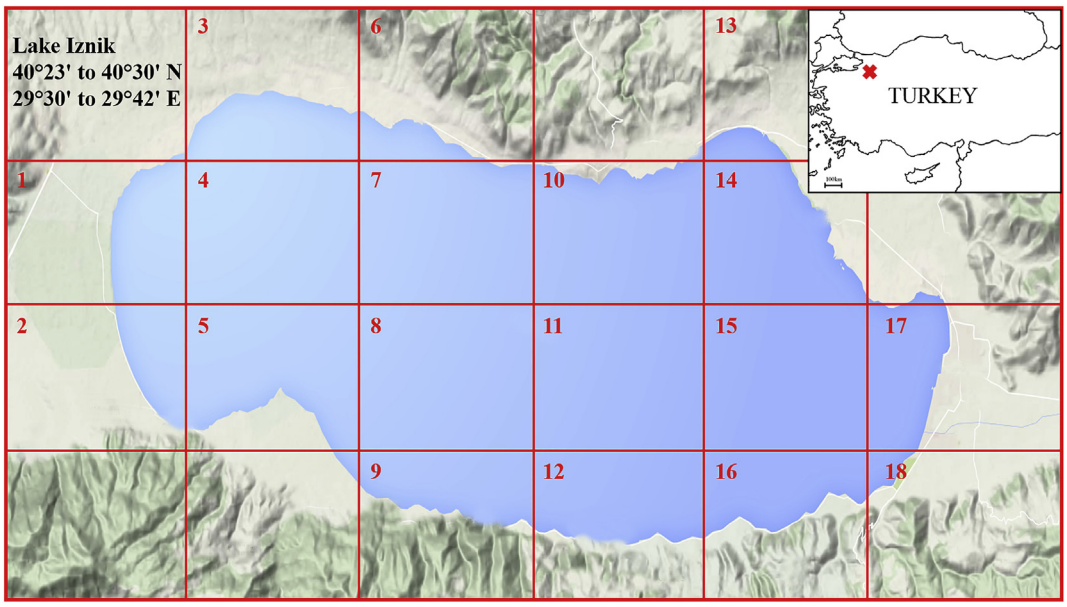

Fig. 1. Sampling grids and station numbers used in water sampling through Lake Iznik.

samples were taken from random points of each grid. Samples were transported to laboratory under dark and cool conditions to minimize further degradation of eDNA. All samples were kept frozen until the filtration. Negative and positive control samples were also taken during the sampling. Water samples were taken from another pond in which no fish species exist and indexed as positive control. Negative controls were taken according to equipment, field and transport which contain only deionized water but treated exactly same with the actual samples (Keskin, 2014).

2.3. Filtration, DNA extraction and PCR

All water samples were filtered within 48 h after the sampling. Water samples were filtered through 0.22 μm pore sized cellulose nitrate filters on stainless steel manifold system operated by a vacuum pump. Filters holding the eDNA were either used in DNA extraction or kept frozen at −20 °C until the DNA extraction process. All filters were subjected to a bead beating prior to DNA extraction. All filters were put into 5 ml skirted tubes with glass beads inside and 500 μl of Lyse BT Buffer (EURx, Poland) added together with 10 μl of Proteinase K. Tubes were vortexed horizontally for 5 min and centrifuged at 4000*g* for one minute. Supernatants were transferred into 2 ml centrifuge tubes. DNA extractions were performed with GeneMATRIX Bio-Trace DNA Purification Kit (EURx, Poland) according to manufacturer's protocol. DNA quantification was performed using Colibri Microvolume Spectrometer (Titertek, Germany).

PCR reactions were performed using fish specific primers targeting 130 bp long fragment of mitochondrial cytochrome *b* gene: FishCBL: TCCTTTTGAGGCGCTACAGT and FishCBR: GGAATGCGAAGAATCGTGTT (Thomsen et al., 2012a). Primers were modified by adding specific tags on 5′ end to make it possible distinguish PCR yields of each sample. PCR amplifications were performed using 8 μl of 5x FIREPol Master Mix Ready to Load (12.5 mM MgCl2) (Solis BioDyne, Estonia), 1 μl of each primer (F, R), 2 μl of template DNA and 28 μl of ultrapure water within a total reaction volume of 40 μl. Thermal cycler conditions were set to 7 min at 95 °C, 50 cycles of 30 s at 94 °C, 30 s at 50 °C, 20 s at 72 °C, completed with a final extension for 5 min at 72 °C. All PCR amplifications were performed in triplicates in correspondence to multiple tubes approach (Taberlet et al., 1996; Goldberg et al., 2013a,b). Positive control samples extracted from tissues of species known not to occur at the site were used to check barcoding integrity. Negative control samples were also included to be sure about contamination is not introduced in pyrosequencing process.

2.4. Pyrosequencing

PCR yields were visualized on 2% agarose gel stained with ethidium bromide and purified using Wizard SV Genomic DNA Purification Kit (Promega, USA) according to the manufacturer's protocol. PCR products were also quantified to standardize the DNA concentration per PCR yield from each grid. Fusion primers for PCR were designed according to GS Junior System Guidelines for Amplicon Experimental Design with the unidirectional sequencing protocol of Lib-L emPCR Kit One Way Reads experimental design. Adaptors for pyrosequencing and multiplex identifier (MID) were used for sample identification after sequencing. Sequencing of amplicon libraries were sequenced with Roche 454 Life Sciences GS Junior platform using NGS Service of Enigma Biotechnology (Ankara, Turkey). PCR yields were pooled and sequenced using Roche GS FLX 454 Pyro-sequencing (Roche Switzerland) following manufacturer's instructions.

To mitigate detection of false positive signals sample extractions and PCR reactions were conducted according to forensic principles and pipelines to avoid cross contamination according to Biggs et al. (2014).

2.5. Data analysis

Sequences were sorted into different files according to multiplex identifier (MID) using the Roche Newblr package. Sff_extract script was used to extract sequences from SFF files and convert them into FASTA, XML and quality files. FASTS and quality files were converted into FASTQ file using a Python script (Python version 3.4.3). Quality and length of reads were examined in FastQC v0.11.4. Lucy DNA sequence quality and vector trimming tool v1.20 was used to trim reads according to quality scores. In order to reduce homopolymer and sequencing errors a consensus sequence was generated from each MOTU and alignments were manually checked in Seaview version 3.2. Finally, Basic Local Alignment Search Tool (BLAST) of NCBI GenBank was used to identify consensus sequences of the MOTUs. Tags on 5′ end of primers were used to distinguish different sequences inside the raw data gathered from large scale pyrosequencing. Therefore, a new data set including only sequences from the relevant grid was generated for each sample. Assignment of sequences to the species was performed according to nucleotide similarity assessment provided by BLAST search engine (Zhang et al., 2000) from the National Center for Biotechnology Information (NCBI) GenBank database. Only sequences exist more than 3 times in the raw data set were used in order to avoid sequencing errors (Huse et al., 2007). BLAST results with only 100% of query coverage and $\geq$98% similarity were set as the threshold.

3. Results

A total of 72 water samples were collected from the largest freshwater lake in Marmara region of Turkey, filtered, DNA amplified using fish specific primers targeting 130 base pairs long fragment of mitochondrial cytochrome *b* and sequenced. eDNA was successfully extracted from all of the sampling stations. None of the replicates turned out to be negative in terms of

eDNA extraction. Positive PCR results with expected size of PCR yield (130 base pairs) were also gathered from every station using DNA's pooled according to stations. No indication of contamination found for DNA and PCR as all the blank samples taken for equipment, field and transport; and the positive samples taken from artificial pond in which no fish species exist were also analysed and found to be negative.

Sequences from raw pyrosequencing data were compared using BLAST tool of NCBI GenBank database were resulted in sequences of 23 different fish species from 6 orders and 8 families (Table 1). All of the species detected have a unique cytochrome *b* sequence that enabled molecular identification at species level. Only some of the species have the same identification level with more than one species but eventually only one of those were found in freshwater ichthyofauna of Turkey (explained in discussion section). Sequences were submitted to GenBank database under accession numbers KT461497–KT461519.

Most of the species (16/23; 70%) were belong to order Cypriniformes. Out of 23 species detected, 8 of them were along commercially important freshwater fish species of Turkey, including: Big-scale sand smelt (*Atherina boyeri*), Freshwater bream (*Abramis brama*), Prussian carp (*Carassius gibelio*), Common carp (*Cyprinus carpio*), Chub (*Squalius cephalus*), Rudd (*Scardinius erythrophthalmus*), Tench (*Tinca tinca*) and Wels catfish (*Silurus glanis*). Some of the species detected were considered as highly invasive species like Prussian carp (*Carassius gibelio*), Common carp (*Cyprinus carpio*) and Eastern mosquitofish (*Gambusia holbrooki*) according to results from invasiveness screening kit assessments by Tarkan et al. (2014).

eDNA metabarcoding results were compared to previously published fish fauna reports conducted by conventional field surveys. eDNA metabarcoding results were better than conventional survey results with 23 to 19 species. Among 23 species detected, 5 of them; Freshwater bream (*Abramis brama*), Dnieper chub (*Petroleuciscus borysthenicus*), Bitterling (*Rhodeus amarus*), South European roach (*Samarutilus rubilio*) and Rudd (*Scardinius erythrophthalmus*) were reported for the first time from the Lake Iznik. Only one species from the list provided from conventional surveys, Angora loach (*Oxynemacheilus angorae*), could not be detected by eDNA metabarcoding. Distribution of the species (based on eDNA samples) according to stations were given in Table 2.

Most of the sequences that were identified at species level were found >100 times through the pyrosequencing data. Sequences with only ≤0.8% variation (or 1/130 base pair) level from the highly represented sequences were considered as sequencing errors based on 5′ tags used in modification of primers used in PCR. The most common sequence found in the dataset was belonging to Prussian carp (*Carassius gibelio*). A total of 156,432 reads were obtained in which 6746 were in good quality. Sequences, BLAST identity percentages and percentage of repeats in the pyrosequencing data were given in Table 3.

4. Discussion

Molecular identification of species becomes an alternative method in biodiversity studies following widespread adoption of DNA barcoding (Hebert et al., 2003). The ability of the method to make identifications at species level is a challenge at the beginning but with the development of DNA libraries it became more useful every day. Related to progress in the field of molecular biology, new tools, such as eDNA barcoding or metabarcoding, based on species level identification approach developed in the recent years which will not only suitable for molecular taxonomy but also applicable to studies in ecology. eDNA allows researchers to detect the species without sampling the species itself, but by taking environmental samples (such

Table 1
Summary of species detected in Lake İznik using eDNA pyrosequencing.

Order	Family	Species
Atheriniformes	Atherinidae	*Atherina boyeri*
Cypriniformes	Cobitidae	*Cobitis vardarensis*
	Cyprinidae	*Abramis brama*
	Cyprinidae	*Alburnus alburnus*
	Cyprinidae	*Alburnus chalcoides*
	Cyprinidae	*Barbus tauricus*
	Cyprinidae	*Capoeta tinca*
	Cyprinidae	*Carassius gibelio*
	Cyprinidae	*Cyprinus carpio*
	Cyprinidae	*Petroleuciscus borysthenicus*
	Cyprinidae	*Rhodeus amarus*
	Cyprinidae	*Rutilus frisii*
	Cyprinidae	*Rutilus rutilus*
	Cyprinidae	*Sarmarutilus rubilio*
	Cyprinidae	*Scardinius erythrophthalmus*
	Cyprinidae	*Squalius cephalus*
	Cyprinidae	*Tinca tinca*
Cyprinodontiformes	Poeciliidae	*Gambusia holbrooki*
Gasterosteiformes	Gasterosteidae	*Gasterosteus aculeatus*
Perciformes	Gobiidae	*Proterorhinus marmoratus*
	Blenniidae	*Salaria fluviatilis*
Siluriformes	Siluridae	*Silurus glanis*

Table 2
Distribution of the species (based on eDNA samples) according to stations.

Species	Stations																	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Abramis brama																	X	
Alburnus alburnus			X	X		X	X		X	X		X	X				X	X
Alburnus chalcoides				X						X				X		X	X	X
Atherina boyeri			X		X	X	X		X	X	X	X	X	X	X	X	X	X
Barbus tauricus	X	X												X				
Capoeta tinca														X			X	
Carassius gibelio	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Cobitis vardarensis	X	X																
Cyprinus carpio			X			X			X	X		X	X	X		X	X	X
Gambusia holbrooki		X													X			
Gasterosteus aculeatus	X	X																
Petroleuciscus borysthenicus	X	X																
Proterorhinus marmoratus	X	X																
Rhodeus amarus	X									X								
Rutilus frisii														X			X	
Rutilus rutilus															X		X	
Salaria fluviatilis	X																	
Sarmarutilus rubilio	X																	
Scardinius erythrophthalmus				X	X													
Silurus glanis				X	X		X			X				X	X		X	
Squalius cephalus	X	X		X	X									X			X	
Tinca tinca															X		X	X
Vimba vimba	X	X																
Total number of species	11	9	4	6	5	4	4	1	4	7	2	4	4	9	6	4	12	6

as soil, water) which includes the DNA released into environment by the species. In species specific eDNA approach, specific primers for the target organisms are designed and used in PCR amplification. The main limitation of the species specific approach is being an *a priori* approach, depending on the previous data for the expected species. This resulted in detection of only expected species which in that case does not allow detection of rare, new or non-native species in the environment (Herder et al., 2014). This could be exceeded by conducting multiple PCRs with different primers or designing multiplex primers to target many species from the same eDNA sample. However this approach is also restricted with the limited amount of eDNA eluted in small volumes through the DNA extraction (Jerde et al., 2011; Thomsen et al., 2012b) to avoid diluting the degraded DNA which is already low in both quality and quantity. The multi-tube approach suggested by Taberlet et al. (1996) is often used in eDNA studies for statistically significant replicates and this method prominently reduces the number of possible PCR amplifications per eDNA extract. As a result, designing many species specific primers could be difficult, expensive and time consuming when the study is targeting a wide range of species, or even impossible to be used for unknown, rare or non-native species found in the environment.

Next generation sequencing (NGS) approach allows the researchers to analyse multiple eDNAs simultaneously. Recent progress in NGS platforms makes it more cost-effective to use eDNA metabarcoding approach to identify all the species found in the environment. This approach is performed by using universal primers specific to a wide range of target species (such as fishes, mammals, plants, amphibians) rather than single organism. PCR yields were sequenced using a NGS platform and the resulting raw data is analysed with bioinformatics tools to select the relevant sequences to be compared with a reference database like NCBI GenBank to make identifications at species level. This method is known as eDNA metabarcoding (Taberlet et al., 2012).

eDNA metabarcoding has been used in many studies including detection of fish species (Thomsen et al., 2012a,b; Kelly et al., 2014; Evans et al., 2016), aquatic plants (Cowart et al., 2015; Visco et al., 2015), macroinvertebrates (Elbrecht and Leese, 2015), amphibians (Evans et al., 2016), arthropods (Yu et al., 2012), protists (Pawlowski et al., 2014; Geisen et al., 2015; Kosakyan et al., 2015), soil biodiversity (Bienert et al., 2012; Taberlet et al., 2012; Yoccoz et al., 2012), diet assessment (Valentini et al., 2009; Deagle et al., 2013; De Barba et al., 2014) and ballast water surveillance (Zaiko et al., 2015). This study presents the first eDNA metabarcoding results of an entire lake ichthyofauna.

The aim of the study was to assess the ichthyofauna of Lake Iznik by performing a large scale pyrosequencing approach using water samples as a source of eDNA and amplifying a short fragment of mitochondrial cytochrome *b* gene. Results are showing that this approach successfully detected more species than any other conventional survey conducted here before, with 23 species including 5 invasive species and 5 new species that haven't been reported in the latest surveys. The only species that could not been detected using eDNA metabarcoding is Angora loach (*Oxynemacheilus angorae*), which was reported during a survey conducted more than a decade ago.

The sampling effort in metabarcoding approach is definitely low when compared to conventional surveys as you can easily finish the sampling of a complete lake (300 km^2 in this study) with 2 person in 2 days, which requires at least 10 times more persons and days in case of conventional survey. The sampling gears that will be used in the field are also important as all you

Table 3
Species detected in Lake Iznik, BLAST identity percentages, number of reads, percentage of repeats in the pyrosequencing data and sequences without the primers.

Species	BLAST %	Reads[a]	Repeat %	Cytochrome *b* sequence
Abramis brama	99	46	0.68	AATCACAAATCTCCTCTCAGCAGTCCCTTATATAGGAGACACTCTTGTCCAATGAATCTGAGGCGG CTTTTCAGTAGACAACGCAACTCTC
Alburnus alburnus	99	523	7.75	AATTACGAACCTCCTCTCAGCAGTTCCCTACATGGGAGATACCCTTGTTCAATGAATTTGGGGCGG TTTCTCAGTAGATAACGCGACTCTT
Alburnus chalcoides	99	275	4.07	AATCACGAACCTCCTCTCAGCAGTCCCCTATATAGGAGACACCCTTGTTCAATGAATTTGAGGCGG TTTCTCAGTAGATAACGCAACCCTT
Atherina boyeri	99	1373	20.35	AATCACTAACCTCCTCTCAGCGGTGCCATACGTTGGTAACTCCTTAGTTCAGTGAATCTGGGGGGG CTTTTCCGTGGATAACGCCACCCTC
Barbus tauricus	99	98	1.45	AATTACAAATCTCCTCTCCGCCGTGCCATATATAGGGGACATACTAGTCCAATGAATCTGAGGCGG CTTTTCAGTAGATAACGCAACACTG
Capoeta tinca	99	78	1.16	AATTACAAATTTCCTATCCGCCGTCCCGTACATAGGCGATATACTAGTCCAGTGAATTTGAGGTGG ATTTTCAGTAGATAATGCAACACTA
Carassius gibelio	100	2353	34.88	AATCACAAACCTTCTATCCGCCGTGCCATACATGGGAGATATGTTAGTTCAATGAATTTGAGGAGG CTTCTCCGTAGACAATGCAACATTA
Cobitis vardarensis	99	65	0.97	AATTACTAACCTTCTTTCAGCCGTCCCCTATGTAGGAAATGCCCTAGTCCAGTGAATCTGAGGTGG ATTCTCAGTAGATAATGCTACACTA
Cyprinus carpio	100	327	4.84	AATCACAAACCTCCTATCTGCCGTACCATACATGGGAGACATGTTAGTCCAATGAATCTGAGGTGG GTTCTCAGTAGACAATGCAACACTA
Gambusia holbrooki	99	65	0.97	GATTACCAACCTCCTGTCTGCTGTCCCTTACATGGGAGACACCCTTGTCCAATGAATTTGAGGGGG ATTTTCAGTTGATAATGCCACCTTA
Gasterosteus aculeatus	99	53	0.78	TATTACCAACCTACTTTCAGCCGTCCCTTACGTTGGTAATTCATTAGTTCAATGAATTTGAGGAGGC TTTTCCGTTGACAACGCCACCTTA
Petroleuciscus borysthenicus	99	78	1.16	AATCACCAATCTACTCTCCGCAGTCCCTTACATAGGAGATACCCTTGTGCAGTGAATCTGAGGCGG CTTCTCCGTAGACAACGCAACCCTC
Proterorhinus marmoratus	100	65	0.97	AATTACAAACCTCCTCTCTGCTATCCCTTATGTAGGAACTGATCTAGTACAGTGAATTTGAGGGGG CTTCTCAGTTGATAACGCAACCCTC
Rhodeus amarus	100	92	1.36	AATTACCAATCTACTCTCAGCGGTCCCTTATATAGGGGACGCTCTGGTTCAATGAATTTGAGGCGG GTTCTCAGTAGACAACGCAACACTA
Rutilus frisii	100	53	0.78	AATCACGAACCTCCTCTCAGCAGTCCCTTACATAGGAGATACCCTCGTTCAATGAATCTGAGGCGG TTTCTCAGTAGACAACGCAACCCTT
Rutilus rutilus	99	65	0.97	AATCACCAACCTCCTCTCAGCGGTCCCTTACATGGGAGATACCCTTGTTCAGTGAATCTGGGGAGG TTTCTCAGTAGATAACGCAACCCTT
Salaria fluviatilis	100	46	0.68	AATTACCAACCTTCTCTCAGCTTTCCCATATGTCGGAAGCACACTTGTTCAATGAATTTGAGGGGGC TTCTCGATTGACAACGCCACCCTC
Sarmarutilus rubilio	99	33	0.49	AATTACAAACCTCCTCTCAGCAGTCCCTTACATAGGAGATACTCTTGTTCAGTGAATCTGAGGCGGT TTCTCCGTAGACAACGCGACCCTT
Scardinius erythrophthalmus	100	105	1.55	AATTACAAACCTCCTCTCAGCAGTCCCCTACATAGGAGATACCCTTGTTCAGTGAATCTGAGGCGG TTTCTCAGTAGACAACGCGACCCTA
Silurus glanis	100	412	6.10	AATTACAAATTTATTATCCGCCGTTCCCTACATAGGAGATGCTCTAGTACAGTGAATCTGAGGGGG CTTTTCTGTAGACAACGCAACTCTC
Squalius cephalus	100	353	5.23	AATTACAAACCTCCTCTCAGCAGTCCCTTACATAGGGGACACTCTTGTTCAATGAATCTGAGGCGG TTTCTCCGTAGATAACGCGACCCTT
Tinca tinca	99	98	1.45	AATTACTAACCTACTATCAGCAGTTCCCTACATAGGAGATGCTTTAGTTCAATGAATCTGAGGGGG CTTCTCAGTAGACAATGCAACTCTT
Vimba vimba	100	92	1.36	AATTACAAACCTCCTCTCAGCAGTCCCTTATATAGGCGACACCCTTGTCCAATGAATCTGAGGCG GTTTCTCAGTAGATAACGCAACTCTC

[a] Number of reads and repeat percentage were calculated from 6746/156432 sequences.

need is sterile containers for water samples in metabarcoding approach while a lot of nets, traps and electrofishing equipment will be needed for the conventional surveys, not even mentioning the conventional method depends heavily on expertise and needs permissions for related water body, species and fishing season. Metabarcoding approach is also cost-effective compared to species specific approach as only one primer pair and single sequencing will be used per sample in order to detect the species. This means 23x less PCR amplification and sequencing for this study, saving a lot of time and expense in order to assess the ichthyofauna of the whole lake. Also, results of molecular identification are more robust and objective when compared to visual identification. Hundreds of samples could be analysed in a single pyrosequencing run and the raw data could be easily analysed using bioinformatics tools (Boyer et al., 2016) developed for metabarcoding data in biodiversity assessment studies.

The limitation about the resolution of the data not reaching the species level (Valentini et al., 2009) is not valid for fish species as the universal cytochrome *b* primers could easily be used and GenBank database covers a wide range of cytochrome *b* sequences belonging to thousands of fish species. All of the sequences were identified to species level (BLAST identification percentages $\geq$ 99) in our study, even the species from the same genus such as *Alburnus alburnus-Alburnus chalcoides* and *Rutilus fresii-Rutilus rutilus*, confirming the resolution of selected fragment of cytochrome *b* region and sufficiency of the reference database for fish species. There is some BLAST comparisons resulted in same identification percentage for more than one species. One of the sequences was resulted in 99% identification with *Abramis brama* and *Blicca bjoerkna*. These species

were very similar and could be misidentified in previous conventional surveys, but it is accepted as *A. brama* as *B. bjoerkna* has never been reported from Lake Iznik whereas *A. brama* is a native species of the lake. Also it is accepted as *A. brama* as it is found out to be related with a possible misidentification of *A. brama* as *B. bjoerkna* in GenBank database as other reference sequences belonging to *B. bjoerkna* (HM560077 and NC020355), in which one of them is from a complete mitochondrial genome (having less chance of sequencing errors compared to partial fragments) resulted with a genetic distance of 7.5%, indicating our sequence is not belong to *B. bjoerkna*. Another result with same BLAST identity percentage is found in *Capoeta tinca-Capoeta baliki*. *C. baliki* has never been reported from Lake Iznik and it is more likely that the sequence we found belong to *C. tinca*, a native species of the lake. Other *C. baliki* sequences were downloaded from GenBank (JF798271 and GQ424020) and the genetic distance is calculated as 4.8 and 3.1 respectively, indicating another misidentification related confusion. As a result, the sequence was accepted as belonging to *C. tinca*. Sequence of *Cyprinus carpio* is also have the same identity level with three other species from the same genus (*C. rubrofuscus*, *C. acutidorsalis* and *C. cultitaeniata*) but none of the species were belong to freshwater fauna of Turkey and tropical species that were found only in China and Vietnam. Another case with same identity rates with species from same genus is found in *Scardinius erythrophtalmus* (*S. dergle* and *S. hesperidicus*) but the other two species were not also ever reported from Turkey and *S. erythrophthalmus* is a well-known native species of the lake, like *C. carpio*. The most challenging one is the sequence with same identification level for *C. gibelio* and *C. auratus* as these species were very similar in terms of both nucleotide sequence and morphological characters and could be easily misidentified. *C. auratus* has never been reported from this lake and *C. gibelio* is the far most intensively caught species in the lake, as the sequence we related with *C. gibelio* has the highest occurrence rate through the metabarcoding metadata were all indicating the sequences were belonging to *C. gibelio*. Complete mitochondrion sequences (KJ874428, KJ874430, KF147851, NC002079) belonging to *C. auratus* were downloaded from GenBank and confirms our result as *C. gibelio* with genetic distance were calculated between 3.2 and 7.4%, indicating another possibly misidentified sequence match.

This approach is applicable for every freshwater lake and different organism groups (with the use proper primer set), with few minor constraints same with the more conventional species specific eDNA studies. Once the NGS raw data processed using bioinformatics tools, it is almost the same procedure to compare the sequences with the ones in the reference database. The same problem with the public database also exist for the metabarcoding data as there is always a probability of encountering reference sequences with errors (Harris, 2003) and misidentifications (Santos and Branco, 2012). Alternative databases, such as BOLD, could be used as a source of verified references but unfortunately it's restricted with only mitochondrial cytochrome oxidase I sequences for vertebrates, which is not suitable for metabarcoding studies because of the primers targeting this region is not suitable for this approach (Herder et al., 2014).

Acknowledgments

The authors acknowledge COST Action TD1209 European Information System on Alien Species and The Scientific and Technological Research Council of Turkey (TUBITAK, Project No. 114R003) for funding.

References

Bienert, F., De Danieli, S., Miquel, C., Coissac, E., Poillot, C., Brun, J.J., Taberlet, P., 2012. Tracking earthworm communities from soil DNA. Mol. Ecol. 21, 2017–2030.
Biggs, J., Ewald, N., Valentini, A., Gaboriaud, C., Griffiths, R.A., Foster, J., Wilkinson, J., Arnett, A., Williams, P., Dunn, F., 2014. Analytical and Methodological Development for Improved Surveillance of the Great Crested Newt. Appendix 5. Technical Advice Note for Field and Laboratory Sampling of Great crested Newt (*Triturus cristatus*) Environmental DNA. Freshwater Habitats Trust, Oxford.
Boyer, F., Mercier, C., Bonin, A., Le Bras, Y., Taberlet, P., Coissac, E., 2016. obitools: a unix-inspired software package for DNA metabarcoding. Mol. Ecol. Resour 16 (1), 176–182.
Comtet, T., Sandionigi, A., Viard, F., Casiraghi, M., 2015. DNA (meta) barcoding of biological invasions: a powerful tool to elucidate invasion processes and help managing aliens. Biol. Invasions 17, 905–922.
Cowart, D.D., Pinheiro, M., Mouchel, O., Maguer, M., Grall, J., Mine, J., Arnaud-Haond, S., 2015. Metabarcoding is powerful yet still blind: a comparative analysis of morphological and molecular surveys of seagrass communities. PLoS One 10 (2) e0117562.
De Barba, M., Miquel, C., Boyer, F., Mercier, C., Rioux, D., Coissac, E., Taberlet, P., 2014. DNA metabarcoding multiplexing and validation of data accuracy for diet assessment: application to omnivorous diet. Mol. Ecol. Resour. 14, 306–323.
Deagle, B.E., Thomas, A.C., Shaffer, A.K., Trites, A.W., Jarman, S.N., 2013. Quantifying sequence proportions in a DNA-based diet study using Ion Torrent amplicon sequencing: which counts count? Mol. Ecol. Resour. 13, 620–633.
Elbrecht, V., Leese, F., 2015. Can DNA-based ecosystem assessments quantify species abundance? Testing primer bias and biomass-sequence relationships with an innovative metabarcoding protocol. PLoS One 10, e0130324.
Evans, N.T., Olds, B.P., Turner, C.R., Renshaw, M.A., Li, Y., Jerde, C.L., Mahon, A.R., Pfrender, M.E., Lamberti, G.A., Lodge, D.M., 2016. Quantification of mesocosm fish and amphibian species diversity via environmental DNA metabarcoding. Mol. Ecol. Resour. 16 (1), 29–41.
Geisen, S., Laros, I., Vizcaíno, A., Bonkowski, M., de Groot, G.A., 2015. Not all are free-living: high-throughput DNA metabarcoding reveals a diverse community of protists parasitizing soil metazoa. Mol. Ecol. 24, 4556–4569.
Goldberg, C.S., Sepulveda, A., Ray, A., Baumgardt, J., Waits, L.P., 2013a. Environmental DNA as a new method for early detection of New Zealand mudsnails (*Potamopyrgus antipodarum*). Freshw. Sci. 32, 792–800.
Goldberg, C.S., Laramie, M.B., Waits, L.P., 2013b. Application of Environmental DNA for Inventory and Monitoring of Aquatic Species. US Department of the Interior, US Geological Survey.
Hajibabaei, M., Smith, M., Janzen, D.H., Rodriguez, J.J., Whitfield, J.B., Hebert, P.D., 2006. A minimalist barcode can identify a specimen whose DNA is degraded. Mol. Ecol. Notes 6, 959–964.
Hambler, C., Henderson, P.A., Speight, M.R., 2011. Extinction rates, extinction-prone habitats, and indicator groups in Britain and at larger scales. Biol. Conserv. 144, 713–721.
Harris, D.J., 2003. Can you bank on GenBank? Trends Ecol. Evol. 18, 317–319.
Hebert, P.D., Cywinska, A., Ball, S.L., 2003. Biological identifications through DNA barcodes. Proc. R. Soc. Lond. B Biol. Sci. 270, 313–321.

Herder, J., Valentini, A., Bellemain, E., Dejean, T., von Delft, J., Thomsen, P.F., Taberlet, P., 2014. Environmental DNA - a review of the possible applications for the detection of (invasive) species. Stichting RAVON. Nijmegen. Rep. 2013–2104.
Huse, S.M., Huber, J.A., Morrison, H.G., Sogin, M.L., Welch, D.M., 2007. Accuracy and quality of massively parallel DNA pyrosequencing. Genome Biol. 8, R143.
Jerde, C.L., Mahon, A.R., Chadderton, W.L., Lodge, D.M., 2011. "Sight-unseen" detection of rare aquatic species using environmental DNA. Conserv. Lett. 4, 150–157.
Kelly, R.P., Port, J.A., Yamahara, K.M., Crowder, L.B., 2014. Using environmental DNA to census marine fishes in a large mesocosm. PLoS One 9, e86175.
Keskin, E., 2014. Detection of invasive freshwater fish species using environmental DNA survey. Biochem. Syst. Ecol. 56, 68–74.
Kosakyan, A., Mulot, M., Mitchell, E.A., Lara, E., 2015. Environmental DNA COI barcoding for quantitative analysis of protists communities: a test using the Nebela collaris complex (Amoebozoa; Arcellinida; Hyalospheniidae). Eur. J. Protistol. 51, 311–320.
Özuluğ, M., Altun, Ö., Meriç, N., 2005. On the fish fauna of Lake Iznik (Turkey). Turkish J. Zool. 29, 371–375.
Pawlowski, J., Esling, P., Lejzerowicz, F., Cedhagen, T., Wilding, T.A., 2014. Environmental monitoring through protist next-generation sequencing metabarcoding: assessing the impact of fish farming on benthic foraminifera communities. Mol. Ecol. Resour. 14, 1129–1140.
Pilliod, D.S., Goldberg, C.S., Laramie, M.B., Waits, L.P., 2013. Application of environmental DNA for inventory and monitoring of aquatic species: US Department of the Interior. US Geological Survey.
Santos, A.M., Branco, M., 2012. The quality of name-based species records in databases. Trends Ecol. Evol. 27, 6–7.
Simberloff, D., 2014. Biological invasions: what's worth fighting and what can be won? Ecol. Eng. 65, 112–121.
Taberlet, P., Griffin, S., Goossens, B., Questiau, S., Manceau, V., Escaravage, N., Waits, L.P., Bouvet, J., 1996. Reliable genotyping of samples with very low DNA quantities using PCR. Nucleic Acids Res. 24, 3189–3194.
Taberlet, P., Coissac, E., Hajibabaei, M., Rieseberg, L.H., 2012. Environmental DNA. Mol. Ecol. 21, 1789–1793.
Tarkan, A.S., Güler Ekmekçi, F., Vilizzi, L., Copp, G.H., 2014. Risk screening of non-native freshwater fishes at the frontier between Asia and Europe: first application in Turkey of the fish invasiveness screening kit. J. Appl. Ichthyol. 30, 392–398.
Thomsen, P.F., Kielgast, J., Iversen, L.L., Wiuf, C., Rasmussen, M., Gilbert, M.T.P., Orlando, L., Willerslev, E., 2012a. Monitoring endangered freshwater biodiversity using environmental DNA. Mol. Ecol. 21, 2565–2573.
Thomsen, P.F., Kielgast, J., Iversen, L.L., Møller, P.R., Rasmussen, M., Willerslev, E., 2012b. Detection of a diverse marine fish fauna using environmental DNA from seawater samples. PLoS One 7, e41732.
Ülgen, U.B., Franz, S.O., Biltekin, D., Çagatay, M.N., Roeser, P.A., Doner, L., Thein, J., 2012. Climatic and environmental evolution of Lake Iznik (NW Turkey) over the last~ 4700 years. Quat. Int. 274, 88–101.
Valentini, A., Miquel, C., Nawaz, M.A., Bellemain, E., Coissac, E., Pompanon, F., Gielly, L., Cruaud, C., Nascetti, G., Wincker, P., Swenson, J.E., Taberlet, P., 2009. New perspectives in diet analysis based on DNA barcoding and parallel pyrosequencing: the trnL approach. Mol. Ecol. Resour. 9, 51–60.
Visco, J.A., Apotheloz-Perret-Gentil, L., Cordonier, A., Esling, P., Pillet, L., Pawlowski, J., 2015. Environmental monitoring: inferring the diatom index from next-generation sequencing data. Environ. Sci. Technol. 49, 7597–7605.
Wester, R., 1989. Wasserwirtschaftlicher Rahmenplan für das Izniksee-Gebiet/Türkei. PhD Thesis. Fachhochschule Mainz.
Yoccoz, N.G., Bråthen, K.A., Gielly, L., Haile, J., Edwards, M.E., Goslar, T., Von Stedingk, H., Brysting, A.K., Coissac, E., Pompanon, F., Sønstebø, J.H., Miquel, C., Valentini, A., De Bello, F., Chave, J., Thuiller, W., Wincker, P., Cruaud, C., Gavory, F., Rasmussen, M., Gilbert, M.T.P., Orlando, L., Brochmann, C., Willerslev, E., Taberlet, P., 2012. DNA from soil mirrors plant taxonomic and growth form diversity. Mol. Ecol. 21, 3647–3655.
Yu, D.W., Ji, Y., Emerson, B.C., Wang, X., Ye, C., Yang, C., Ding, Z., 2012. Biodiversity soup: metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. Methods Ecol. Evol. 3, 613–623.
Zaiko, A., Martinez, J.L., Schmidt-Petersen, J., Ribicic, D., Samuiloviene, A., Garcia-Vazquez, E., 2015. Metabarcoding approach for the ballast water surveillance–An advantageous solution or an awkward challenge? Mar. Pollut. Bull. 92, 25–34.
Zhang, Z., Schwartz, S., Wagner, L., Miller, W., 2000. A greedy algorithm for aligning DNA sequences. J. Comput. Biol. 7, 203–214.