Assessment of fish biodiversity in four Korean rivers using environmental DNA metabarcoding

Md Jobaidul Alam [1], **Nack-Keun Kim** [1], **Sapto Andriyono** [2], **Hee-kyu Choi** [3], **Ji-Hyun Lee** [4], **Hyun-Woo Kim** [Corresp. 1, 4]

[1] Interdisciplinary Program of Biomedical, Mechanical and Electrical Engineering, Pukyong National University, Busan, Republic of Korea

[2] Fisheries and Marine Faculty, C Campus Jl. Mulyorejo Surabaya, Universitas Airlangga, Surabaya, East Java, Indonesia

[3] Molecular Ecology and Evolution Laboratory, Department of Biological Science, College of Science & Engineering, Sangji University, Wonju, Republic of Korea

[4] Department of Marine Biology, Pukyong National University, Busan, Republic of Korea

Corresponding Author: Hyun-Woo Kim
Email address: kimhw@pknu.ac.kr

Environmental DNA (eDNA) metabarcoding is a cost-effective novel approach to estimate biodiversity in an ecosystem. In this study, the MiFish pipeline was employed to test if the system methodology is sufficiently reliable to estimate fish biodiversity in Korean rivers. A total of 125 unique haplotypes and 73 species were identified at the species level from 16 water samples collected from a single survey in four Korean rivers (Hyeongsan, Taehwa, Seomjin, and Nakdong). Among the four rivers, the highest species richness was recorded in the Seomjin River (52 species), followed by the Taehwa (42 species) and Hyeongsan (40 species) rivers. The Nakdong River (26 species) presented the lowest species richness and number of endemic species, presumably due to its metropolitan location and anthropogenic impacts, such as dams or weirs. We were also able to detect that five exotic species (*Carassius cuvieri, Cyprinus carpio, Cyprinus megalophthalmus, Lepomis macrochirus,* and *Micropterus salmoides*) are widely distributed in all surveyed rivers, a situation that might be problematic in terms of conservation. Our findings indicate that the eDNA metabarcoding technique is one of the most cost-effective scientific tools available for the management and conservation of the freshwater fish resources available in Korea. However, the low number of 12S sequences of endemic species in the database and low resolution of the MiFish region for differentiating several taxa should be upgraded for their wide use.

1 **Assessment of fish biodiversity in four Korean rivers using environmental DNA**

2 **metabarcoding**

3 Md. Jobaidul Alam[1], Nack-Keun Kim[1], Sapto Andriyono[1,2], Hee-kyu Choi[3], Ji-Hyun Lee[4], and

4 Hyun-Woo Kim[1,4]*

6 [1]Interdisciplinary Program of Biomedical, Mechanical and Electrical Engineering, Pukyong

7 National University, Busan, 48513, Republic of Korea

8 [2]Fisheries and Marine Faculty, C Campus Jl. Mulyorejo Surabaya 60115. Universitas Airlangga,

9 Surabaya, East Java, Indonesia

10 [3]Molecular Ecology and Evolution Laboratory, Department of Biological Science, College of

11 Science & Engineering, Sangji University, Wonju 26339, Republic of Korea

12 [4]Department of Marine Biology, Pukyong National University, Busan 48513, Republic of

13 Korea

16 * Corresponding author:

17 Hyun-Woo Kim, Ph. D

18 Department of Marine Biology

19 Pukyong National University

20 48513, Republic of Korea

21 Tel: 82-51-629-5926

22 Fax: 82-51-629-5930

23 E-mail: kimhw@pknu.ac.kr

28 **ABSTRACT**

29 Environmental DNA (eDNA) metabarcoding is a cost-effective novel approach to estimate

30 biodiversity in an ecosystem. In this study, the MiFish pipeline was employed to test if the

31 system methodology is sufficiently reliable to estimate fish biodiversity in Korean rivers. A total

32 of 125 unique haplotypes and 73 species were identified at the species level from 16 water

33 samples collected from a single survey in four Korean rivers (Hyeongsan, Taehwa, Seomjin, and

34 Nakdong). Among the four rivers, the highest species richness was recorded in the Seomjin

35 River (52 species), followed by the Taehwa (42 species) and Hyeongsan (40 species) rivers. The

36 Nakdong River (26 species) presented the lowest species richness and number of endemic

37 species, presumably due to its metropolitan location and anthropogenic impacts, such as dams or

38 weirs. We were also able to detect that five exotic species (*Carassius cuvieri, Cyprinus carpio,*

39 *Cyprinus megalophthalmus, Lepomis macrochirus,* and *Micropterus salmoides*) are widely

40 distributed in all surveyed rivers, a situation that might be problematic in terms of conservation.

41 Our findings indicate that the eDNA metabarcoding technique is one of the most cost-

42 effective scientific tools available for the management and conservation of the freshwater fish

43 resources available in Korea. However, the low number of 12S sequences of endemic species in

44 the database and low resolution of the MiFish region for differentiating several taxa should be

45 upgraded for their wide use.

47 Keywords: biodiversity, Korea, next-generation sequencing, MiFish, metabarcoding, eDNA

49 **INTRODUCTION**

50   Fish communities have been considered as reliable bioindicators of ecosystem status due to their

51   vulnerability to environmental or anthropogenic stresses such as pollution, climate change, or

52   other disturbances in habitats (Dudgeon, 2010). Traditional monitoring methods for fish

53   biodiversity, which have relied on the direct capture or observation of specimens, are often

54   costly and time-consuming due to a lack of taxonomic expertise and the necessity of extensive

55   fieldwork. Environmental DNA (eDNA) metabarcoding (detection of multispecies by using

56   degraded DNA from environmental samples) has been proposed as an alternative strategy to

57   analyze fish biodiversity, demonstrating the potential to improve the traditional methods in a

58   cost-effective way (Foote et al., 2012; Kelly et al., 2017; Kelly et al., 2014; Shaw et al., 2016;

59   Stoeckle et al., 2017; Yamamoto et al., 2017).This technique has been shown to be sensitive as it

60   allows the identification of rarely identified (Pilliod et al., 2013), invasive (Ardura et al., 2015;

61   Cai et al., 2017; Clusa et al., 2017; Dejean et al., 2012; Klymus et al., 2017; Takahara et al., 2013;

62   Williams et al., 2018), or migratory species (Gustavson et al., 2015; Pont et al., 2018; Yamamoto

63   et al., 2016; Yamanaka and Minamoto, 2016).

64   Since eDNA metabarcoding analysis of fish biodiversity is mainly based on the amplicon of

65   homologous genes by PCR, universal primers with high taxon-specificity and wide taxon-

66   coverage are essential. Three fish-specific universal primer sets are currently reported: two sets

67   for 12S rRNA regions [EcoPrimers (Riaz et al., 2011) and MiFish (Miya et al., 2015b)] and one

68   for the 16S rRNA region (Shaw et al., 2016). Among them, the MiFish primer set demonstrated

69   reliability for eDNA metabarcoding analysis of fish biodiversity in both marine (Ushio et al.,

70   2017; Yamamoto et al., 2017) and continental waters (Sato et al., 2018). More recently, the web-

71   based MiFish pipeline in MitoFish was publicly open (http://mitofish.aori.u-tokyo.ac.jp/mifish/),

72   alleviating the time-consuming bioinformatic analysis for the users (Sato et al., 2018).

73   Although metabarcoding analysis by the MiFish pipeline is one of the most reliable tools at

74   the moment, numbers of MiFish sequences in the database are still one of the last hurdles to

75   overcome for the global use of the MiFish pipeline. Since the average length of the MiFish

76   region is approximately 170 bp, which is much smaller than the typically used 670 bp of the COI

77   barcodes, a high-quality database is critical for successful species assignment. Species

78   identification using the MiFish primer could not discriminate closely related species in several

79   genera, including *Sebastes* spp. and *Takifugu* spp. (Yamamoto et al., 2017). In particular,

80   considering the tremendous diversity of freshwater fishes, the direct application of the MiFish

81   platform may produce a high amount of 'unidentified' records. In addition, a relatively much

82   lower amount of MiFish sequence data (12S region) is currently deposited compared with those

83   of the COI region. Therefore, before the direct application of the MiFish pipeline, the MiFish

84   DNA sequence data for the local freshwater species should be tested for accurate fish

85   biodiversity analysis using eDNA metabarcoding.

86   In this study, we first employed eDNA metabarcoding analysis of water samples collected

87   from four rivers using the MiFish primer set in order to improve the knowledge regarding

88   freshwater fish biodiversity in Korea. Next, we analyzed the haplotypes obtained by the MiFish

89   pipeline to assess their compatibilities in the identification of endemic species of fishes

90   inhabiting Korean rivers. We also calculated the Shannon-Wiener (H') indices derived from the

91   eDNA metabarcoding results to estimate fish biodiversity in four Korean rivers. Finally, the

92   relationship between the fish assemblage according to the locations in the river was analyzed

93   using heat-map clustering analysis.

94

95   **MATERIALS AND METHODS**

96 **Sample collection and environmental DNA extraction**

97 The eDNA water samples were collected on June 11 and 12, 2018 from 16 stations in the

98 Hyeongsan, Taehwa, Seomjin, and Nakdong rivers, which are four large rivers in the southern

99 part of the Korean Peninsula (Fig.1 and Table 1). In this study, the sampling stations of each

100 river were categorized as upstream (stations 1 and 2), midstream (station 3), and downstream

101 (Station 4). One liter of water was collected at each station using disposable plastic bottles. After

102 collecting the water, the bottles were immediately stored in an icebox and taken to the laboratory

103 for filtration. Water temperature and salinity were measured with a conductivity meter (CD-

104 4307SD, LUTRON). The water collected was filtered (250 mL× 4) with a 0.45 μm pore-sized

105 GN-6 membrane (PALL Life Sciences, Mexico). The filtration system was cleaned with 10 %

106 commercial bleach containing sodium hypochlorite to prevent cross-contamination. After

107 filtration, the membranes were put into 2.0 ml tubes and stored at -20 °C before DNA

108 purification.

109   Genomic DNA was extracted directly from the membrane filters using the DNeasy® Blood

110 and Tissue Kit (Qiagen, Germany), according to the manufacturer's manual. The membrane

111 filters were cut into smaller pieces before homogenization using a TissueLyser II motorized

112 homogenizer (QIAGEN, Hilden, Germany). The extracted genomic DNA was quantified using a

113 ND-1000 NanoDrop (Thermo Scientific, Waltham, MA, USA), aliquoted, and stored at -20 °C.

114

115 **Construction of the library and MiSeq sequencing**

116 In order to assess the fish biodiversity, amplicon libraries of partial 12S rRNA region using the

117 MiFish universal primer sets were constructed (Miya et al., 2015a). The first PCR was performed

118 to amplify the MiFish regions with an overhanging linker sequence for each Nextera XT index

119   (Illumina, USA). The PCR mixture (20 µL) contained 1.0 µL of the MiFish (forward & reverse)

120   primers (5pmol each), 2.0 µL template, 2.0 µL dNTPs (2.5mM), 2.0 µL of 10X EX Taq buffer,

121   0.6 µL DMSO (3 %), 0.2 µL of EXTaq Hot Start polymerase (TaKaRa Bio Inc. Japan), and

122   11.20 µL ultra-pure water. The PCR reaction began with denaturation at 95 °C for 3 min;

123   followed by 30 cycles at 94 °C for 20 s, 65 °C for 15 s, and 72 °C for 15 s; and a final extension

124   at 72 °C for 5 min. The amplicon with the expected size (250–350 bp) was purified with the

125   AccuPrep® Gel Purification Kit (Bioneer, Republic of Korea) after 1.5 % agarose gel

126   electrophoresis. The purified amplicons were subjected to additional PCR to link each amplicon

127   with the corresponding Nextera XT index. The second PCR mixture (20 µL) contained 5 µL

128   template, 1 µL of a couple of index primers (10 pmol), 0.5 µL dNTPs (10 mM), 4 µL 5X

129   Phusion HF Buffer, 8.3 µL ultrapure water, and 0.2 µL Phusion Hot Start Flex DNA polymerase

130   (New England Biolabs, Hitchen, UK). The second PCR started at 94 °C for 5 min; followed by

131   15 cycles at 94 °C for 30 s, 55 °C for 30 s, and 72 °C for 30 s; and an additional 5 min at 72 °C.

132   No noticeable bands were detected in the desired ranges for 16 field negative controls in 1.5 %

133   agarose gel electrophoresis. Consequently, the 16 negative controls were discarded from the

134   following analyses. After gel purification, the quality and quantity of the indexed PCR products

135   with the expected sizes were analyzed using the Qubit dsDNAHS Assay Kit (Invitrogen,

136   Carlsbad, CA, USA), followed by sequencing using the MiSeq platform (2 × 300 bp).

137

138   **Bioinformatic analysis of the NGS data**

139   The MiSeq raw reads were paired using Python 2.7 (Zhang, 2015), and the paired reads were

140   uploaded to the MiFish pipeline (http://mitofish.aori.u-tokyo.ac.jp/mifish/) for further analyses.

141   In the MiFish pipeline, a low-quality tail of reads (QV ≤ 20) was trimmed in FASTQC. After

142 taxonomic assignments from the MiFish pipeline, the sequences assigned to OTUs were

143 compared with the GenBank database. If the sequence identity of the query sequence and top

144 BLASTN hit was ≥ 99 %, the sequence was ascertained as a particular species. If the sequence

145 identity ranged from 97 % to 99 %, the sequence was ascertained to the genus level, whereas

146 sequences ranging from 97 % to 95 % were assigned as 'unidentified' genera. The geographic

147 distribution of each species was assessed on the FishBase website (https://www.fishbase.org/).

148 Alpha biodiversity was measured using the normalized read numbers from each sampling station

149 of the four rivers sampled. The Shannon-Wiener (H') index indicates the heterogeneity of species

150 or the richness of species in an ecosystem (Gray, 2000; Magurran, 1988). The H′ index and the

151 heat map clustering analysis were calculated using the PRIMER® v7 software (Clarke and

152 Gorley, 2015).

153

154 **RESULTS**

155 **Physicochemical parameters**

156 The water temperature of the sample sites ranged from 18.6 °C to 24.20 °C (Table 1). The

157 Hyeongsan River showed the highest temperature difference (5.4 °C) between upstream (HS1) to

158 downstream (HS4), whereas the lowest levels of temperature variation were observed in the

159 Seomjin (0.8 °C) and Nakdong (1.5 °C) rivers. The lowest salinity (0.15 PSU) was measured at

160 station 1 (upstream) of the Seomjin River, while the highest (20.20 PSU) was recorded at station

161 4 (downstream) of the Hyeongsan River. The salinity level increased from upstream to

162 downstream in all rivers, except in the Nakdong River, where an artificial dam was constructed

163 to block water from the ocean (Table 1).

164

165 **Analysis of fish haplotypes obtained using the MiFish pipeline**

166 The reliability of the MiFish pipeline (http://mitofish.aori.u-tokyo.ac.jp/mifish/workflows/ new)

167 for the biodiversity assessment of fish species inhabiting the sampled rivers was analyzed (Table

168 2). From 2,315,605 raw reads, 2,280,850 merged reads were obtained by the MiFish pipeline,

169 with a 98.50 % yield from the raw reads. A total of 238 representative haplotypes were assigned

170 to the default cutoff sequence identity. Among the 238 haplotypes, 125 unique haplotypes were

171 identified using the phylogenetic tree analysis in the MEGA 7 software (Kumar et al., 2016) with

172 a maximum likelihood algorithm (Fig. 2-5). A total of 2,241,130 reads (98.26 %) were assigned

173 to 73 ~~confirmed~~ species, 46 genera, and 13 families of Teleostei, with 99 % as cutoff identity.

174 The remaining 39,720 reads (49 haplotypes), which showed less than 99 % identity, were further

175 assigned to11 genera and eight unidentified genera (Table 3). A total of 34,755 reads (1.50 %)

176 were discarded from further analyses. The highest species number was identified in the family

177 Cyprinidae (35), followed by Gobiidae (11), and Cobitidae (8), while the remaining (19) were

178 from other families of Teleostei. Among them, the highest species number (4 species) was

179 identified in the genus *Acheilognathus*, followed by *Carassius, Misgurnus, Squalidus,* and

180 *Tridentiger* with three species in each of those genera (Table S1).

181

182 **Cyprinidae**

183 A total of 65 haplotypes were identified in the ~~family~~ Cyprinidae. Among the 65 haplotypes, 51

184 were assigned to 35 species of fishes with ≥ 99 % of sequence identity to the GenBank database

185 (Fig. 2). Two haplotypes in the genus *Hemibarbus* from the Seomjin River (SJ1) and the

186 Nakdong River (ND2) showed 100 % and 99 % identity to the sequences of *Hemibarbus labeo*

187 (GenBank Number: DQ347953) and *Hemibarbus maculatus* (LC146032) sampled in Korea and

188   Japan, respectively. Among the four endemic species in the genus *Hemibarbus*, *H. labeo* and *H.*

189   *longirostris* are the most widely distributed species in Korea (Lee et al., 2012). Two haplotypes

190   identified from the Seomjin River (SJ1 and SJ2) and one from Taehwa River (TH1) showed 97 %

191   and 95 % identity to a sequence of *H. longirostris* (LC049889), respectively, which suggests that

192   these three haplotypes may be either *H. longirostris* or *H. mylodon* (Fig.2).

193   Five haplotypes were identified in the genus *Squalidus*. Four species of the genus have been

194   reported from Korean waters: *Squalidus gracilis*, *S. japonicus, S. multimaculatus,*

195   and *S. chankaensis* (Kim and Park, 2002). Two haplotypes from the Taehwa (TH3) and

196   Hyeongsan rivers (HS1) showed 100 % identity to sequences of *S. japonicas coreanus* (GenBank

197   Number: KR075134) and *S. multimaculatus* (GenBank Number: KT948081), respectively.

198   Another haplotype from the Hyeongsan River (HS3) showed 100 % identity to a sequence of *S.*

199   *japonicas* (GenBank Number: LC277782) sampled in Japan. Two haplotypes from the Seomjin

200   River showed 99 % identity to a sequence of *S. chankaensis tsuchigae* (GenBank Number:

201   KT948082) sampled in Korea.

202   Fishes of the subfamily Acheilognathinae, commonly known as bitterlings, deposit eggs in

203   the gill cavities of freshwater mussels (Kitamura, 2007; Kitamura et al., 2012). Approximately

204   60 species of bitterlings are considered valid in the genera *Acheilognathus*, *Tanakia*, and

205   *Rhodeus* (Arai, 1988). *Acheilognathus intermedia, A. macropterus, A. majusculus, A. rhombeus,*

206   *Rhodeus suigensis, R. uyekii, Tanakia somjinensis,* and *T. signifier* were identified with a

207   sequence identity > 99 % when compared to the GenBank database. Three haplotypes from the

208   Seomjin River showed 99 % sequence identity to the respective haplotypes of *A. intermedia*

209   (EF483933), *T. somjinensis* (FJ515921), and *T. signifier* (EF483930) sampled in Korea. Among

210   them, *T. somjinensis* and *T. signifier* are endemic to Korea (Kim and Park, 2002). One haplotype

211   from the Taehwa River (TH3) showed 100 % identity to a sequence of *Rhynchocypris semotilus*

212   (KT748874) sampled in Korea. This species is currently categorized as Critically Endangered in

213   the Red Data Book of endangered fishes in Korea (Ko et al., 2011).

214   Two sub-species of *Sarcocheilichthys* are known in Korea: *S. nigripinnis morii* and *S.*

215   *variegates wakiyae* (Kim and Park, 2002). Two haplotypes from the Seomjin (SJ2) and

216   Hyeongsan (HS2) rivers showed 100 % and 97 %, respectively, identity to a sequence of *S.*

217   *variegatus wakiyae* (GenBank Number: KU301744) sampled in Korea. One haplotype from the

218   Hyeongsan River (HS2) showed 100 % and 99.43 % identity to a sequence of *S. soldatovi*

219   (LC146036) and the Korean haplotype of *S. nigripinnis morii* (AP017653) sampled in Japan and

220   Korea, respectively. However, *S. soldatovi* is not currently reported for Korean waters. Therefore,

221   further studies are needed to confirm the occurrence of this species in the Hyeongsan River for

222   conservation purposes.

223

224   **Gobiidae**

225   We identified 16 haplotypes of the family Gobiidae, representing seven genera and 11 species

226   (Fig. 3). Five haplotypes were identified in the genus *Tridentiger*, which represents the five

227   known species of the genus recorded in Korea (Kim et al., 2005). One haplotype from the

228   Taehwa River (TH4) showed 100 % identity with a sequence of *T. obscures* (GenBank Number:

229   KT601092) sampled in Korea. One haplotype from the Hyeongsan River (HS4) showed 100 %

230   identity to a sequence of *T. trigonocephalus* (GenBank Number: LC385175) sampled in Japan,

231   and another haplotype from the Seomjin River (SJ3) showed 100 % identity to a sequence of *T.*

232   *trigonocephalus* (GenBank Number: KM030481) sampled in Korea. According to the recovered

233   phylogenetic tree, the *T. trigonocephalus* haplotype from the Seomjin River is different from that

234   of the Hyeongsan River (Fig. 3). All three haplotypes of the genus *Rhinogobius* showed 100 %

235   identity to the database. The first and second haplotypes showed 100 % identity to sequences of

236   *R. brunneus* sampled in Korea (KM030471) and Japan (LC049760), respectively. The third

237   haplotype showed 100 % identity to a sequence of *R. giurinus* sampled in Korea (KM030475).

238   Two haplotypes of *Gymnogobius* sp. from the Taehwa and Hyeongsan rivers showed 98 %

239   sequence identity to *G. taranetzi* (GenBank Number: LC385155). Nine species of the genus

240   *Gymnogobius* are currently reported in Korea (Kim et al., 2005), and their MiFish sequences

241   should be supplemented to the GenBank database.

242

243   **Cobitidae**

244   Sixteen species in five genera of the family Cobitidae are currently reported from Korean

245   rivers (Kim, 2009). A total of 18 haplotypes, representing five genera of the family, were

246   identified (Fig. 4). Two haplotypes in the genus *Cobitis* identified in the Seomjin River were

247   most closely related to *C. tetralineata* (LC146139) sampled in Japan, with 100 % and 99 %

248   sequence identity. Two haplotypes from the Taehwa River showed 98 % and 97 % identity to *C.*

249   *hankugensis* (LC146140). Two species of *Misgurnus* are reported from the Korean waters, *M.*

250   *mizolepis* and *M. anguillicaudatus* (Kim, 2009). Interestingly, two phylogenetically distinct

251   clades in *M. anguillicaudatus* were identified in the phylogenetic analysis (Fig. 4). One of them

252   was grouped with the haplotype of *M. bipartitus* (KF562047) sampled in China, while the other

253   was clustered with the *M. mizolepis* (AP017654) sampled in Korea. *Misgurnus bipartitus* is

254   currently reported ~~to be~~ endemic to China, and sequence data of Korean freshwater fishes in

255   GenBank data should be reexamined.

256   Two haplotypes from the Hyeongsan River (HS1; KJ699181) and the Taehwa River

257   (TH4; KM186182) showed 100% identity with haplotypes of *Paramisgurnus dabryanus* sampled

258   in China (Fig. 4). This species is regarded as endemic to China, but *P. dabryanus* is often

259   imported to Korea together with *Misgurnus anguillicaudatus* due to their phenotypic similarity.

260   Shimizu and Takagi (2010) concluded that there are different populations of *P. dabryanus*

261   (Shimizu and Takagi, 2010), and the two haplotypes of the species identified herein suggest that

262   *P. dabryanus* has been imported from various locations in China. One haplotype from the

263   Taehwa River (TH1) showed 100 % identity to a sequence of *Niwaella multifaciata* (EU670806)

264   sampled in Korea, while another from the Hyeongsan River (HS1) showed a lower (96 %)

265   identity to *Niwaella* sp. Therefore, further studies should be conducted to confirm the presence of

266   species of this genus in the Hyeongsan River.

267

268   **Other families of Teleostei**

269   In addition to the three main families of Teleostei identified in this study, 27 additional

270   haplotypes were found in the samples. These haplotypes represented 19 species belonging to14

271   genera and 11 families, namely Amblycipitidae (1), Anguillidae (1), Bagridae (5 ~~haplotypes~~),

272   Centrarchidae (3), Channidae (1), Clupeidae (2), Mugilidae (4), Odontobutidae (3),

273   Pleuronectidae (1), Siluridae (3), and Sinipercidae (3). All ~~the~~ haplotypes of the family Bagridae

274   were ~~clearly~~ identified ~~and included;~~ *Pseudobargrus ussuriensis*, *P. koreanus*, *Tachysurrus*

275   *nitidus*, and *T. fulvidraco* (Fig. 5). Two species of *Silurus* are currently known in Korean rivers,

276   *S. microdorsalis* and *S. asotus* (Park and Kim, 1994). One haplotype from the Taehwa River

277   (TH1) showed 99 % identity to a sequence of *Silurus microdorsalis* (GenBank Number:

278   KT350610) sampled in Korea, whereas another haplotype from the Seomjin River (SJ1) showed

279   a lower identity (96 %) with *S. microdorsalis* (KT350610) sampled in Korea.

280   One haplotype of the Amblycipitidae from the Seomjin River showed 97 % and 96 %

281   identity to *Liobagrus styani* (KX096605) and *L. mediadiposalis* (KR075136), sampled in China

282   and Korea, respectively. These results indicate that haplotypes of the family should be

283   supplemented for accurate identification. Three species of *Odontobutis* are currently known in

284   Korea: *O. interrupta, O. platycephala*, and *O. obscura* (Kim et al., 2005). Two of them (*O.*

285   *interrupta* and *O. platycephala*) were identified in this study with 100 % identity to the

286   sequences of *O. interrupta* and *O. platycephala* sampled in Korea (KR364945 and KM030426).

287   Two haplotypes of the genus *Coreoperca* showed 100 % and 97 % sequence identity to

288   *Coreoperca herzi* (KR075132) sampled in Korea. Since two species of *Coreoperca* are reported

289   to be endemic to the Korean Peninsula (Kim et al., 2005), the second haplotype is most likely *C.*

290   *kawamebari*, but further studies should be conducted to confirm this identification. Two invasive

291   species of the family Centrarchidae, the Bluegill (*Lepomis macrochirus*) and the Largemouth

292   bass (*Micropterus salmoides*) were also identified in this study. These two species are endemic

293   to North America but were introduced in the Korean Peninsula for aquaculture purposes without

294   considering their impact on local ecosystems.

295

296   **Fish biodiversity in the four rivers**

297   Fish assemblages in the four rivers included in this study were analyzed. Among the 73

298   confirmed fish species detected in this study, 13 were identified in all four rivers: *Anguilla*

299   *japonica, Hemibarbus labeo, Konosirus punctatus, Micropterus salmoides, Misgurnus mizolepis,*

300   *Mugil cephalus, Opsariichthys uncirostris, Pseudorasbora parva, Rhinogobius brunneus,*

301   *Rhynchocypris lagowskii, Silurus asotus, Tridentiger obscurus,* and *Zacco platypus* (Fig. 6).

302   Regardless of sample stations, species of the Cyprinidae appear to be dominant, with average

303   proportions of 47.02 ± 6.73 %, followed by the Gobiidae (15.24 ± 3.07 %) and Cobitidae (9.95 ±

304   4.09 %; Fig.7). However, the proportions of species in those families were different between

305   upstream and downstream stations. The proportion of Cyprinidae species was higher (45.27±

306   9.1%) upstream (stations 1 and 2) than downstream (33.78 ± 18 % at station 4). In contrast, the

307   proportion of Gobiidae was lower (14.53 ± 8.28 %) upstream than downstream (station 4, 19.90

308   ± 14 %).

309   The highest number of species was recorded in the Seomjin River (52 species), followed by

310   the Taehwa (42 ~~species~~), Hyeongsan (40 ~~species~~), and Nakdong (26 ~~species~~) rivers. A total of 17

311   species were exclusively recorded in the Seomjin River: *Acanthogobius hasta, Acheilognathus*

312   *intermedia, A. majusculus, A. rhombeus, Cobitis tetralineata, Coreoleuciscus splendidus,*

313   *Kareius bicoloratus, Microphysogobio yaluensis, Phoxinus oxycephalus, Pseudobagrus*

314   *koreanus, Rhodeus suigensis, R. uyekii, Sarcocheilichthys variegatus, Siniperca scherzeri,*

315   *Squalidus gracilis, Tanakia somjinensis,* and *T. signifier.* Five species were only recorded in the

316   Taehwa River: *Acanthogobius lactipes,Mugilogobius abei, Pseudogobius masago,*

317   *Rhynchocypris semotilus,* and *Silurus microdorsalis,* whereas four species were ~~only identified~~ in

318   the Nakdong River: *Plagiognathops microlepis, Pseudobagrus ussuriensis, Rhinogobius giurinus,*

319   and *Tachysurus nitidus.* Finally, only three species (*Nipponocypris koreanus, Sarcocheilichthys*

320   *soldatovi,* and *Squalidus multimaculatus*) were exclusively recorded in the Hyeongsan River (Fig.

321   6).

322   The highest Shannon index (SI) was identified in the Seomjin River (3.480), followed by the

323   Taehwa (3.067), Hyeongsan (2.954), and Nakdong (2.864) rivers. Among the 16 surveyed

324   stations, station 1 of the Seomjin River (SJ1) showed the highest species richness (2.197),

325   whereas the lowest richness (1.008) was recorded atthe station 4 of the Nakdong River (ND4).

326   From upstream to downstream, average species richness decreased from 1.951 to 1.415 (Table 4).

327

328   **Clustering analysis**

329   In order to assess the correlation between the fish assemblage and sample stations, we conducted

330   a heat-map analysis with the 30 most abundant species using Primer software (Clarke and Gorley,

331   2015). The results indicate the species distribution in different sampling stations (Fig. 8). In

332   upstream sites (Stations 1 and 2), the dominant species were *A. intermedia, Coreoperca herzi,*

333   *Misgurnus mizolepis, Nipponocypris temminckii, Rhynchocypris lagowskii, Odontobutis*

334   *interrupta, O. platycephala, Tanakia signifier,* and *Zacco platypus.* At station 3, the dominant

335   species were *Gymnogobius breunigii, Mugil cephalus, Pseudorasbora parva, Rhinogobius*

336   *giurinus,* and *R. brunneus.* Finally, in the downstream sample (Station 4), *Anguilla japonica,*

337   *Konosirus punctatus, Mugil cephalus, Planiliza haematocheila, Tridentiger obscurus,* and *T.*

338   *trigonocephalus* were identified as the dominant species, all of which were either euryhaline or

339   anadromous (https://www.fishbase.org).

340

341   **DISCUSSION**

342   The results indicate that eDNA metabarcoding using the MiFish pipeline is a useful tool for

343   assessing fish biodiversity in Korean freshwater ecosystems, since a total of 125 unique

344   haplotypes, including at least 73 species, were successfully identified by a single-day survey of

345   16 sampling stations in four rivers (Fig. 2-5). According to the "Survey and Evaluation of

346   Aquatic Ecosystem Health (SEAEH)," a total of 130 freshwater ~~fish species~~ were identified from

347   953 sampling sites that covered most of the Korean rivers and lakes (Yoon et al., 2012). The total

348   number of species ~~confirmed~~ by eDNA metabarcoding was equivalent to approximately 56 % of

349   those obtained by the year-long conventional surveys. The efficiency of eDNA metabarcoding

350   might actually be even higher, especially considering the number of haplotypes successfully

351   identified at the genus and/or family level. This result indicates that eDNA metabarcoding with

352   the MiFish pipeline can significantly contribute to the assessment of freshwater fish biodiversity

353   in Korea, especially considering its relatively lower cost of implementation when compared with

354   more conventional morphology-based surveys. Although the methodology in each research

355   group may be slightly different, similar conclusions have been reached in other studies (Bista et

356   al., 2017; Deiner et al., 2016). eDNA metabarcoding analysis is also adequate for surveying

357   aquatic species in protected areas, as it minimizes disturbance of vulnerable communities

358   (Fernandez et al., 2018).

359   Despite its relevance as a methodology for the assessment of biodiversity, there are still a

360   few shortcomings for a more widespread use of eDNA metabarcoding by the MiFish pipeline.

361   First, MiFish sequence data for endemic species of Korea should be supplemented to the

362   GenBank database. According to the Archive of Korean species (https://species.nibr.go.kr), 67

363   species of freshwater fishes are endemic to Korea, and many of their MiFish sequences are still

364   not available in the GenBank database. In addition to the lack of sequence data, freshwater fishes

365   typically have intra-species genetic distances that are generally higher than those of marine

366   species (Seehausen and Wagner, 2014). Second, the MiFish primer amplifies the 12S rRNA gene

367   (163–185 bp) region of mitochondrial DNA, which is smaller and less variable than the COI

368   region, which is typically used in species identification (IVANOVA et al., 2007). In fact, the

369   MiFish region was unable to differentiate several closely related marine fish taxa, such as

370   those in the genus *Sebastes* and *Takifugu* (Sato et al., 2018; Yamamoto et al., 2017). We also

371   found that the average genetic distance of several genera in the family Cyprinidae was low in the

372   MiFish region. For example, the average genetic distance of *Carassius* species was too low (0.01)

373   and the identification at the species level was not possible (Fig. 2).

374   Further studies using eDNA metabarcoding might also be relevant to obtain more than

375   biodiversity data, such as the quantitative analysis of fish species. It is difficult to estimate the

376   spatial abundance of eDNA in lotic environments. In fact, many factors should be considered for

377   the quantitative analysis of eDNAs in rivers, including water dynamics (Deiner and Altermatt,

378   2014; Jerde et al., 2016; Wilcox et al., 2016) or different decaying times due to different physical,

379   chemical, or biological factors (Shapiro, 2008). ~~It is generally known that~~ ~~shorter~~ fragments of

380   DNA are degraded slower than larger ones, increasing their probability of detection in natural

381   environments (Deagle et al., 2006). Therefore, it is still too early to adopt eDNA metabarcoding

382   for the quantitative analysis of fish species under natural conditions. For the quantitative study,

383   standardized collection methods and pretreatment procedures for NGS sequencing analysis

384   should also be established. One of the strongest points in the biodiversity survey by eDNA

385   metabarcoding is the quantity of information it can generate compared with more conventional

386   surveys since large datasets are useful for statistical analyses. However, large amounts of data

387   have been produced using different water collection methods, eDNA preparation, sequencing,

388   and bioinformatic analysis platforms by different research groups in different countries.

389   Therefore, the interconversion of data is currently not possible. The establishment of an

390   international standard regarding the overall methodology of eDNA metabarcoding would help

391   researchers to produce more comparable data.

392   According to the results obtained in this study, the highest species richness was found in the

393   Seomjin River (3.48) compared with those of the other three rivers: the Taehwa River (3.06),

394   Hyeongsan River (2.95), and Nakdong River (2.86). The lower values of species richness

395   detected in the Nakdong, Hyeongsan, and Taehwa rivers are presumably related to the higher

396   anthropogenic alteration of the natural conditions in those rivers. Like most other Korean rivers,

397   these three rivers run through highly populated metropolitan cities, in which rivers are exposed

398   to various human impacts that directly or indirectly promote changes in the diversity and

399   distribution of freshwater fishes (Finkenbine et al., 2000). In particular, the lowest species

400   richness (2.86) and number of endemic species (only one, *Odontobutis interrupta*) were

401   identified in the Nakdong River, where the highest number of constructions and population exist

402   among the sampled rivers. Lee et al. (2015) reported only two endemic species (*Coreoperca*

403   *herzi* and *Odontobutis platycephala*) in the Nakdong River using a conventional catch survey.

404   Moreover, eight endemic species(*Coreoleuciscus splendidus, Iksookimia longicorpa,*

405   *Microphysogobio koreensis, M. yaluensis, Odontobutis interrupta, O. platycephala,*

406   *Pseudobagrus koreanus,* and *Squalidus gracilis*) were identified in this study in the Seomjin

407   River, a number that is similar to those obtained in previous studies (Jang et al., 2003; Lee et al.,

408   2015). Several constructions along urbanized watersheds, including dams and weirs, have caused

409   the simplification and reduction of habitats, decreasing the biodiversity in the river (Nilsson et al.,

410   2005; Riley et al., 2005). In contrast, there is no metropolitan city along the Seomjin River,

411   which is, therefore, less exposed to anthropogenic impacts. A long-term survey should be

412   conducted to establish a clear correlation between anthropogenic factors and fish assemblages in

413   the Korean rivers.

414   The eDNA metabarcoding analysis also indicates that some exotic fish species are widely

415   distributed in Korean rivers. We were able to identify at least five exotic ~~fish~~ species: *Carassius*

416   *cuvieri, Cyprinus carpio, C. megalophthalmus, Lepomis macrochirus*, and *Micropterus*

417   *salmoides* (Table S3). These exotic species may affect native fishes in terms of shelter and

418   spawning sites. They can also disturb the food chain, preying on native ~~fish,~~ In addition, these

419   species have a high reproductive capacity, which makes them important potentially invasive

420   species (Keller & Lake, 2007; Koster *et al.*, 2002; Nico & Fuller, 2010). Surprisingly, our results

421   also revealed that the largemouth bass, *M. salmoides*, and the bluegill, *L. macrochirus*, are likely

422   present in all the sampled rivers. These two species, which are native to North America, were

423   artificially introduced in the 1970s in Korea as freshwater fish stock, without any further

424   consideration of the effects on the freshwater ecosystems of the country. They are now widely

425   distributed throughout the Korean Peninsula, competing with the native species. A long-term

426   survey of these rivers should be conducted to properly assess the potential impacts of these

427   introduced species (Jang et al., 2002; Yoon et al., 2012). Freshwater ecosystems are much more

428   vulnerable to invasive species, causing biodiversity loss and global climate change (Clavero and

429   García-Berthou, 2005), and eDNA metabarcoding analyses would be useful for monitoring the

430   distribution patterns of invasive species in Korean rivers.

431

Acknowledgments

432

433   The authors are thankful to the Ministry of Oceans and Fisheries of the Republic of Korea. The

434   authors also thank the reviewers for their valuable comments and suggestions for the manuscript.

435

Additional information and declarations

436

437 **Funding:**

438 This research was a part of the project titled "Long-term change of structure and function in

439 marine ecosystems of Korea", funded by the Ministry of Oceans and Fisheries, Korea.

440

441 **Role of funding**

442 The funding sources had no role in the research design, sample collection, data analysis,

443 manuscript writing, or the decision to submit the article for publication.

444

445 **Competing Interests**

446 The authors declare that they have no competing interests.

447

448 **Author Contributions**

449 • Md. Jobaidul Alam collected the samples, performed the experiments, analyzed the data, prepared

450   figures and/or tables, and wrote the manuscript.

451 • Nack-Keun Kim collected the samples and analyzed the data

452 • Sapto Andriyono performed the experiments, analyzed the data, prepared figures, and/or tables

453 • Hee-Kyu Choi analyzed the data, prepared figures, and/or tables

454 • Ji-Hyun Lee analyzed the data, prepared figures, and/or tables

455 • Hyun-Woo Kim conceived and designed the experiments, analyzed the data, contributed

456   reagents/materials/analysis tools, authored or reviewed drafts of the manuscript, approved the

457   final draft, and wrote the manuscript.

458

459

460

461 **REFERENCES**

462 Arai, R., 1988. Acheilognathus melanogaster, a senior synonym of A moriokae, with a revision of the
463 genera of the subfamily Acheilogathinae (Cypriniformes, Cyprinidae). Bul Nat Sci Mus Tokyo Ser A 14,
464 199-213.

465 Ardura, A., Zaiko, A., Martinez, J.L., Samulioviene, A., Semenova, A., Garcia-Vazquez, E., 2015. eDNA
466 and specific primers for early detection of invasive species–a case study on the bivalve Rangia cuneata,
467 currently spreading in Europe. Marine Environmental Research 112, 48-55.

468 Bista, I., Carvalho, G.R., Walsh, K., Seymour, M., Hajibabaei, M., Lallias, D., Christmas, M., Creer, S.,
469 2017. Annual time-series analysis of aqueous eDNA reveals ecologically relevant dynamics of lake
470 ecosystem biodiversity. Nature communications 8, 14087.

471 Cai, W., Ma, Z., Yang, C., Wang, L., Wang, W., Zhao, G., Geng, Y., Douglas, W.Y., 2017. Using eDNA
472 to detect the distribution and density of invasive crayfish in the Honghe-Hani rice terrace World Heritage
473 site. PloS one 12, e0177724.

474 Clarke, K., Gorley, R., 2015. Getting started with PRIMER v7. PRIMER-E: Plymouth, Plymouth Marine
475 Laboratory.

476 Clavero, M., García-Berthou, E., 2005. Invasive species are a leading cause of animal extinctions. Trends
477 in Ecology & Evolution 20, 110.

478 Clusa, L., Miralles, L., Basanta, A., Escot, C., García-Vázquez, E., 2017. eDNA for detection of five
479 highly invasive molluscs. A case study in urban rivers from the Iberian Peninsula. PloS one 12, e0188126.

480 Deagle, B.E., Eveson, J.P., Jarman, S.N., 2006. Quantification of damage in DNA recovered from highly
481 degraded samples–a case study on DNA in faeces. Frontiers in zoology 3, 11.

482 Deiner, K., Altermatt, F., 2014. Transport distance of invertebrate environmental DNA in a natural river.
483 PloS one 9, e88786.

484 Deiner, K., Fronhofer, E.A., Mächler, E., Walser, J.-C., Altermatt, F., 2016. Environmental DNA reveals
485 that rivers are conveyer belts of biodiversity information. Nature communications 7, 12544.

486 Dejean, T., Valentini, A., Miquel, C., Taberlet, P., Bellemain, E., Miaud, C., 2012. Improved detection of
487 an alien invasive species through environmental DNA barcoding: the example of the American bullfrog
488 Lithobates catesbeianus. Journal of applied ecology 49, 953-959.

489 Dudgeon, D., 2010. Prospects for sustaining freshwater biodiversity in the 21st century: linking
490 ecosystem structure and function. Current Opinion in Environmental Sustainability 2, 422-430.

491 Fernandez, S., Sandin, M.M., Beaulieu, P.G., Clusa, L., Martinez, J.L., Ardura, A., García-Vázquez, E.,
492 2018. Environmental DNA for freshwater fish monitoring: insights for conservation within a protected
493 area. PeerJ 6, e4486.

494 Finkenbine, J.K., Atwater, J., Mavinic, D., 2000. STREAM HEALTH AFTER URBANIZATION 1.
495 JAWRA Journal of the American Water Resources Association 36, 1149-1160.

496 Foote, A.D., Thomsen, P.F., Sveegaard, S., Wahlberg, M., Kielgast, J., Kyhn, L.A., Salling, A.B.,
497 Galatius, A., Orlando, L., Gilbert, M.T.P., 2012. Investigating the potential use of environmental DNA
498 (eDNA) for genetic monitoring of marine mammals. PloS one 7, e41781.

499 Gray, J.S., 2000. The measurement of marine species diversity, with an application to the benthic fauna of
500 the Norwegian continental shelf. Journal of experimental marine biology and ecology 250, 23-49.

501 Gustavson, M., Collins, P., Finarelli, J., Egan, D., Conchúir, R., Wightman, G., King, J., Gauthier, D.,
502 Whelan, K., Carlsson, J., 2015. An eDNA assay for Irish Petromyzon marinus and Salmo trutta and field
503 validation in running water. Journal of Fish Biology 87, 1254-1262.

504 IVANOVA, N.V., ZEMLAK, T.S., HANNER, R.H., HEBERT, P.D.N., 2007. Universal primer cocktails
505 for fish DNA barcoding. Molecular Ecology Notes 7, 544-548.

506 Jang, M.-H., Lucas, M.C., Joo, G.-J., 2003. The fish fauna of mountain streams in South Korean national
507 parks and its significance to conservation of regional freshwater fish biodiversity. Biological
508 Conservation 114, 115-126.

509 Jang, M.H., Kim, J.G., Park, S.B., Jeong, K.S., Cho, G.I., Joo, G.J., 2002. The current status of the
510 distribution of introduced fish in large river systems of South Korea. International Review of
511 Hydrobiology 87, 319-328.

512 Jerde, C.L., Olds, B.P., Shogren, A.J., Andruszkiewicz, E.A., Mahon, A.R., Bolster, D., Tank, J.L., 2016.
513 Influence of stream bottom substrate on retention and transport of vertebrate environmental DNA.
514 Environmental science & technology 50, 8770-8779.

515 Kelly, R.P., Closek, C.J., O'Donnell, J.L., Kralj, J.E., Shelton, A.O., Samhouri, J.F., 2017. Genetic and
516 manual survey methods yield different and complementary views of an ecosystem. Frontiers in Marine
517 Science 3, 283.

518 Kelly, R.P., Port, J.A., Yamahara, K.M., Martone, R.G., Lowell, N., Thomsen, P.F., Mach, M.E., Bennett,
519 M., Prahler, E., Caldwell, M.R., 2014. Harnessing DNA to improve environmental management. Science
520 344, 1455-1456.

521 Kim, I.-S., 2009. A Review of the Spined Loaches, Family Cobitidae (Cypriniformes) in Korea Korean
522 Journal of Ichthyology 21, 7-28.

523 Kim, I.-S., Park, J.-Y., 2002. Freshwater fishes of Korea. Kyo hak sa.

524 Kim, I., Choi, Y., Lee, C., Lee, Y., Kim, B., Kim, J., 2005. Illustrated book of Korean fishes. Kyo-Hak
525 Publishing, 417-418.

526 Kitamura, J.-i., 2007. Reproductive ecology and host utilization of four sympatric bitterling
527 (Acheilognathinae, Cyprinidae) in a lowland reach of the Harai River in Mie, Japan. Environmental
528 Biology of Fishes 78, 37-55.

529 Kitamura, J., Nagata, N., Nakajima, J., Sota, T., 2012. Divergence of ovipositor length and egg shape in a
530 brood parasitic bitterling fish through the use of different mussel hosts. Journal of evolutionary biology
531 25, 566-573.

532 Klymus, K.E., Marshall, N.T., Stepien, C.A., 2017. Environmental DNA (eDNA) metabarcoding assays
533 to detect invasive invertebrate species in the Great Lakes. PloS one 12, e0177643.

534 Ko, M., Kim, K., Park, J., 2011. Red Data Book of endangered fishes in Korea. National Institute of
535 Biological Resources, Incheon.(in Korean).

536 Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: molecular evolutionary genetics analysis version 7.0
537 for bigger datasets. Molecular biology and evolution 33, 1870-1874.

538 Lee, J.W., Yoon, J.D., Kim, J.H., Park, S.H., Baek, S.H., Yu, J.J., Jang, M.H., Min, J.I., 2015.
539 Length-weight relationships for 18 freshwater fish species from the Nakdong River in South Korea.
540 Journal of Applied Ichthyology 31, 576-577.

541   Lee, W.-O., Zhang, M.-M., Oh, C.-W., Baek, J.-M., Song, K.-J., 2012. Age and Growth of Barbel Steed
542   Hemibarbus labeo in Goe-san Lake in Korea. Fisheries and aquatic sciences 15, 353-359.

543   Magurran, A.E., 1988. Ecological diversity and its measurement. Princeton university press.

544   Miya, M., Sato, Y., Fukunaga, T., Sado, T., Poulsen, J.Y., Sato, K., Minamoto, T., Yamamoto, S.,
545   Yamanaka, H., Araki, H., 2015a. MiFish, a set of universal PCR primers for metabarcoding
546   environmental DNA from fishes: detection of more than 230 subtropical marine species. Royal Society
547   open science 2, 150088.

548   Miya, M., Sato, Y., Fukunaga, T., Sado, T., Poulsen, J.Y., Sato, K., Minamoto, T., Yamamoto, S.,
549   Yamanaka, H., Araki, H., Kondoh, M., Iwasaki, W., 2015b. MiFish, a set of universal PCR primers for
550   metabarcoding environmental DNA from fishes: detection of more than 230 subtropical marine species.
551   Royal Society Open Science 2.

552   Nilsson, C., Reidy, C.A., Dynesius, M., Revenga, C., 2005. Fragmentation and flow regulation of the
553   world's large river systems. Science 308, 405-408.

554   Park, S.-W., Kim, Y.-G., 1994. Studies on disease of catfish, Silurus asotus, in Korea. III. Edwardsiella
555   ictaluri infection. Journal of fish pathology 7, 105-112.

556   Pilliod, D.S., Goldberg, C.S., Laramie, M.B., Waits, L.P., 2013. Application of environmental DNA for
557   inventory and monitoring of aquatic species. US Department of the Interior, US Geological Survey.

558   Pont, D., Rocle, M., Valentini, A., Civade, R., Jean, P., Maire, A., Roset, N., Schabuss, M., Zornig, H.,
559   Dejean, T., 2018. Environmental DNA reveals quantitative patterns of fish biodiversity in large rivers
560   despite its downstream transportation. Scientific reports 8, 10361.

561   Riaz, T., Shehzad, W., Viari, A., Pompanon, F., Taberlet, P., Coissac, E., 2011. ecoPrimers: inference of
562   new DNA barcode markers from whole genome sequence analysis. Nucleic Acids Research 39, e145-
563   e145.

564   Riley, S.P., Busteed, G.T., Kats, L.B., Vandergon, T.L., Lee, L.F., Dagit, R.G., Kerby, J.L., Fisher, R.N.,
565   Sauvajot, R.M., 2005. Effects of urbanization on the distribution and abundance of amphibians and
566   invasive species in southern California streams. Conservation Biology 19, 1894-1907.

567   Sato, Y., Miya, M., Fukunaga, T., Sado, T., Iwasaki, W., 2018. MitoFish and MiFish Pipeline: A
568   Mitochondrial Genome Database of Fish with an Analysis Pipeline for Environmental DNA
569   Metabarcoding. Molecular Biology and Evolution 35, 1553-1555.

570   Seehausen, O., Wagner, C.E., 2014. Speciation in Freshwater Fishes. Annual Review of Ecology,
571   Evolution, and Systematics 45, 621-651.

572   Shapiro, B., 2008. Engineered polymerases amplify the potential of ancient DNA. Trends in
573   biotechnology 26, 285-287.

574   Shaw, J.L.A., Clarke, L.J., Wedderburn, S.D., Barnes, T.C., Weyrich, L.S., Cooper, A., 2016.
575   Comparison of environmental DNA metabarcoding and conventional fish survey methods in a river
576   system. Biological Conservation 197, 131-138.

577   Shimizu, T., Takagi, M., 2010. Two genetic clades in populations of Paramisgurnus dabryanus, an exotic
578   invader in ehime prefecture. Jpn J Ichthyol 57, 125-134.

579   Stoeckle, M.Y., Soboleva, L., Charlop-Powers, Z., 2017. Aquatic environmental DNA detects seasonal
580   fish abundance and habitat preference in an urban estuary. PLOS ONE 12, e0175186.

581   Takahara, T., Minamoto, T., Doi, H., 2013. Using environmental DNA to estimate the distribution of an
582   invasive fish species in ponds. PloS one 8, e56584.

583   Ushio, M., Murakami, H., Masuda, R., Sado, T., Miya, M., Sakurai, S., Yamanaka, H., Minamoto, T.,
584   Kondoh, M., 2017. Quantitative monitoring of multispecies fish environmental DNA using high-
585   throughput sequencing. bioRxiv.

586   Wilcox, T.M., McKelvey, K.S., Young, M.K., Sepulveda, A.J., Shepard, B.B., Jane, S.F., Whiteley, A.R.,
587   Lowe, W.H., Schwartz, M.K., 2016. Understanding environmental DNA detection probabilities: A case
588   study using a stream-dwelling char Salvelinus fontinalis. Biological Conservation 194, 209-216.

589   Williams, K.E., Huyvaert, K.P., Vercauteren, K.C., Davis, A.J., Piaggio, A.J., 2018. Detection and
590   persistence of environmental DNA from an invasive, terrestrial mammal. Ecology and evolution 8, 688-
591   695.

592   Yamamoto, S., Masuda, R., Sato, Y., Sado, T., Araki, H., Kondoh, M., Minamoto, T., Miya, M., 2017.
593   Environmental DNA metabarcoding reveals local fish communities in a species-rich coastal sea.
594   Scientific reports 7, 40368.

595   Yamamoto, S., Minami, K., Fukaya, K., Takahashi, K., Sawada, H., Murakami, H., Tsuji, S., Hashizume,
596   H., Kubonaga, S., Horiuchi, T., 2016. Environmental DNA as a 'snapshot'of fish distribution: A case
597   study of Japanese jack mackerel in Maizuru Bay, Sea of Japan. PLoS One 11, e0149786.

598   Yamanaka, H., Minamoto, T., 2016. The use of environmental DNA of fishes as an efficient method of
599   determining habitat connectivity. Ecological indicators 62, 147-153.

600   Yoon, J.-D., Jang, M.-H., Kim, H.-W., Joo, G.-J., 2012. Fish Biodiversity Monitoring in Rivers of South
601   Korea, The Biodiversity Observation Network in the Asia-Pacific Region. Springer, 175-191.

602   Zhang, Y., 2015. An Introduction to Python and computer programming, An Introduction to Python and
603   Computer Programming. Springer, 1-11.

604

Figure 1

Water sample collection sites of four Korean rivers

Figure 1 Water sample collection sites for environmental DNA metabarcoding study from four
Korean rivers

Figure 2

Phylogenetic tree of ~~the fish~~ species ~~under~~ the family Cyprinidae

Figure 2 Phylogenetic tree ~~analysis~~ of ~~fish~~ species ~~under~~ the family Cyprinidae ~~detected~~ ~~from~~ four Korean rivers. Phylogenetic tree was constructed by Maximum likelihood (ML) algorithm (MEGA 7.0) under ~~the~~ 1000 ~~replication~~ bootstrap.
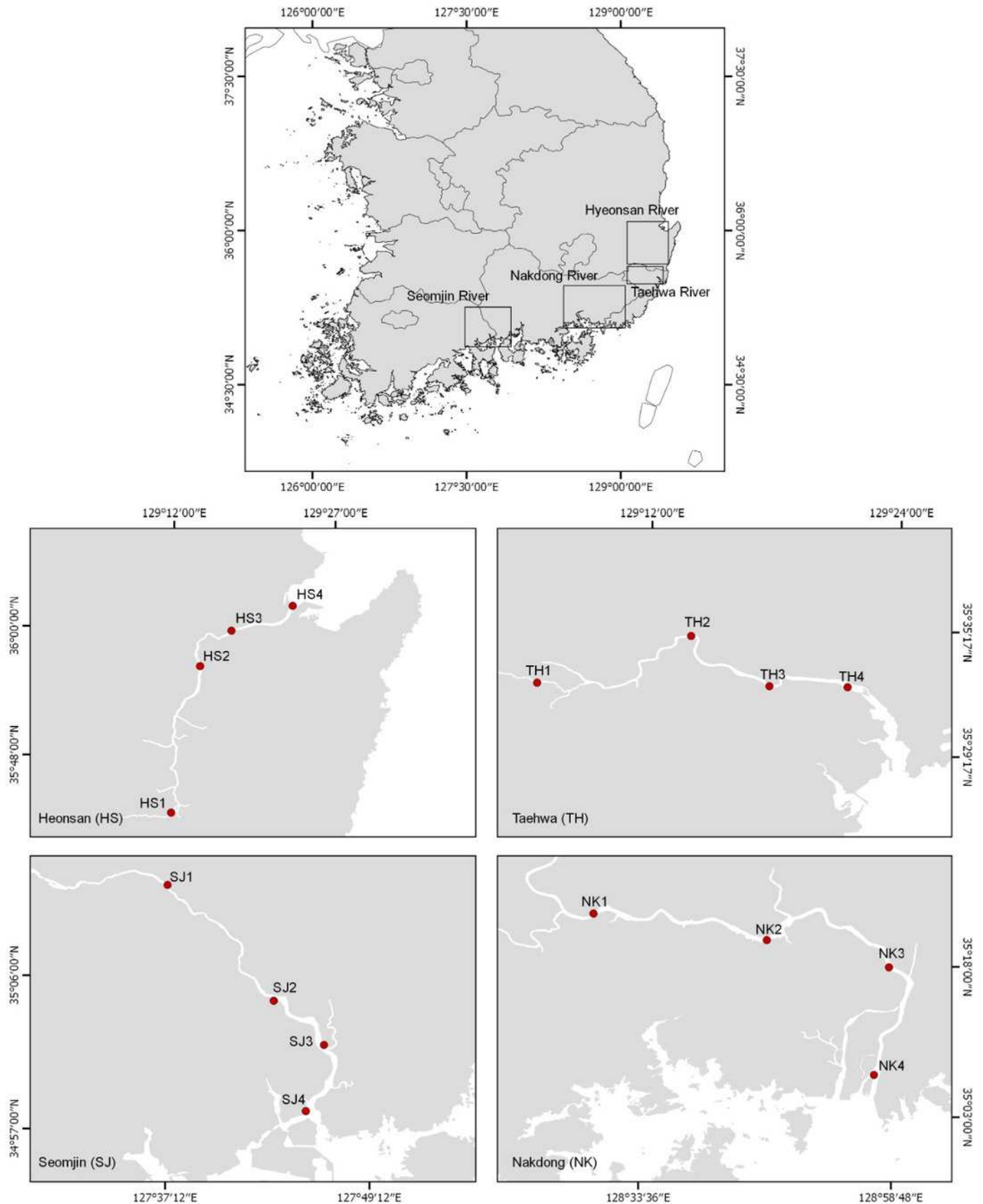
Figure 3

Phylogenetic tree of the fish species under the family Gobiidae

Figure 3 Phylogenetic tree ~~analysis~~ of ~~fish~~ species ~~under~~ the family Gobiidae. Phylogenetic tree was constructed by Maximum likelihood (ML) algorithm (MEGA 7.0) under ~~the~~ 1000 ~~replication~~ bootstrap.
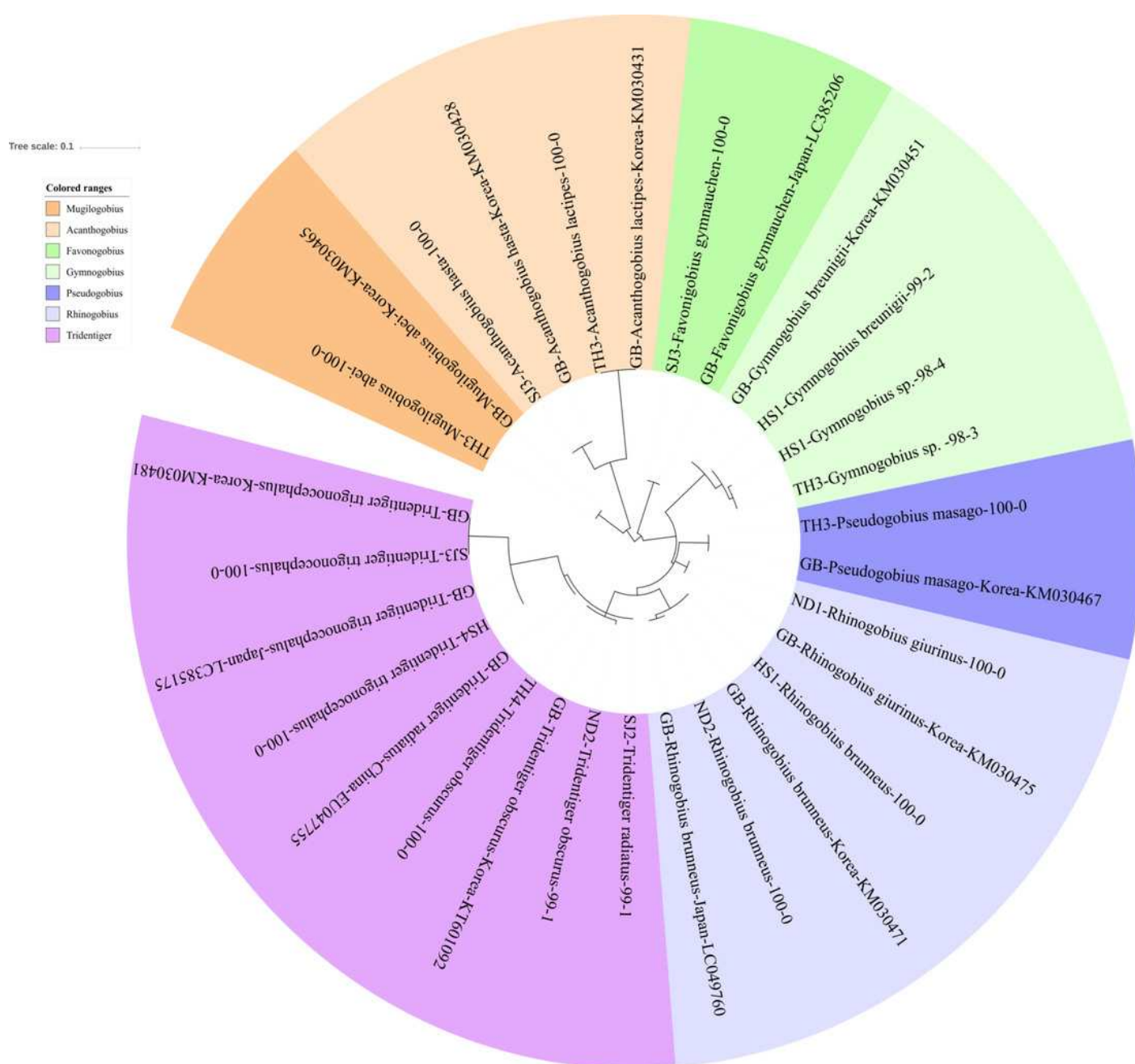
Figure 4

Phylogenetic tree of the fish species under the family Cobitidae

Figure 4 Phylogenetic tree analysis of fish species under the family Cobitidae. Phylogenetic tree was constructed by Maximum likelihood (ML) algorithm (MEGA 7.0) under the 1000 replication bootstrap.
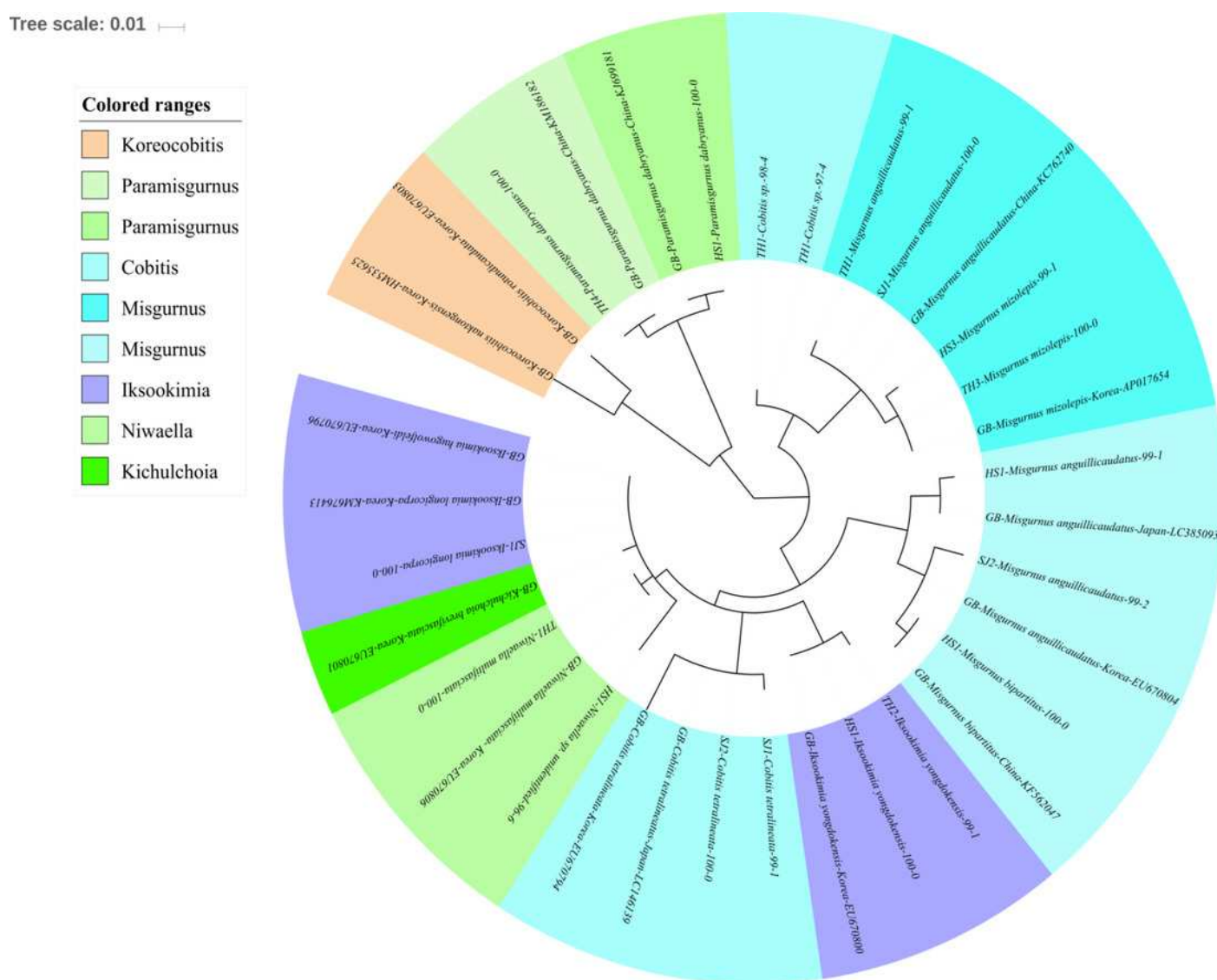
Figure 5

Phylogenetic tree of the fish species under the other families

Figure 5 Phylogenetic tree analysis of fish species under the other families of Teleostei. Phylogenetic tree was constructed by Maximum likelihood (ML) algorithm (MEGA 7.0) under the 1000 replication bootstrap.
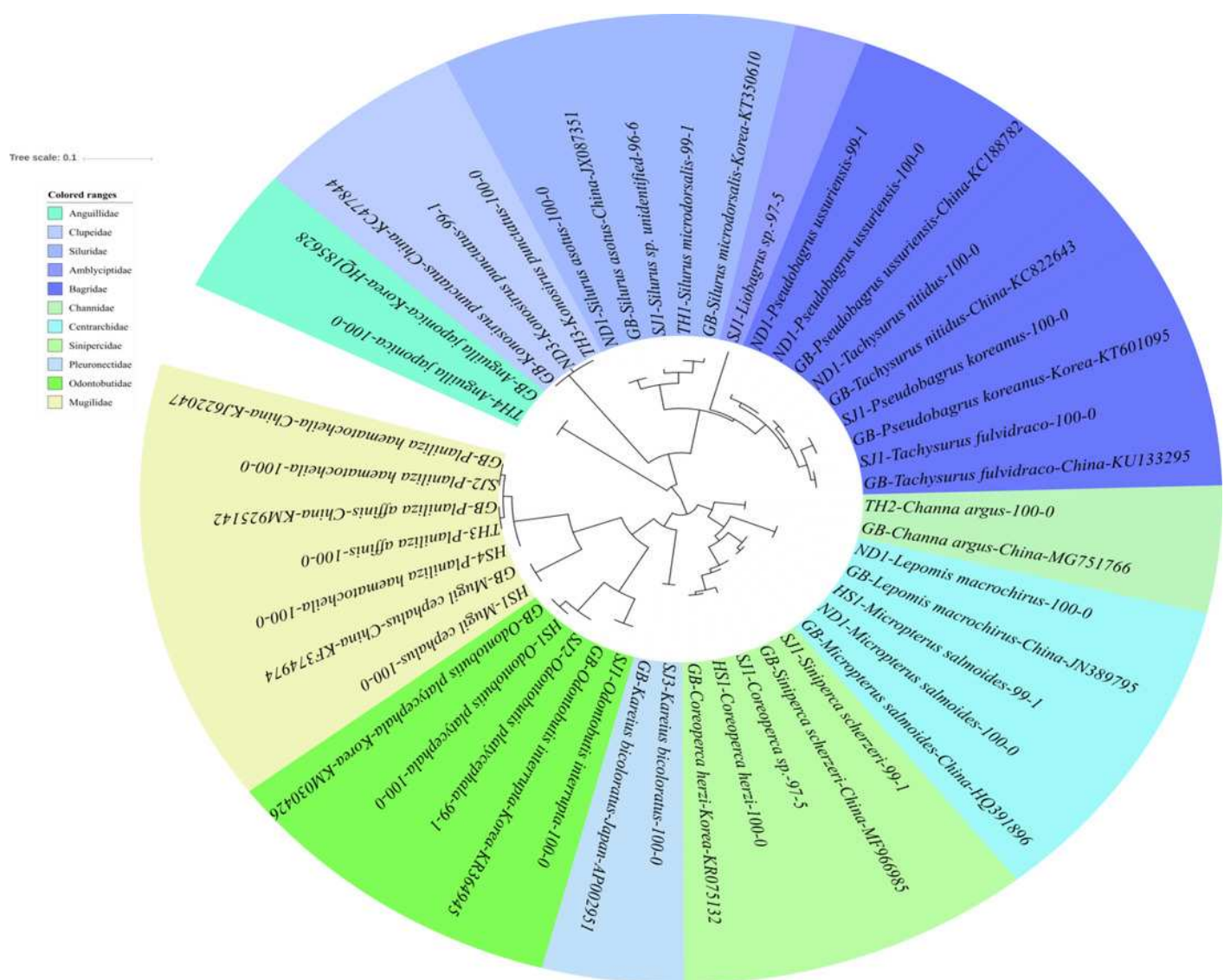
Figure 6

Venn diagram of fish species identified in the four Korean rivers.

Figure 6 Venn diagram of identified species of fishes in the four Korean rivers. Venn diagram was constructed by an online program (http://bioinformatics.psb.ugent.be/webtools/Venn/).
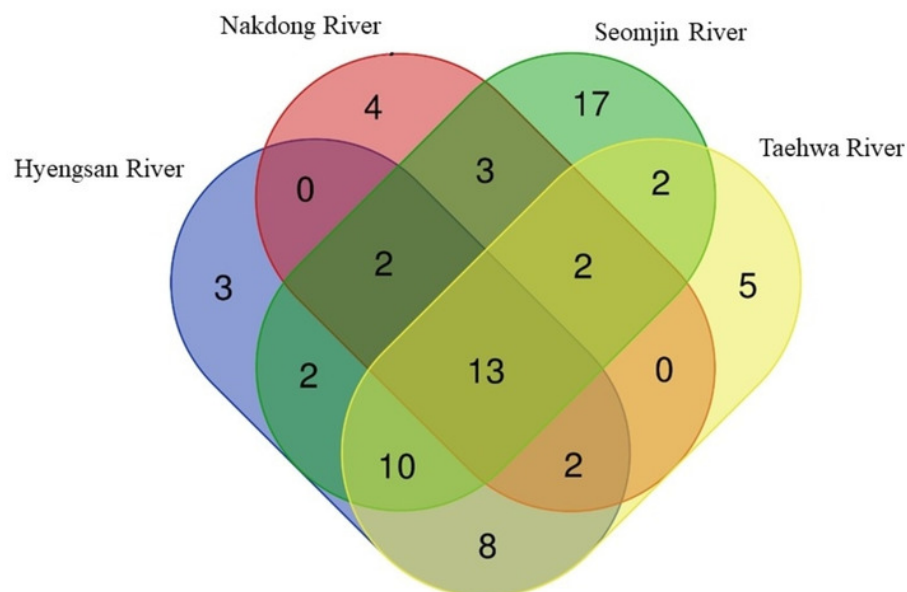
Figure 7

Proportion of families detected from the four Korean rivers

Figure 7 Proportion of families detected ~~from the~~ four Korean rivers by environmental DNA metabarcoding.
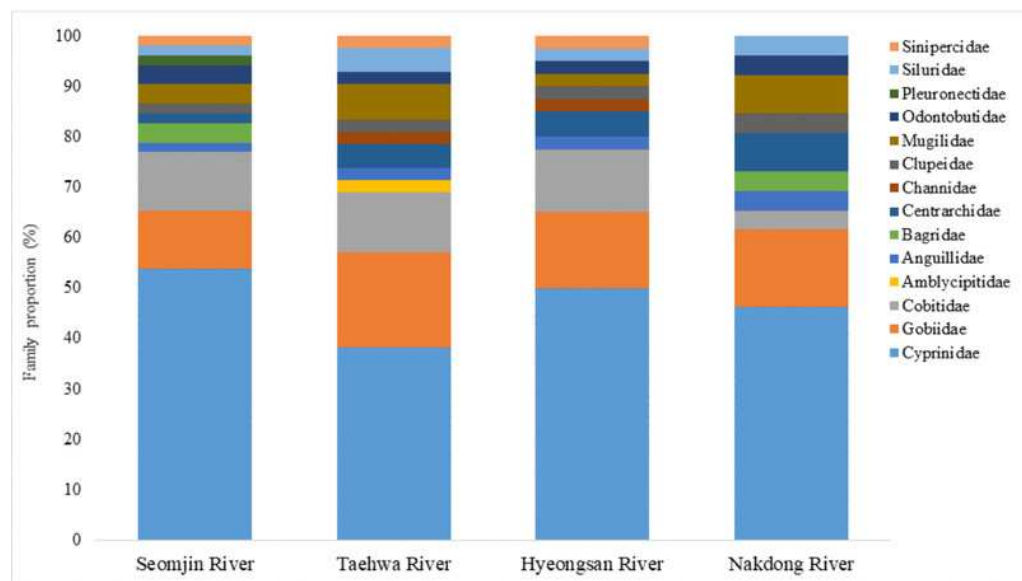
Figure 8

Heat map analysis of top 30 fish species identified in 16 sampling stations of the four
Korean rivers.

Figure 8 Heat map analysis of ~~top 30 fish species~~ identified in 16 sampling stations of ~~the~~
four Korean rivers. Heat map analysis was constructed by Primer v7 program.
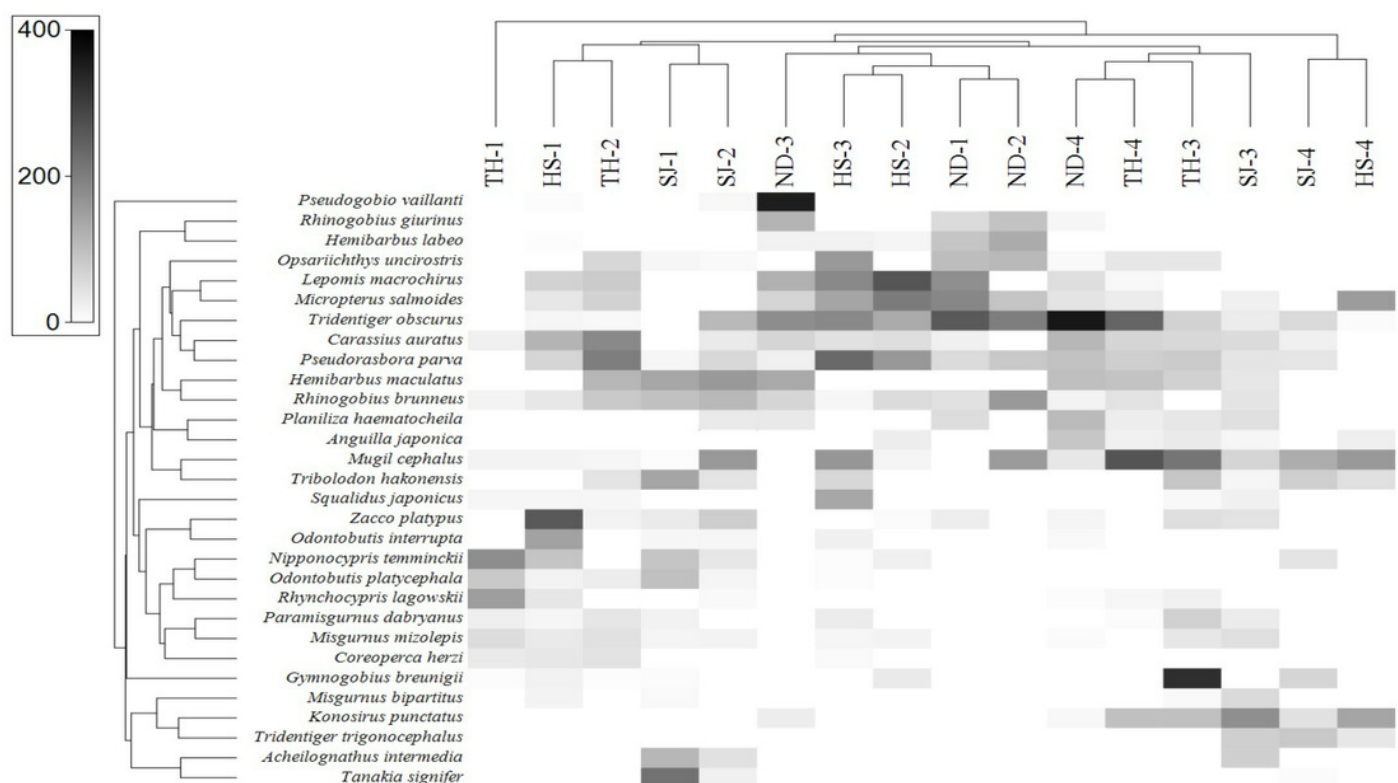
Table 1 Environmental DNA sample collection sites with physico-chemical parameters of ~~the~~ four Korean rivers

1 Table 1 Environmental DNA sample collection sites with physico-chemical parameters of ~~the~~ four
2 Korean rivers

River	Date	Station	GPS location	Temp. (0c)	Salinity (PSU)
Hyeongsan	2018.06.11	HS1	N 35° 42' 36", E 129° 11' 42"	18.6	1.00
		HS2	N 35° 56' 14", E 129° 14' 24"	19.5	2.02
		HS3	N 35° 59' 32", E 129° 17' 19"	20.0	3.20
		HS4	N 36o 01' 51", E 129° 23' 01"	24.0	20.20
Taehwa	2018.06.11	TH1	N 35° 32' 52", E 129° 06' 27"	19.4	1.02
		TH2	N 35° 35' 07", E 129° 13' 52"	19.8	2.04
		TH3	N 35° 32' 42", E 129° 17' 38"	22.7	14.02
		TH4	N 35° 32' 39", E 129° 21' 24"	19.2	17.80
Seomjin	2018.06.12	SJ1	N 35° 11' 18", E 127° 37' 21"	24.2	0.15
		SJ2	N 35° 04' 30", E 127° 43' 35"	23.4	2.01
		SJ3	N 35° 01' 54", E 127° 46' 32"	23.0	12.9
		SJ4	N 34° 58' 01", E 127° 45' 28"	23.25	16.8
Nakdong	2018.06.12	ND1	N 35° 23' 19", E 128° 29' 09"	24.0	1.92
		ND2	N 35° 20' 40", E 128° 46' 26"	24.1	2.40
		ND3	N 35° 17' 57", E 128° 58' 37"	23.2	2.78
		ND4	N 35° 07' 13", E 128° 57' 07"	22.5	4.50

5

Table 2 Taxonomic assignment summary of of the MiSeq reads from four Korean rivers

1   Table 2. Summary of taxonomic assignment of the MiSeq reads from four Korean rivers

	Seomjin River	Taehwa River	Hyeongsan River	Nakdong River	Total
Raw reads	561,473	609,755	601,165	543,212	2,315,605
Processed Merged reads	553,175	600,744	592,281	534,650	2,280,850
Total Haplotypes	76	67	53	42	238 (125)*
Haplotypes with species name	61	49	48	31	189 (105)*
Total species	52	42	40	26	160 (73)*

3   * Final number, after removal of duplicated one in brackets

Table 3 List of haplotypes of fishes identified by eDNA metabarcoding study in four Korean rivers

1   **Table 3: List of fish haplotypes with the GenBank numbers identified from the eDNA metabarcoding study of the four rivers**

No.	Family	Haplotype ID	Haplotypes	Identity (%)	Korean haplotype	Chinese haplotype	Japanese haplotype	Others
1	Gobiidae	SJ3	*Acanthogobius hasta*	100	KM030428	KM891736	-	
2	Gobiidae	TH3	*Acanthogobius lactipes*	100	KM030431	-	LC385140	
3	Cyprinidae	SJ1	*Acheilognathus intermedia*	99	EF483933	-	-	
4	Cyprinidae	HS1	*Acheilognathus macropterus*	99	EF483935	KJ499466	LC092100	
5	Cyprinidae	SJ1	*Acheilognathus majusculus*	99	-	-	LC006056	
6	Cyprinidae	SJ2	*Acheilognathus rhombeus*	99	KT601094	-	LC146100	
7	Cyprinidae	SJ1	*Acheilognathus* sp. (unidentified)	95			LC006056	
8	Anguillidae	TH4	*Anguilla japonica*	100	HQ185628	MH050933	LC193417	
9	Cyprinidae	HS1	*Carassius auratus*	100	-	KX505165		
10	Cyprinidae	TH2	*Carassius auratus*	100				Turkey KM657132
11	Cyprinidae	TH3	*Carassius auratus*	99		AY771781	LC193299	
12	Cyprinidae	SJ2	*Carassius auratus*	99	-	AY771781	LC193299	
13	Cyprinidae	TH3	*Carassius cuvieri*	100	-	-	AP011237	
14	Cyprinidae	SJ3	*Carassius cuvieri*	100			AP011237	
15	Channidae	TH1	*Channa argus*	100	-	MG751766	AB972107	
16	Cobitidae	TH1	*Cobitis* sp.	97	EU670794	-	LC146139	
17	Cobitidae	TH1	*Cobitis* sp.	97	EU670794	-	LC146139	
18	Cobitidae	SJ2	*Cobitis tetralineata*	100	EU670794	-	LC146139	
19	Cobitidae	SJ1	*Cobitis tetralineata*	99	EU670794	-	LC146139	
20	Cyprinidae	SJ1	*Coreoleuciscus* sp. (unidentified)	96	JN831358	-	AP011258	
21	Cyprinidae	SJ1	*Coreoleuciscus splendidus*	100	JN831358	-	AP011258	
22	Sinipercidae	HS3	*Coreoperca herzi*	100	KR075132	-	-	
23	Sinipercidae	SJ1	*Coreoperca* sp.	97	KR075132	-	-	
24	Cyprinidae	ND4	*Cyprinus carpio*	100	-	KX710076	AP017363	
25	Cyprinidae	HS2	*Cyprinus carpio*	100	-	KX710076	AP017363	
26	Cyprinidae	ND3	*Cyprinus carpio*	99	-	KX710076	AP017363	
27	Cyprinidae	TH2	*Cyprinus megalophthalmus*	100	-	KR869143	-	

28	Gobiidae	SJ3	*Favonigobius gymnauchen*	100	-	-	LC385206	
29	Gobiidae	HS1	*Gymnogobius breunigii*	99	KM030451	-	-	
30	Gobiidae	HS1	*Gymnogobius* sp.	98	KM030451	-	-	
31	Gobiidae	TH3	*Gymnogobius* sp.	98	KM030451	-	-	
32	Cyprinidae	SJ1	*Hemibarbus labeo*	100	DQ347953	KP064328	LC049898	
33	Cyprinidae	ND2	*Hemibarbus maculatus*	99	-	NC018534		
34	Cyprinidae	SJ1	*Hemibarbus* sp.	97	DQ347953	KP064328	LC049898	
35	Cyprinidae	SJ2	*Hemibarbus* sp.	97	DQ347953	KP064328	LC049898	
36	Cyprinidae	TH4	*Hemibarbus* sp. (unidentified)	95	DQ347953	KP064328	LC049898	
37	Cyprinidae	ND1	*Hemiculter leucisculus*	100	-	-	LC340359	
38	Cobitidae	SJ1	*Iksookimia longicorpa*	100	KM676413	-	LC146135	
39	Cobitidae	HS1	*Iksookimia yongdokensis*	100	EU670800	-	-	
40	Cobitidae	TH2	*Iksookimia yongdokensis*	99	EU670800	-	-	
41	Pleuronectidae	SJ3	*Kareius bicoloratus*	100	-	-	AP002951	
42	Clupeidae	TH3	*Konosirus punctatus*	100	-	KC477844	LC020951	Taiwan AP011612
43	Clupeidae	ND3	*Konosirus punctatus*	99	-	KC477844	LC020951	Taiwan AP011612
44	Centrarchidae	TH4	*Lepomis macrochirus*	100	-	JN389795	AP005993	USA KP013118
45	Amblycipitidae	SJ1	*Liobagrus* sp.	97	KR075136	KX096605	AP012015	
46	Cyprinidae	SJ2	*Microphysogobio koreensis*	100	FJ515920	-	-	
47	Cyprinidae	SJ1	*Microphysogobio yaluensis*	99	KR075133	-	AP012073	
48	Centrarchidae	ND1	*Micropterus salmoides*	100	-	HQ391896	LC069536	USA DQ536425
49	Centrarchidae	HS1	*Micropterus salmoides*	99	-	HQ391896	LC069536	USA DQ536425
50	Cobitidae	SJ1	*Misgurnus anguillicaudatus*	100	-	KC762740	-	
51	Cobitidae	TH1	*Misgurnus anguillicaudatus*	99	-	KC762740	-	
52	Cobitidae	SJ2	*Misgurnus anguillicaudatus*	99	EU670804	-	-	

53	Cobitidae	HS1	*Misgurnus anguillicaudatus*	99	-	-	LC385093	
54	Cobitidae	HS1	*Misgurnus bipartitus*	100	-	KF562047	LC091592	
55	Cobitidae	TH3	*Misgurnus mizolepis*	100	AP017654	-	-	
56	Cobitidae	HS3	*Misgurnus mizolepis*	99	AP017654	-	-	
57	Mugilidae	HS1	*Mugil cephalus*	100	-	KF374974	LC278014	
58	Gobiidae	TH3	*Mugilogobius abei*	100	KM030465	-	LC421743	Taiwan KF128984
59	Cyprinidae	TH1	*Nipponocypris koreanus*	100	-	KJ427719	-	
60	Cyprinidae	HS1	*Nipponocypris temminckii*	100	-	-	AP012116	
61	Cobitidae	TH1	*Niwaella multifasciata*	100	EU670807	-	LC146133	
62	Cobitidae	HS1	*Niwaella* sp. (unidentified)	96	EU670807	-	LC146133	
63	Odontobutidae	SJ1	*Odontobutis interrupta*	100	KR364945	-	-	
64	Odontobutidae	HS1	*Odontobutis platycephala*	100	KM030426	-	-	
65	Odontobutidae	SJ2	*Odontobutis platycephala*	99	KM030426			
66	Cyprinidae	HS1	*Opsariichthys* sp. (unidentified)	96	-	-	AB218897	
67	Cyprinidae	TH3	*Opsariichthys uncirostris*	99	-	-	AB218897	
68	Cobitidae	TH4	*Paramisgurnus dabryanus*	100	-	KM186182	LC146125	
69	Cobitidae	HS1	*Paramisgurnus dabryanus*	100	-	KJ699181	LC146125	
70	Cyprinidae	SJ2	*Phoxinus oxycephalus*	99	MK208924	-	AB626852	
71	Cyprinidae	SJ3	*Phoxinus oxycephalus*	99	MK208924	-	AB626852	
72	Cyprinidae	TH3	*Phoxinus semotilus*	100	KT748874	-	-	
73	Mugilidae	TH3	*Planiliza affinis*	100	-	KM925142	LC277843	
74	Mugilidae	SJ2	*Planiliza haematocheila*	100	-	KJ622047	LC021099	
75	Mugilidae	HS4	*Planiliza haematocheila*	100	-	KJ622047	LC021099	
76	Bagridae	SJ1	*Pseudobagrus koreanus*	100	KT601095	-	-	
77	Bagridae	ND1	*Pseudobagrus ussuriensis*	100	-	KC188782	-	
78	Bagridae	ND2	*Pseudobagrus ussuriensis*	99	-	KC188782	-	
79	Cyprinidae	ND2	*Pseudogobio esocinus*	100	-	-	LC340042	
80	Cyprinidae	ND1	*Pseudogobio esocinus*	99	-	-	LC340042	
81	Cyprinidae	ND3	*Pseudogobio vaillanti*	100	-	KU314695	LC146041	
82	Cyprinidae	SJ2	*Pseudogobio vaillanti*	99	-	KU314695	LC146041	
83	Gobiidae	TH3	*Pseudogobius masago*	100	KM030467	-	LC049791	

84	Cyprinidae	TH1	*Pungtungia herzi*	99	KF006339	-	AB239598	
85	Cyprinidae	SJ1	*Pungtungia* sp.	97	KF006339	-	AB239598	
86	Cyprinidae	TH1	*Pungtungia* sp. (unidentified)	96	KF006339	-	AB239598	
87	Gobiidae	HS1	*Rhinogobius brunneus*	100	KT601096	-		
88	Gobiidae	ND2	*Rhinogobius brunneus*	100			LC049760	
89	Gobiidae	ND1	*Rhinogobius giurinus*	100	KM030475	KP892753	LC049748	
90	Cyprinidae	SJ2	*Rhodeus suigensis*	100	EF483934	-	-	
91	Cyprinidae	SJ1	*Rhodeus uyekii*	100	EF483937	-	-	
92	Cyprinidae	HS1	*Rhynchocypris lagowskii*	99	-	KJ641843	-	
93	Cyprinidae	TH3	*Rhynchocypris lagowskii*	99		KJ641843		
94	Cyprinidae	TH4	*Rhynchocypris lagowskii*	99		KJ641843		
95	Cyprinidae	SJ2	*Rhynchocypris oxycephalus*	99	-	-	LC193377	
96	Cyprinidae	SJ3	*Rhynchocypris oxycephalus*	99			LC193377	
97	Cyprinidae	HS4	*Rhynchocypris* sp.	98			LC193377	
98	Cyprinidae	HS2	*Sarcocheilichthys soldatovi*	100	-	-	LC146036	
99	Cyprinidae	HS2	*Sarcocheilichthys* sp.	97	KU301744	-	AP012067	
100	Cyprinidae	ND3	*Sarcocheilichthys* sp.	97	KU301744	-	AP012067	
101	Cyprinidae	SJ2	*Sarcocheilichthys variegatus*	100	KU301744	-	AP012067	
102	Siluridae	ND1	*Silurus asotus*	100	-	JX087351	NC015806	
103	Siluridae	TH1	*Silurus microdorsalis*	99	KT350610	-	-	
104	Siluridae	SJ1	*Silurus* sp. (unidentified)	96	KT350610			
105	Sinipercidae	SJ1	*Siniperca scherzeri*	100	-	MF966985	-	Taiwan AP014527
106	Cyprinidae	SJ2	*Squalidus chankaensis*	100	KT948082	-	-	
107	Cyprinidae	HS3	*Squalidus japonicus*	100			LC277782	
108	Cyprinidae	SJ3	*Squalidus japonicus*	99			LC277782	
109	Cyprinidae	TH3	*Squalidus japonicus coreanus*	100	KR075134	-		
110	Cyprinidae	HS1	*Squalidus multimaculatus*	100	KX495606	-	-	
111	Bagridae	SJ1	*Tachysurus fulvidraco*	100	-	KU133295	LC193372	
112	Bagridae	ND2	*Tachysurus nitidus*	100	-	KC822643	-	
113	Cyprinidae	SJ1	*Tanakia signifer*	99	EF483930	-	-	

114	Cyprinidae	SJ2	*Tanakia somjinensis*	99	FJ515921	-	-
115	Cyprinidae	SJ1	*Tanakia* sp.(unidentified)	96	FJ515921		
116	Cyprinidae	TH2	*Tribolodon hakonensis*	100	-	-	AB626855
117	Cyprinidae	SJ3	*Tribolodon hakonensis*	99	-	-	AB626855
118	Gobiidae	TH4	*Tridentiger obscurus*	100	KT601092	MF663787	LC193168
119	Gobiidae	SJ2	*Tridentiger radiatus*	99	-	EU047755	-
120	Gobiidae	ND2	*Tridentiger radiatus*	99			
121	Gobiidae	SJ3	*Tridentiger trigonocephalus*	100	KM030481		
122	Gobiidae	HS4	*Tridentiger trigonocephalus*	100		KT282115	LC385175
123	Cyprinidae	SJ1	*Zacco platypus*	100	-		LC277796
124	Cyprinidae	HS1	*Zacco platypus*	99		KF683339	
125	Cyprinidae	TH1	*Zacco* sp.	97		KF683339	

Table 4 Shannon Index (SI) measured from four Korean rivers by eDNA metabarcoding

Table 4 Shannon Index (SI) measured from four Korean rivers by eDNA metabarcoding

	Seomjin River	Taehwa River	Hyeongsan River	Nakdong River	Average
Station 1	2.197	2.073	1.755	1.777	1.951
Station 2	2.182	1.941	1.709	1.734	1.892
Station 3	2.125	1.631	1.691	1.465	1.728
Station 4	2.105	1.443	1.102	1.008	1.415
Overall SI index	3.48	3.067	2.954	2.864	-