

Title

Comparing local- and regional-scale estimations of the diversity of stream fish using eDNA metabarcoding and conventional observation methods

Authors

Hikaru Nakagawa¹, Satoshi Yamamoto², Yukuto Sato³, Tetsuya Sado⁴, Toshifumi Minamoto⁵ and Masaki Miya⁴

¹ Field Science Education and Research Center, Kyoto University, Kyoto, Japan

² Department of Zoology, Graduate School of Science, Kyoto University, Kyoto, Japan

³ The Ryukyu University, Okinawa, Japan

⁴ Natural History Museum and Institute, Chiba, Japan

⁵ Graduate School of Human Development and Environment, Kobe University, Kobe, Japan

Corresponding Author

Hikaru Nakagawa

Address: Onojya 1, Miyama, Nantan, Kyoto, Japan

Telephone: +81 771-77-0321

Fax: +81 771-77-0323

Email: hikaunakagawa@icloud.com

Running title

Fish fauna estimation by eDNA metabarcoding

Keywords

Assemblage pattern, Environmental DNA, Metabarcoding, MiFish primers, Stream fish

Summary

1. We present a performance evaluation of environmental DNA (eDNA) metabarcoding with MiFish-U/E primers to investigate local and regional diversities of stream fish species to examine potential effectiveness, limits, and future remedies of this technique in large-scale monitoring. We hypothesized that eDNA inferences are more consistent with fish assemblages observed upstream than downstream due to a directional flow of river water.

2. River water was sampled at 102 sites in 51 rivers around Lake Biwa in the central part of Honshu Island, Japan, within 10 person-days; and fish species compositions inferred from eDNA and existing observational data were compared. Observation sites were chosen from the observational data that were within a certain distance (buffer range) of a water-sampling site along a river trajectory. The hypothesis of the detection bias of eDNA toward upstream assemblage was tested by comparing results with all of the observational data, data from a higher elevation, and data from a lower elevation. The Jaccard dissimilarity index was plotted between the observational data and the eDNA estimates against the buffer range; the buffer range with minimum dissimilarity was chosen.

3. When using existing observational data from within 6 km upstream of the eDNA sampling sites, the eDNA results were the most consistent with the observational data and inferred 88.6% of the species reported (38/44), as well as two additional species. eDNA results also showed patterns consistent with known upstream–downstream turnover of related species and biogeographical assemblage patterns of certain species.

4. Our 10-person-days survey using the metabarcoding technique enabled us to obtain as much regional fish diversity data including the hypothesized pattern of eDNA detection with an upstream bias as the accumulated observational data obtained through greater amounts of time, money, and labor. The problems regarding false-positive/negative detection were suggested in our survey, however, these should be decreased or removed by modifying the sampling methods and experimental procedures in future works. Therefore, we concluded this

new tool to enable monitoring that has never been implemented, such as cross-nation, and even whole-Earth monitoring with the data at yearly, seasonal, or finer temporal scales.

Introduction

DNA shed from fishes via metabolic waste, damaged tissue, or sloughed skin cells has been detected in various aquatic environments (Kelly *et al.*, 2014a), including ponds (Takahara *et al.*, 2012, 2013; Sigsgaard *et al.*, 2015), rivers (Jerde *et al.*, 2013; Mahon *et al.*, 2013; Wilcox *et al.*, 2013; Yamanaka & Minamoto, 2016), and marine waters (Yamamoto *et al.*, 2016). Genetic material found in the water column is referred to as environmental DNA (eDNA). The ubiquitous presence of fish eDNA in the water column has led to its use as a tool to detect invasive (Takahara *et al.*, 2013; Jerde *et al.*, 2013; Mahon *et al.*, 2013) and rare or threatened species (Wilcox *et al.*, 2013; Sigsgaard, 2015), as well as to determine biomass (Takahara *et al.*, 2012; Doi *et al.*, 2017), the movement and spawning activity of fish (Erickson *et al.*, 2016; Yamanaka & Minamoto, 2016), and parameters of population genetics (Sigsgaard *et al.*, 2016). These pioneering studies have shown that eDNA can be used as a noninvasive genetic monitoring tool in various fields of fish biology.

While these studies detected known target species using specific PCR primers, recent advances in the technology of metabarcoding using universal PCR primers enabled us to elucidate the complete biodiversity using eDNA (Thomsen *et al.*, 2012a; Kelly *et al.*, 2014a; Valentini *et al.*, 2016). At present, there are a few universal primer sets for eDNA metabarcoding of fishes (12S-V5: Riaz *et al.*, 2011; Kelly *et al.*, 2014b; MiFish-U/E: Miya *et al.*, 2015; Ac12s and Ac16s: Evans *et al.* 2016; teleo: Valentini *et al.*, 2016). The performance of the MiFish-U/E primers has been examined using eDNA samples from large aquariums with known species compositions and has demonstrated a high potential to identify fishes (>90% of expected species) from diverse taxonomic species (168 species from 14 orders) (Miya *et al.*, 2015). However, only a few field evaluations have been conducted using

MiFish-U/E in marine ecosystems (Miya *et al.*, 2015; Yamamoto *et al.*, 2017). To establish this technology as a practical tool for field monitoring, the MiFish-U/E primers must be evaluated in the field in various aquatic ecosystems (e.g., Valentini *et al.*, 2016; Civade *et al.*, 2016).

Here, we present the first performance evaluation of this technique in rivers. Rivers differ from other aquatic habitats (e.g., oceans, coasts, swamps, or lakes) in several ways, such as the continuous and directional water flow and one-dimensional species turnover along a river continuum (Allan & Castillo, 2007). In lotic ecosystems, genetic materials shed from organisms are expected to flow downstream until they are chemically or biologically decomposed (Deiner & Altermatt 2014; Jane *et al.* 2014; Wilcox *et al.* 2016). As such, eDNA collected from fishes should represent the combination of upstream and local fish assemblages at a river sampling site (Civade *et al.*, 2016; Deiner *et al.*, 2016). In contrast, direct observations and sampling provide more pinpoint data compared with eDNA methods.

Therefore, we aimed (i) to test the hypothesis that estimations of fish species composition using eDNA will better reflect the assemblages observed at sites upstream from the water sampling location than those observed downstream, and (ii) to examine potential effectiveness, limits, and future remedies of the technique in large-scale monitoring. Lake Biwa and the surrounding watersheds encompass one of the most species-rich freshwater regions in Japan (Kawanabe & Mizuno, 2001; Uonokai, 2005). The diversity and distribution of fish in the region has also been intensively surveyed by a local museum and corporate volunteers (Uonokai, 2005). This makes the region ideal for a field test of MiFish-U/E to assess river fish diversities and examine the above hypothesis. We sampled river water from 102 sites in 51 rivers and tested the performance of the primers by comparing data inferred from eDNA to existing observational data.

Methods

Sampling sites and previous surveys of fish distributions

Lake Biwa, the largest and oldest lake in Japan, is located in the central part of Honshu Island and has more than 100 river inflows. With the exception of the southern side, Lake Biwa is surrounded by highlands: the Hira Mountain range to the west, the Tanba Highlands to the northwest, the Suzuka Mountain range to the east, and the Nosaka and Ibuki Highlands to the northeast. A relatively large plain lies to the east of the lake (Fig. 1). In Japan, these geological structures separating watersheds generate complex faunal structures of freshwater fish species (i.e., biogeographical borders; Watanabe, 2012). Those faunal structures are not explained by contemporary differences in river environments (Avice, 2000). Indeed, despite the existence of rivers with similar scales and environments, several spatial differences of species composition are known at the sampling area, such as those at the eastern and western sides of the lake (Kawanabe & Mizuno, 2001; Uonokai, 2005).

Previous studies of stream fish distribution have been conducted in both major river systems (Matsumiya *et al.*, 2001; Uonokai, 2005; Ministry of Land, Infrastructure, Transport and Tourism, 2005; Nakagawa, 2014) and small and minor rivers (Ministry of Land, Infrastructure, Transport and Tourism Kinki Regional Development Bureau Biwako Office, 2004; Uonokai, 2005) of this watershed (see Appendix S1 in Supporting Information). In addition, direct underwater observations of fishes have been conducted at several sites by one of the authors of the present paper (H. Nakagawa, unpublished) (see Appendix S1). We compared our eDNA data to these previously collected sources of data.

Water sampling

Water samples (one sample per site) were collected by H. Nakagawa at 102 sites in 51 rivers around Lake Biwa (Fig. 1, see Appendix S2 in Supporting Information). Sampling was done over 10 days in the period from August 1, 2014 to October 10, 2014. Sampling sites were mainly in the upper-middle reaches of each river to detect fish species inhabiting lotic habitats.

Sampling sites were selected to cover all rivers flowing into Lake Biwa with catchment areas $\geq 2.0 \text{ km}^2$, except plain rivers as possible. The elevation, slope, and extent of the catchment areas of the rivers ranged from 42.4–650.3 m a.s.l, 0.1–30.5%, and 2.3–4,646.5 km^2 , respectively (see Appendix S2). For rivers or tributaries with catchment areas $\geq 50.0 \text{ km}^2$ within the Lake Biwa watershed, two or three sampling sites were designated along a river/tributary trajectory. Sites in large rivers/tributaries that contained habitats of the majority of lotic species, according to the observational data, were selected. In rivers neighboring the watershed of Lake Biwa, water was collected at reaches with 10–20 m^2 flow-widths where several changes in the presence/absence of fish species (e.g., *Hemibarbus longirostris* (Cyprinidae) and *Niwaella delicata* (Cobitidae)) were expected because of biogeographical borders. Apart from a few exceptions, *Hemibarbus longirostris* does not inhabit the northern and eastern rivers around Lake Biwa (Kawanabe & Mizuno, 2001; Uonokai, 2005), and *Niwaella delicata* does not inhabit rivers that flow into the northern and western sides of Lake Biwa (Kawanabe & Mizuno, 2001; Kitagawa *et al.*, 2001; Uonokai, 2005).

A 1-L sample of surface water from a riffle habitat with a water velocity $> 20 \text{ cm s}^{-1}$ was collected in a plastic bottle and immediately filtered using a glass filter (GF/F, Whatman) at the edge of the stream at each sampling site. A water sampling point with no stagnation or backward water flow was selected by eye. The glass filters were placed on ice, transported to the laboratory, and preserved in a -20°C freezer prior to experiments. The plastic bottle and all filtration equipment were bleached immediately after each sampling in a 20-L tank filled with 10 L 200 ppm sodium hypochlorite in water for at least 30 min and then washed with 4 L freshwater in a bucket from the river at the next sampling site just before the next sampling to reduce the risk of cross-contamination. Water used to rinse the bleached equipment was diluted to $< 5 \text{ ppm}$ sodium hypochlorite (i.e. weaker than tap water in Japan) and disposed of on site. The waste bleach in the 20-L tank was brought back to and disposed of in the laboratory.

DNA extraction, amplification, and high-throughput sequencing

DNA extractions followed the procedure of Yamamoto *et al.* (2016). Briefly, DNA was extracted from the filters using the DNeasy Blood and Tissue Kit (QIAGEN, Hilden, Germany) in combination with a Salivette tube (Sarstedt, Nümbrecht, Germany). The DNA solution (ca. 440 μ L) was purified using the DNeasy Blood and Tissue Kit following a modified version of the manufacturer's instructions. To check for cross-contamination during extractions, an empty column was simultaneously treated using the same procedures in each experiment (one blank per 40 samples, five extraction blanks in total).

DNA amplifications followed the procedure of Miya *et al.* (2015). Extracted eDNA samples were used for multiplex PCR with two universal primer pairs (MiFish-U/E). Before preparing a DNA library, work spaces and equipment were sterilized. Filtered pipette tips were used, and pre- and post-PCR products were separated to safeguard against contamination (Miya *et al.*, 2015). To monitor contamination, PCR blanks were included for each experiment through the first and second-round PCR (two PCR blanks in total). PCR procedures were duplicated for all samples to avoid missing values due to experimental error.

The first PCR was carried out with 35 cycles in a 12.0- μ L reaction volume containing 6.0 μ L 2 $\times$ KAPA HiFi HotStart ReadyMix (KAPA Biosystems, Wilmington, MA, USA), 0.36 μ L each MiFish-U/E primer (10 μ M), 4.28 μ L sterile distilled H₂O, and 1.0 μ L template. The first PCR product (150 bp paired-end sequences) was diluted 10 times using Milli-Q water and used as a template for the second PCR. The second PCR was carried out in a 12.0- μ L reaction volume containing 6.0 μ L 2 $\times$ KAPA HiFi HotStart ReadyMix, 1.8 μ L each primer (2 μ M), 1.4 μ L sterile distilled H₂O, and 1.0 μ L template. Different combinations of sequencing primers and sequencing adapters with sample indices with two or more differences in base pairs were used for different templates and PCR replicates for massively parallel sequencing using the MiSeq platform (Hamady *et al.*, 2008). The indexed second PCR products that were

pooled in equal volumes were purified by agarose gel electrophoresis. The DNA concentration of the pooled library was adjusted to a final concentration of 12.0 pM for sequencing on the MiSeq platform.

Taxonomic assignment

Data pre-processing and taxonomic assignments followed Miya *et al.* (2015) using the publicly available bioinformatics pipeline (<http://datadryad.org/resource/doi:10.5061/dryad.54v2q>; <http://mitofish.aori.u-tokyo.ac.jp/mifish>). The top BLAST hit with a sequence identity of $\geq 97\%$ and an E-value threshold of 10^{-5} was used for species assignments of each representative sequence. Sequences represented by at least 10 identical reads were subjected to subsequent analyses. Sequences representing the negative control were eliminated from the dataset.

Comparing eDNA and existing observational data

Existing observational data were sorted by geographic location (Fig. 1), elevation, and watershed boundaries. We chose observational data that were within a certain distance of our eDNA sampling site along a river trajectory (Fig. 2). Buffer zones of 1–10-km diameter were used, with 1-km intervals, to test the hypothesis that eDNA inferences are more consistent with fish assemblages observed upstream rather than downstream by comparing our results with all of the observational data (All), data only from the same site or at a higher elevation than that site (Upstream), and data from the same site or at a lower elevation than that site (Downstream). We also identified eDNA sampling sites that shared at least once species inference with previous studies in their buffer range and used them for later analysis. GIS data were provided by the National Land Numerical Information download service of Japan (<http://nlftp.mlit.go.jp/ksj/>) (elevation, rivers, lakes and ponds, and coastlines) and Conservation GIS Consortium Japan (<http://cgisj.jp>) (watershed boundaries). Data were

compiled using QGIS 2.2.0 and the GDAL Georeferencer plugin (Quantum GIS project, 2013).

We compared previous observations to the eDNA estimates using the Jaccard dissimilarity index for sampling sites. We plotted the dissimilarity index against the buffer range in each comparison (All, Upstream, and Downstream) and chose the buffer range with the minimum dissimilarity. We also plotted the number of sampling sites available for comparison, the median number of observational data points within the buffer zone around each sampling site, and the total number of fish species inferred by the observational data against the buffer ranges. These were used as supplementary data to determine the best buffer range.

To examine the potential effect of sequence depth on the false negative detection of eDNA, in the best matching dataset of eDNA and previous observations, we compared the pattern of consistency of the two estimations between the entire dataset and the data subset with a ≥ 1000 sequence depth in which most sites were saturated in species number (see Appendix S4 in Supporting Information). We categorized sampling sites into categories (i)–(iv) for each fish species as follows: the presence of fish was inferred (i) only by eDNA, (ii) only by the observational data, (iii) by both, or (iv) by neither (shown as e, L, B and blank, respectively, in Appendix S5). Then, the total numbers for the categories were compared between two datasets by Fisher's exact test.

The interspecific variation of the total number of sites within the categories (i)–(iii) mentioned above was examined for both the entire dataset and the data subset with ≥ 1000 sequence depth. The multi-dimensional Euclidian distance based on the total numbers was calculated among fishes, and was used as the index of interspecific dissimilarity in the pattern of the consistency between eDNA and the observational data. Then, a cluster dendrogram was drawn with the Ward method using the dissimilarities, and the five highest clusters were determined visually; these patterns of consistency were compared between the two estimations. All statistical analyses were performed using the 'MASS' (Venables & Ripley,

2002) and ‘vegan’ (Oksanen, 2017) packages of the R 3.2.0 software (R Developmental Core Team, 2015).

Results

MiSeq sequencing and data analysis

The MiSeq paired-end sequencing of the 218 libraries (102 sampling sites and 7 negative controls, each with duplicate PCRs) yielded a total of 1,662,988 reads (DNA Data Bank of Japan Sequence Read Archive accession no. DRA005106). After demultiplexing and subsequent preprocessing of the raw data from MiSeq, 1,178,682 reads were retained for subsequent analyses. Of these, 901,997 reads (76.5%) were assigned to known species with $\geq 97\%$ identity to reference sequences in the database, and 864,420 reads (73.3%) were identified as fishes recorded in the study region.

Fish species inferred by eDNA

Among 82 of the 102 sampling sites, assigned sequences of 55 fish species were identified (see Appendix S3). A total of 15 marine fish species that have never been expected to inhabit the sampling sites were detected across 11 sites. The number of species excepting marine fishes detected at each sampling site ranged from 1 to 23. The three most frequently detected taxa were *Rhynchocypris oxycephalus* (Cyprinidae) (60 sites), *Nipponocypris temminckii* (Cyprinidae) (51 sites), and *Rhinogobius* spp (Gobiidae). (46 sites), all of which had high numbers of MiFish reads (162,559, 176,051, and 146,029 reads, respectively). The genera *Oncorhynchus* (Salmonidae), *Carassius* (Cyprinidae), *Cobitis* (Cobitidae), *Rhinogobius* (Gobiidae), and a part of *Hemibarbus* (*Hemibarbus labeo* and *Hemibarbus barbus* (Cyprinidae)) were not sorted into lower taxonomic levels because of limited variations in the short sequences between the MiFish primers. One extraction from negative control yielded 45 reads of a sequence from *Rhinogobius* spp. (34 reads) and 11 reads from *Tribolodon*

hakonensis (Cyprinidae), and these two sequences were excluded from the dataset. These sequences were commonly detected throughout samples (*Tribolodon hakonensis*, 16 samples; *Rhinogobius* spp., 60 samples), and the source of the contamination could not be identified.

The mean $\pm$ SD of sequence depth was $8,478 \pm 13,741$, and the median sequence depth was 302 per sampling site for sequences assigned to fishes recorded in the study region. Of the 102 sampling sites, 45 sites had a sequence depth $\geq 1,000$.

Similarity between eDNA and existing observational data

eDNA and observational data were most similar when only upstream observational data within a 6 km buffer zone were compared (Fig. 3a; Jaccard dissimilarity index = 0.63 ± 0.00 , Median $\pm$ SE), with 48 sampling sites sharing one or more species with the observational data. The number of available observational data sites and number of total fish species were minimal when only data within a 1 km buffer zone of the eDNA collection point were considered, and this remained constant in the datasets from larger <2 km buffer zones (Fig. 3b, d). By contrast, the number of fish species in the observational data did not increase or decrease with buffer zone range (Fig. 3c).

In datasets using ‘All’ observational data and ‘Downstream’ only data, the similarities between the eDNA and the literature were maximized when using data within a <2 km buffer zone (Fig. 3a). The available observational data and the number of fish species detected increased with increasing buffer range.

In the dataset using upstream observational data within a 6 km buffer zone, the number of species that were detected by eDNA covered 86.4% (38/44) of the fish species that were reported in the observational data (see Appendix S5 in Supporting Information). In addition, eDNA analysis identified two species (*Opsariichthys uncirostris* (Cyprinidae) and *Gasterosteus aculeatus* (Gasterosteidae)) not recorded in the observational data. The species recorded by previous observations alone were *Lethenteron reissneri* (Petromyzontidae),

Tanakia lanceolata (Cyprinidae), *Acheilognathus rhombeus* (Cyprinidae), *Lefua echigonia* (Balitoridae), *Oryzias latipes* (Adrianichthyidae), and *Poecilia reticulata* (Poeciliidae).

Therefore, the habitation of a total of 46 fish species was inferred by eDNA or observational data at the sampling sites.

In the comparison between the dataset of all sampling sites sharing one or more species with previous reports and that of the sites with ≥ 1000 sequence depth, patterns of consistency between the presence/absence of detection by eDNA and the observational data were not significantly different from each other (Fisher's exact test, $p > 0.99$; Table 1).

Based on a cluster analysis using the data from the 48 (of 102) sampling sites sharing more than one species with previous reports, fishes were categorized into five clusters as follows: Cluster 1 (three species), detected at ≥ 38 of the 48 available sites using eDNA and/or the observational data, $\geq 65.1\%$ of which were detected using both methods; Cluster 2 (11 species), detected at 18–31 of the available sites using eDNA and/or the observational data (the detection rates using the observational data alone [$44.6 \pm 15.2\%$] and using both the observational data and eDNA [$41.6 \pm 14.0\%$] were higher than that using eDNA alone [$13.8 \pm 8.5\%$]); Cluster 3 (six species), detected at 9–13 of the available sites using eDNA and/or the observational data (the detection rate using eDNA alone [$75.4 \pm 18.0\%$] was higher than that using both eDNA and the observational data [$17.4 \pm 11.2\%$] or the observational data alone [$7.1 \pm 11.2\%$]); Cluster 4 (11 species), detected at 4–14 of the available sites using eDNA or the observational data (the detection rate using the observational data alone [$\geq 57.1\%$] was higher than that of the other scenarios [$\geq 40.0\%$]); and Cluster 5 (15 species), with only rare detection (1–10 of the available sites using eDNA and/or the observational data) (Fig. 4a).

In the dataset with ≥ 1000 sequence depth, the five highest clusters were defined approximately the same as the clusters in the full dataset based on the pattern of consistency between eDNA or the observational data inferences (Fig. 4b). The species compositions of Clusters 1, 2, and 3 did not change from those in the full dataset except for the following

cases: *Zacco platypus* (Cyprinidae) from Cluster 2 to Cluster 1, *Liobagrus reinii* (Amblycipitidae) from Cluster 2 to Cluster 3, *Pseudogobio esocinus* (Cyprinidae) from Cluster 2 to Cluster 5, and *Cottus reinii* (Cottidae) from Cluster 3 to Cluster 5. By contrast, the species composition of clusters with a small number of detections changed as the available sampling sites (26 sites) decreased. Approximately half of the species within Clusters 4 and 5 in the full dataset composed other clusters in the dataset with ≥ 1000 sequence depth.

Overall, 7 of 11 detections of *Hemibarbus longirostris* and 16 of 21 detections of *Niwaella delicata* inferred by eDNA were consistent with biogeographic patterns from the observational data (Fig. 5a, b). Detection sites of *Cottus pollux* (Cottidae) were skewed toward the upper reaches compared with those of *Cottus reinii* (Fig. 5c, d).

Discussion

What is the extent of eDNA reflects the existing observational data?

On the whole, eDNA metabarcoding using the universal primer MiFish-U/E successfully detected most fish species near the sampling sites reported in the existing observational data. We hypothesized that fish species compositions estimated by eDNA would better reflect compositions upstream than downstream from a sampling site. Our results supported this hypothesis. The estimation of fish species composition matched best with observational data from sites ≤ 6 km upstream from a sampling site. Civade et al. (2016) used the universal Teleo primers to evaluate the spatial representativeness of eDNA metabarcoding, and found that eDNA signals of lake-inhabiting fishes were detectable in stream water at a site just downstream of a dam but not at a site 2 km downstream. The observational data on the stream fish distribution around Lake Biwa include data obtained over several years, which likely reflect the temporal fluctuations in fish distributions. Therefore, the estimate of the potential fish habitat based on the observational data might be broader than that based on snapshot monitoring, such as the direct observation by Civade et al. (2016).

The extent of the upstream area that contributes to eDNA detection presumably depends on the speed of flow of the stream water and the initial concentration and decomposition rate of genetic material (Deiner & Altermatt, 2014; Deiner *et al.*, 2016). In experiments that artificially introduced a caged animal into a river without a focal species, detection of eDNA reads decreased and reads disappeared a few hundred meters from the source (Jane *et al.*, 2015; Wilcox *et al.*, 2016), whereas the detectable distance reached a few kilometers downstream from a dam as an eDNA source in studies that examined a natural population of lentic animals (Deiner & Altermatt, 2014; Civade *et al.*, 2016). This difference might be due to downstream transport or storage of DNA in the stream bed (Jane *et al.*, 2015) or live individuals or carcasses of lentic species from upstream habitats (Deiner & Altermatt, 2014). In addition, previous studies have indicated the effect of many environmental factors on the efficiency of eDNA detection, such as water temperature, pH, UV, organic materials as PCR inhibitors, and the activity of microorganisms (Takahara *et al.*, 2012; Barnes *et al.*, 2014; Strickler *et al.*, 2014; Jane *et al.*, 2015; Tsuji *et al.*, 2017). However, field samplings always involve many factors assumed to affect the decomposition of eDNA. The relative importance of factors that determine patterns of bias of eDNA toward upstream assemblages may be an important research area for future studies.

Patterns in the consistency between eDNA and the existing observational data

We inferred two species (*Gasterosteus aculeatus* and *Opsariichthys uncirostris*) by eDNA alone, both of which were species reasonably detected in the sampling area. The native population of *Gasterosteus microcephalus* (Gasterosteidae), a closely related species of *Gasterosteus aculeatus*, is endangered and detailed information on its distribution is not publicly available due to conservation efforts (Uonokai, 2005). However, we do know that a hybrid population of *Gasterosteus microcephalus* and non-native *Gasterosteus aculeatus* exists at one of our sites (No. 38) (T. Kokita, personal communication). *Opsariichthys*

uncirostris uses river habitats only during early to mid-summer for spawning (Uonokai, 2005), and the small number of sampling records in the observational data may reflect this seasonality. The six species reported in the observational data alone were also reasonable, with the exception of the lamprey *Lethenteron reissneri* and the loach *Lefua echigonia*. Three of these species are pond or swamp species (*Tanakia lanceolata*, *Acheilognathus rhombeus*, and *Oryzias latipes*), which rarely inhabit upper-middle reaches of a river like the sampling sites (Kawanabe & Mizuno, 2001; Uonokai, 2005). The last species, *Poecilia reticulata*, is a nonnative species without confirmed establishment (Kawanabe & Mizuno, 2001; Uonokai, 2005).

In the results of cluster analysis, the lower number of species inferred by eDNA than by the observational data at some sites (i.e., Cluster 4) may be explained by habitat preferences. With the exception of *Salvelinus leucomaenis* (Salmonidae) in the full dataset, all fishes in Cluster 4 mainly inhabit lower reaches or lentic habitats (Kawanabe & Mizuno, 2001; Uonokai, 2005), whereas we mainly sampled river water at upper-middle reaches of rivers. The mismatch between sampling sites and the habitat preferences of fishes may have caused this discrepancy between our findings and the species reported in the observational data. In contrast to Cluster 4, fishes detected more frequently by eDNA than in previous reports (i.e., Cluster 3) may reflect differences in sampling methods. For example, in Uonokai (2005), local volunteers mainly sampled fish and this included elementary school students using hand nets. Therefore, fishes such as *Cyprinus carpio* (Cyprinidae), *Silurus asotus* (Siluridae), *Hemibarbus longirostris*, and *Hemibarbus* spp., which mainly inhabit deep pools or runs in large rivers, were less likely to be caught.

MiFish-U/E and its designated pipeline provide not only information on local and regional stream fish fauna but also data on interspecific differences in distribution along a river in pairs of closely related species (e.g., *Cottus pollux* and *C. reinii*), similar to data often reported as the result of interspecific competition and/or differences in habitat preferences (e.g.,

Taniguchi & Nakano, 2000). Previously reported differences in longitudinal distributions between closely related species (Kawanabe & Mizuno, 2001; Matsumiya *et al.*, 2001; Uonokai, 2005) were consistently inferred by eDNA in the genus *Cottus* (Fig. 5c–f). In addition, the inter-river system patterns of the eDNA-inferred presence/absence of *Hemibarbus longirostris* and *Niwaella delicata* were similar to, but not completely consistent with, the patterns driven by biogeographical processes (Kawanabe & Mizuno, 2001; Kitagawa *et al.*, 2001; Uonokai, 2005). Our results suggest that eDNA metabarcoding may be able to contribute to the analysis of assemblage patterns on which ecologists have traditionally focused, such as niche segregation of species within the same guild (Hutchinson, 1959), correlations between species composition and environmental factors (Townsend & Hildrew, 1994), and spatiotemporal dynamics driven by ecological and biogeographical processes (Leibold *et al.*, 2004).

We also detected eDNA sequences attributed to unexpected marine fishes at 11 sites. These marine fishes do not include species that are confamiliar to expected native species in the sampling sites, with the exception of *Acanthogobius flavimanus* (Gobiidae), and even this species has a ≥ 50 bp difference within the target sequence of MiFish-U/E to native species of the family Gobiidae at the sampling sites. Therefore, it is unlikely that misidentification of native species as marine species occurred. We instead suspect that eDNA assigned to marine fishes originated from domestic sewage. All marine fishes detected in our experiments are commonly used in Japanese food such as sushi and sashimi; they were detected mainly at sites near urban areas. For example, there is a famous hot springs resort in the upper reaches of site No. 3, at which the highest number of marine fish eDNA detections occurred. The area has many hotels that serve traditional Japanese foods and are likely sources of sewage containing the related genetic material. We investigated only the upper-middle reaches of rivers and thus could easily interpret the source of marine fish eDNA. However, artificial introduction of eDNA may be a concern for estimating fish fauna in areas potentially

inhabited by marine fishes, such as the lower reaches of rivers, estuaries, and oceans (e.g., Yamamoto *et al.*, 2017).

Potential limits and future remedies of eDNA metabarcoding

Although we followed decontamination procedures for laboratory spaces and equipment (see Materials and Methods) that are known to significantly limit contamination (Willerslev & Cooper, 2005), detection of *Gasterosteus aculeatus* at sites No. 37 and 70 was most likely due to contamination or tag-jump (Carlson *et al.*, 2012; Schnell *et al.*, 2015), given the environmental conditions of the sampling site and the habitat preference of this species. The former site was a small stream, highly fragmented by sand dams without fish-ways, and the latter was a small mountainous tributary, dominated by step-pool structures and separated from lower reaches by a large dam. *Gasterosteus aculeatus* prefers gentle plain habitats, and likely would not be able to survive in the habitats observed at sites No. 37 and 70 (Kawanabe & Mizuno, 2001). We suggest that this was contaminated during extraction or PCR from a neighboring sample (site No. 38), or accidental detection by tag-jump.

Ideally, the false-negative detection of rare species and factors related to false-positive detection caused by cross-contamination would be distinguished by multiple samplings per site (Willerslev & Cooper, 2005; Carlson *et al.*, 2012). To identify the step in the procedure at which cross-contamination occurred, blanks should be included for all steps in the protocol (i.e. water sampling, filtering, DNA extractions, 1st and 2nd PCRs) (Willerslev & Cooper, 2005). The sequence depth of eDNA was <1,000 reads at approximately half of the sampling sites in this study, and the false-negative detection of rare species was a concern. This problem may be decreased by quantitative PCR, which enables uniform DNA concentrations prior to sequencing (e.g., Wittwer *et al.*, 1997). The PCR and sequencing procedures performed in our study (i.e., the samples were identified by the combination of the index sequence that tagged by the two step PCR) are commonly used for eDNA metabarcoding

because of their simplicity and the shared use of tag sequences (Miya *et al.* 2015; Yamamoto *et al.* 2016). However, the one PCR step procedure (e.g., Civade *et al.*, 2016) may reduce the risk of cross-contamination, while that take more cost comparing with our procedure. In addition, we might reduce the effect of primer-bias (i.e., the variation in amplification efficiency among species) by using multiple universal primers (Elbrecht & Leese, 2015; Vallentini *et al.*, 2016). False-positive detection by tag-jump might be determined by using a unique tag sequence for all 3' and 5' primers when the number of samples in parallel sequencing is not large (Schnell *et al.*, 2015). Our sampling and experimental procedures did not follow these ideal methods exactly because of the constraints of costs and equipment, and thus the potential effects of cross-contamination could not be removed completely. Therefore, the interpretation of our results is limited, especially in terms of the local species diversity at each sampling site.

Concluding remarks

Overall, in an attempt to minimize false-positive detection, such as the elimination of sequences from a negative control, our 10-person-days survey using the metabarcoding technique enabled us to obtain at least as much regional fish diversity data as the accumulated observational data of traditional observations obtained through greater amounts of time, money, and labor. We also demonstrated a reasonable pattern of eDNA detection with a bias toward the upstream assemblages in the comparisons with observational data. For the local species diversity at each sampling site, the eDNA results showed patterns consistent with known upstream–downstream turnover of related species and the biogeographical assemblage patterns of certain species, but underestimation of the species diversity because of failure to detect rare species was also suggested. However, these problems regarding false-positive/negative detection should be decreased or removed by modifying the sampling methods and experimental procedures in future works. Therefore, we expect this new tool to

enable monitoring that has never been implemented, such as whole-assemblage, cross-nation, and even whole-Earth monitoring with the data at yearly, seasonal, or finer temporal scales.

Acknowledgements

We thank Tomoyuki Kokita for providing useful information on fish distribution. This study was supported by Sasagawa Scientific Research Grants in 2013 and 2014 from the Japan Science Society, by the Japan Society for the Promotion of Science KAKENHI Grant no. 14444453, and by JST CREST Grant no. JPMJCR13A2, Japan.

References

- Allan, J.D. & Castillo, M.M. (2007) *Stream ecology: Structure and function of running waters* 2nd edn. Springer, Dordrecht.
- Avice, J. C. (2000) *Phylogeography: the history and formation of species*. Harvard University Press, Cambridge.
- Barnes, M.A., Turner, C.R., Jerde, C.L., Renshaw, M.A., Chadderton, W.L. & Lodge, D.M. (2014) Environmental Conditions Influence eDNA Persistence in Aquatic Systems. *Environmental Science & Technology*, **48**, 1819–1827.
- Carlson, T., Aas, B.A., Lindner, D., Vrålstad, A., Schumacher, T. & Kauserud, H. (2012) Don't make a mista(g)ke: is tag switching an overlooked source of error in amplicon pyrosequencing studies? *Fungal Ecology*, **5**, 747–749.
- Civade, R., Dejean, T., Valentini, A., Roset, N., Raymond, J., Bonin, A., Taberlet, P. & Pont, D. (2016) Spatial representativeness of environmental DNA metabarcoding signal for fish biodiversity assessment in a natural freshwater system. *PLoS One*, **11**, e0157366.
- Deiner, K. & Altermatt, F. (2014) Transport distance of invertebrate environmental DNA in a natural river. *PloS One*, **9**, e88786.

- 491 Deiner, K., Fronhofer, E.A., Mächler, E., Walser, J., Altermatt, F. (2016) Environmental DNA
 492 reveals that rivers are conveyor belts of biodiversity information. *Nature*
 493 *Communications*, **7**, 12544.
- 494 Doi, H., Inui, R., Akamatsu, Y., Kanno, K., Yamanaka, H., Takahara, T. & Minamoto, T.
 495 (2017) Environmental DNA analysis for estimating the abundance and biomass of stream
 496 fish. *Freshwater Biology*, **62**, 30–39.
- 497 Elbrecht, V. & Leese, F. (2015) Can DNA-based ecosystem assessments quantify species
 498 abundance? Testing primer bias and biomass-sequence relationships with an innovative
 499 metabarcoding protocol. *PLoS One*, **10**, e0130324.
- 500 Erickson, R.A., Rees, C.B., Coulter, A.A., Merkes, C.M., McCalla, S.G., Touzinsky, K.F.,
 501 Wallesser, L., Goforth, R.R. & Amberg, J.J. (2016) Detecting the movement and
 502 spawning activity of bigheaded carps with environmental DNA. *Molecular Ecology*, **16**,
 503 957–965.
- 504 Evans, N., Olds, B. P., Renshaw, M. A., Turner, C. R., Li, Y., Jerde, C. L., Mahon, A. R.,
 505 Pfrender, M. E., Lamberti, G. A. & Lodge, D. M. (2016) Quantification of mesocosm
 506 fish and amphibian species diversity via environmental DNA metabarcoding. *Molecular*
 507 *Ecology Resources*, **16**, 29–41.
- 508 Hamady, M., Walker, J.J., Harris, K., Gold, N.J. & Knight, R. (2008) Error-correcting
 509 barcoded primers for pyrosequencing hundreds of samples in multiplex. *Nature Methods*,
 510 **5**, 235–237.
- 511 Hutchinson, G.E. (1959) Homage to Santa Rosalia or why are there so many kind of animals.
 512 *American Naturalist*, **95**, 145–159.
- 513 Jane, S.F., Wilcox, T.M., McKelvey, K.S., Young, M.K., Schwartz, M.K., Lowe, W.H.,
 514 Letcher, B.H. & Whiteley, A.R. (2014) Distance, flow, and PCR inhibition: eDNA
 515 dynamics in two headwater streams. *Molecular Ecology Resources*, **15**, 216–227.
- 516 Japan Ministry of Land, Infrastructure, Transport and Tourism (2015) *Census of Rivers and*

- 517 *Riparian Areas in 2013*. Ministry of Land, Infrastructure, Transport and Tourism, Tokyo,
 518 JP. (in Japanese)
- 519 Japan Ministry of Land, Infrastructure, Transport and Tourism Kinki Regional Development
 520 Breau Biwako Office (2004) *The survey of aquatic organs in lower–middle reaches of*
 521 *the Takatoki and Ane River*. Ministry of Land, Infrastructure, Transport and Tourism
 522 Kinki Regional Development Breau Biwako Office. Ohtsu, JP. (in Japanese)
- 523 Jerde, C.L., Chadderton, W.L., Mahon, A.R., Renshaw, M.A., Corush, J., Budny, M.L.,
 524 Mysorekar, S. & Lodge, D.M. (2013) Detection of Asian carp DNA as part of a Great
 525 Lakes basin-wide surveillance program. *Canadian Journal of Fisheries and Aquatic*
 526 *Science*, **70**, 522–526.
- 527 Kawanabe, H. & Mizuno, N. (2001) *Freshwater fishes of Japan*. Yamakei publication, Tokyo.
- 528 Kelly, B.P., Port, J.A., Yamahara, K.M., Martone, R.G., Lowell, N., Thomsen, P.F., Mach,
 529 M.E., Bennett, M., Prahler, E., Caldwell, M.R. & Crowder, L.B. (2014a) Harnessing
 530 DNA to improve environmental management. *Science*, **344**, 1455–1456.
- 531 Kelly, R.P., Port, J.A., Yamahara, K.M. & Crowder, L.B. (2014b) Using environmental DNA
 532 to census marine fishes in a large mesocosm. *PLoS One*, **9**, e86175.
- 533 Kitagawa, T., Yoshioka, M., Kashiwagi, M. & Okazaki, T. (2001) Population structure and
 534 local differentiation in the delicate loach, *Niwaella delicata*, as revealed by
 535 mitochondrial DNA and morphological analyses. *Ichthyological Research*, **48**,
 536 127–135.
- 537 Leibold, M.A., Holyoak, M., Mouquet, N., Amarasekare, P., Chase, J.M., Hoopes, M.F., Holt,
 538 R.D., Shurin, J.B., Law, R., Tilman, D., Loreau, M. & Gonzalez, A. (2004) The
 539 metacommunity concept: a framework for multi-scale community ecology. *Ecology*
 540 *Letters*, **7**, 601–613.
- 541 Mahon, A.R., Jerde, C.L., Galaska, M., Bergner, J.L., Chadderton, W.L., Lodge, D.M., Hunter,
 542 M.E. & Nico, L.G. (2013) Validation of eDNA surveillance sensitivity for detection of

- Asian carps in controlled and field experiments. *PLoS One*, **8**, e58316.
- Matsumiya, Y., Watanabe, K., Iguchi, K., Iwata, H., Yamamoto, G. & Nishida, M. (2001) Freshwater fishes of the Minami River system in Reinan Region, Fukui Prefecture, Japan. *Japanese Journal of Ichthyology*, **48**, 93–107. (in Japanese)
- Miya, M., Sato, Y., Fukunaga, T., Sado, T., Poulsen, J.Y., Sato, K., Minamoto, T., Yamamoto, S., Yamanaka, H., Araki, H., Kondoh, M. & Iwasaki, W. (2015) MiFish, a set of universal PCR primers for metabarcoding environmental DNA from fishes: Detection of >230 subtropical marine species. *Royal Society Open Science*, **2**, 150088.
- Nakagawa, H. (2014) Contribution of environmental and spatial factors to the structure of stream fish assemblages at different spatial scales. *Ecology of Freshwater Fish*, **23**, 208–223.
- Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Szoecs, E. & Wagner, H. (2017) *vegan: Community Ecology Package. R package version 2.4-2*. <https://CRAN.R-project.org/package=vegan>.
- Quantum GIS project (2013) *QGIS version 2.2.0*. [http:// www.qgis.org](http://www.qgis.org). Accessed 5 March 2014.
- R Development Core Team (2014) *R: a language and environment for statistical computing, Ver. 3.1.2*. R Foundation for Statistical Computing, Vienna, Austria, <http://cran.r-project.org/>. Accessed 5 May 2015.
- Riaz, T., Shehzad, W., Viari, A., Pompanon, F., Taberlet, P. & Coissac, E. (2011) ecoPrimers: inference of new DNA barcode makers from whole genome sequence analysis. *Nucleic Acids Research*, **39**, e145.
- Schnell, I.B., Bohmann, K. & Gilbert, M.T.P. (2015) Tag jumps illuminated – reducing sequence-to-sample misidentifications in metabarcoding studies. *Molecular Ecology Resources*, **15**: 1289–1303.

- 569 Sigsgaard, E.E., Carl, H., Møller, P.R. & Thomsen, P.F. (2015) Monitoring the near-extinct
 570 European weather loach in Denmark based on environmental DNA from water samples.
 571 *Biological Conservation*, **183**, 48–52.
- 572 Sigsgaard, E.E., Nielsen, I.B., Bach, S.S., Lorenzen, E.D., Robinson, D.P., Knudsen, S.W.,
 573 Pedersen, M.W., Jaidah, M.A., Orlando, L., Willerslev, E., Møller, P.R. & Thomsen, P.F.
 574 (2016) Population characteristics of a large whale shark aggregation inferred from
 575 seawater environmental DNA. *Nature Ecology & Evolution*, **1**, 0004.
- 576 Strickler, K.M., Fremier, A.K. & Goldberg, C.S. (2014) Quantifying effects of UV-B,
 577 temperature, and pH on eDNA degradation in aquatic microcosms. *Biological*
 578 *Conservation*, **183**, 85–92.
- 579 Takahara, T., Minamoto, T., Yamanaka, H., Doi, H. & Kawabata, Z. (2012) Estimation of fish
 580 biomass using environmental DNA. *PLoS One*, **7**, e35868.
- 581 Takahara, T., Minamoto, T. & Doi, H. (2013) Using environmental DNA to estimate the
 582 distribution of an invasive fish species in ponds. *PLoS One*, **8**, e56584.
- 583 Taniguchi, Y. & Nakano, S. (2000) Condition-specific competition: implication for the
 584 altitudinal distribution of stream fishes. *Ecology*, **81**, 2027–2039.
- 585 Thomsen, P.F., Kielgast, J., Iversen, L.L., Møller, P.R., Rasmussen, M. & Willerslev, E.
 586 (2012) Detection of a diverse marine fish fauna using environmental DNA from seawater
 587 samples. *PLoS One*, **7**, e41732.
- 588 Thomsen, P.F. & Willerslev, E. (2015) Environmental DNA – An emerging tool in
 589 conservation for monitoring past and present biodiversity. *Biological Conservation*, **187**,
 590 4–18.
- 591 Townsend, C.R. & Hildrew, A.G. (1994) Species traits in relation to a habitat templet for river
 592 systems. *Freshwater Biology*, **31**, 265–275.
- 593 Tsuji, S., Ushio, M., Sakurai, S., Minamoto, T. & Yamanaka, H. (2017) Water temperature-
 594 dependent degradation of environmental DNA and its relation to bacterial abundance.

- 595 *Plos One*, **12**, e0176608.
- 596 Uonokai (2005) Activities of Uonokai (fish survey group) - Fish distribution in the watershed
597 of Lake Biwa. *Research Report of Lake Biwa Museum*, **23**. (in Japanese)
- 598 Yamamoto, S., Minami, K., Fukaya, K., Takahashi, K., Sawada, H., Murakami, H., Tsuji, S.,
599 Hashizume, H., Kubonaga, S., Horiuchi, T., Hongo, M., Nishida, J., Okugawa, Y.,
600 Fujiwara, A., Fukuda, M., Hidaka, S., Suzuki, K.W., Miya, M., Araki, H., Yamanaka, H.,
601 Maruyama, A., Miyashita, K., Masuda, R., Minamoto, T. & Kondoh, M. (2016)
602 Environmental DNA as a ‘snapshot’ of fish distribution: A case study of Japanese jack
603 mackerel in Maizuru Bay, Sea of Japan. *PLoS One*, **11**, e0153291.
- 604 Yamamoto, S., Masuda, R., Sato, Y., Sado, T., Araki, H., Kondoh, M., Minamoto, T. & Miya,
605 M. (2017) Environmental DNA metabarcoding reveals local fish communities in a
606 species-rich coastal sea. *Scientific Reports*, 40368.
- 607 Yamanaka, H. & Minamoto, T. (2016) The use of environmental DNA of fishes as an efficient
608 method of determining habitat connectivity. *Ecological Indicators*, **62**, 147–153.
- 609 Valentini, A., Taberlet, P., Miaud, C., Civade, R., Herder, J., Thomsen, P.F., Bellemain, E.,
610 Besnard, A., Coissac, E., Boyer, F., Gaboriaud, C., Jean, P., Poulet, N., Roset, N., Copp,
611 G.H., Geniez, P., Pont, D., Argillier, C., Baudoin, J., Peroux, T., Crivelli, A.J., Olivier, A.,
612 Acqueberge, M., Burn, M.L., Møller, P.R., Willerslev, E. & Dejean, T. (2016)
613 Next-generation monitoring of aquatic biodiversity using environmental DNA
614 metabarcoding. *Molecular Ecology*, **25**, 929–942.
- 615 Venables, W.N. & Ripley, B.D. (2002) *Modern Applied Statistics with S* 4th edn. Springer,
616 New York.
- 617 Watanabe, K. (2012) Faunal structure of Japanese freshwater fishes and its artificial
618 disturbance. *Environmental Biology of Fishes*, **94**, 533–547.
- 619 Wilcox, T.M., McKelvey, K.S., Young, M.K., Jane, S.F., Lowe, W.H., Whiteley, A.R. &
620 Schwartz, M.K. (2013) Robust detection of rare species using environmental DNA: the

importance of primer specificity. *PLoS One*, **8**, e59520.

Wilcox, T.M., McKelvey, K.S., Young, M.K., Sepulveda, A.J., Shepard, B.B., Jane, S.F., Whiteley, A.R., Lowe, W.H. & Schwartz, M.K. (2016) Understanding environmental DNA detection probabilities: a case study using a stream-dwelling char (*Salvelinus fontinalis*). *Ecological Applications*, **194**, 209–216.

Willerslev, E. & Cooper, A. (2005) Review paper: Ancient DNA. *Proceedings of the Royal Society series B*, **272**, 3–16.

Wittwer, C.T., Herrmann, M.G., Moss, A.A. & Rasmussen, R.P. (1997) Continuous fluorescence monitoring of rapid cycle DNA amplification. *Bio Techniques*, **22**, 130–138.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Appendix S1 List of literature on fish distribution in the river systems surrounding Lake Biwa during 2001–2014.

Appendix S2 Location and geometric characters of the sampling sites. Geometric information was obtained from fifth-order mesh data (250 × 250 m) on elevation and slope and from the mesh data on river catchment area provided by the National Land Numerical Information download service of Japan (<http://nlftp.mlit.go.jp/ksj/>).

Appendix S3 Number of eDNA reads assigned to fish species detected in MiSeq analyses from each sampling site.

Appendix S4 Rarefaction curves of the number of detected fish species identified by the bioinformatics pipeline against the sequence depth (number of reads) for each sampling site. Bold lines and polygons indicate means and standard deviations estimated from 1000 bootstrap resampling simulations. The different colors of the lines and polygons indicate

647 differences in sequence depth: 1–99 reads, yellow; 100–999 reads, blue; 1,000–9,999 reads,
648 green; 10,000–100,000 reads, red.

649 **Appendix S5** Consistency between the presence/absence of each fish species inferred by
650 eDNA and observational data from within 5- km upstream of each eDNA sampling site. B, e,
651 and O indicate species that were inferred by both eDNA and the observational data, only by
652 eDNA, and only by the observational data, respectively.

653

654

TABLE

Table 1. The number of consistent/inconsistent cases of the presence/absence of detections by eDNA and observational data in the dataset of all sampling sites sharing one or more species with previous reports (full dataset) and that of sites with ≥ 1000 sequence depth.

Full data set	eDNA\Observational data	Presence	Absence
	Presence	243 (11.0%)	145 (6.6%)
	Absence	236 (10.7%)	1584 (71.7%)
Samples with ≥ 1000 sequences only	eDNA\Observational data	Presence	Absence
	Presence	85 (7.1%)	171 (14.3)
	Absence	136 (11.4)	804 (67.2)

FIGURE LEGENDS

Figure 1. Sampling sites of eDNA (red circles) and previous fish survey sites from the observational data (yellow circles). The number on each sampling site corresponds to the site No. in Appendix S2 in Supporting Information. Gray areas show the major mountain ranges and highlands (≥ 500 m elevation): the Hira Mountain range on the western side, Tanba Highland on the northwestern side, the Suzuka Mountain range on the eastern side, and the Nosaka and Ibuki Highlands on the northeastern side.

Figure 2. Schematic image of the comparison of eDNA and observational data. We chose existing observational data that were within a certain distance (buffer range) of our eDNA sampling site along a river trajectory at 1 km intervals. The comparisons were conducted for all of the observational data (All), data only from the same site or at a higher elevation (Upstream), and data from the same site or at a lower elevation (Downstream).

Figure 3. Relationships between buffer ranges from sampling sites and (a) the Jaccard dissimilarity index between eDNA and observational data, (b) the number of available sampling sites to compare eDNA and existing data, (c) the number of observational data points per sampling site, and (d) the total number of fish species inferred by the observational data at sampling sites where one or more species had been observed previously.

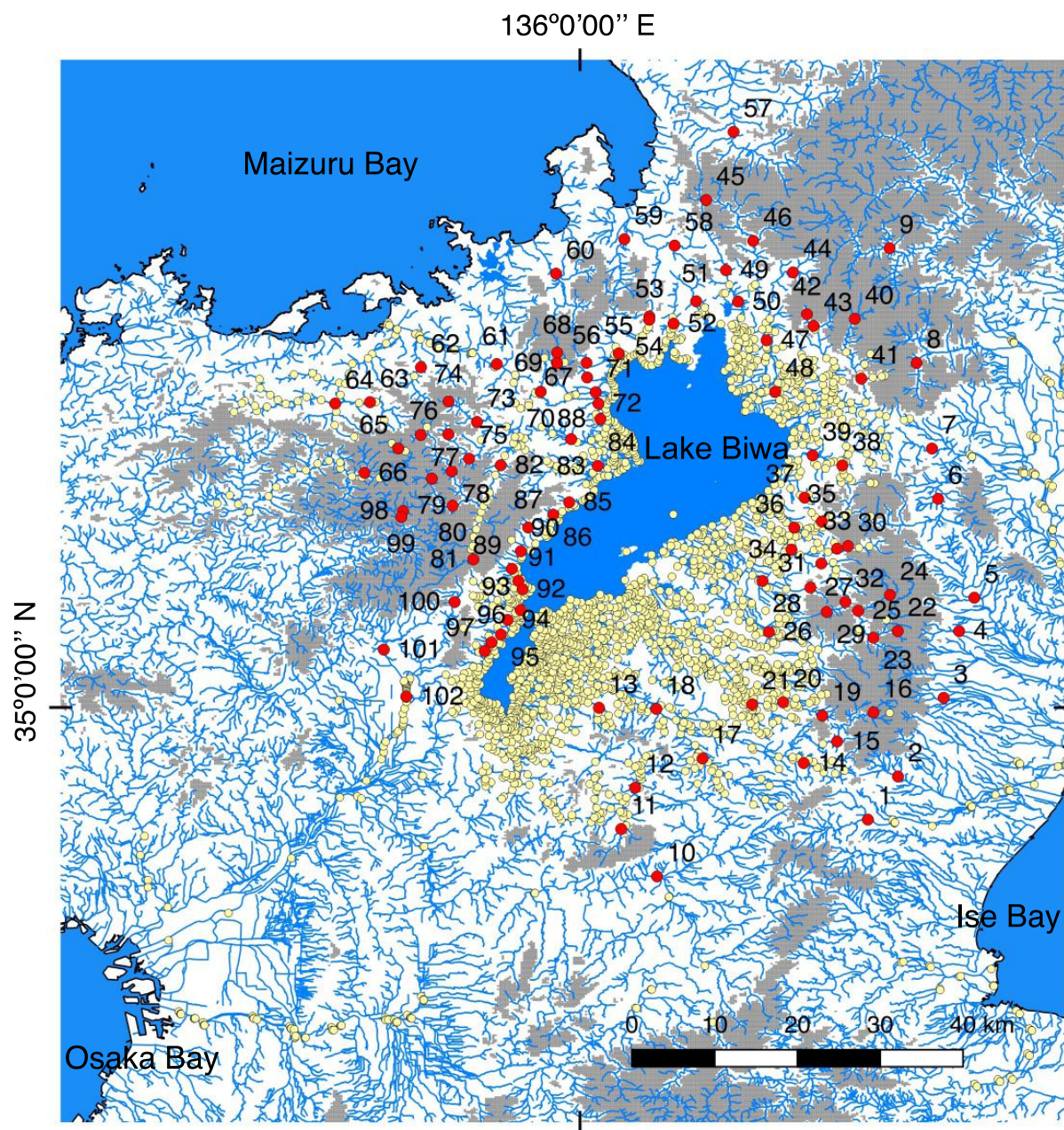
Figure 4. Results of a cluster analysis based on the presence/absence of each fish species inferred by eDNA and the observational data at each sampling site in (a) the full dataset and (b) the data subset of sites with ≥ 1000 sequence depth.

Figure 5. Detections of assigned eDNA sequences of (a) *Hemibarbus longirostris*, (b) *Niwaella delicata*, (c) *Cottus pollux*, and (d) *Cottus reinii* and observational data of (e) *Cottus pollux* and (f) *Cottus reinii* in the sampling area. Red circles indicate sampling sites with eDNA detections or observational records of focal species. Yellow circles indicate sampling sites without eDNA detections or observational records. Red shaded areas indicate watersheds with observational records of focal species (Kawanabe & Mizuno, 2001; Kitagawa *et al.*,

687 2001; Matsumiya *et al.*, 2001; Ministry of Land, Infrastructure, Transport and Tourism, Kinki
688 Regional Development Bureau, Biwako Office, 2004; Ministry of Land, Infrastructure,
689 Transport and Tourism, 2005; Uonokai, 2005; Nakagawa, 2014; H. Nakagawa, unpublished).
690
691

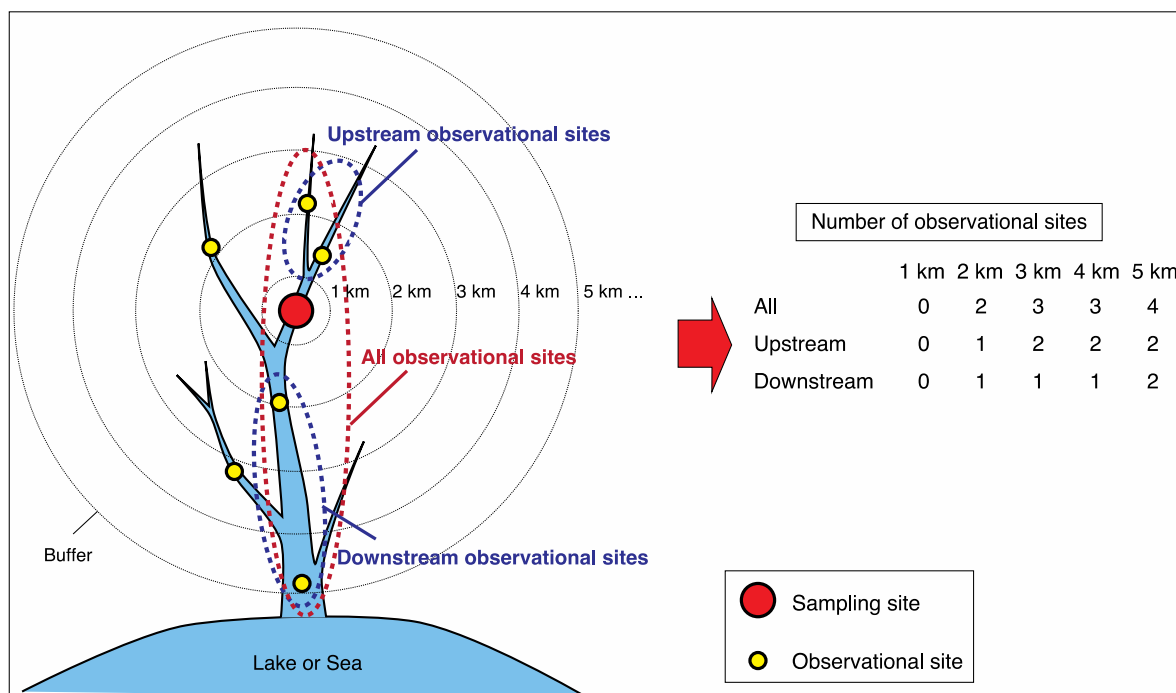
692 **Figures**

693 Figure 1.



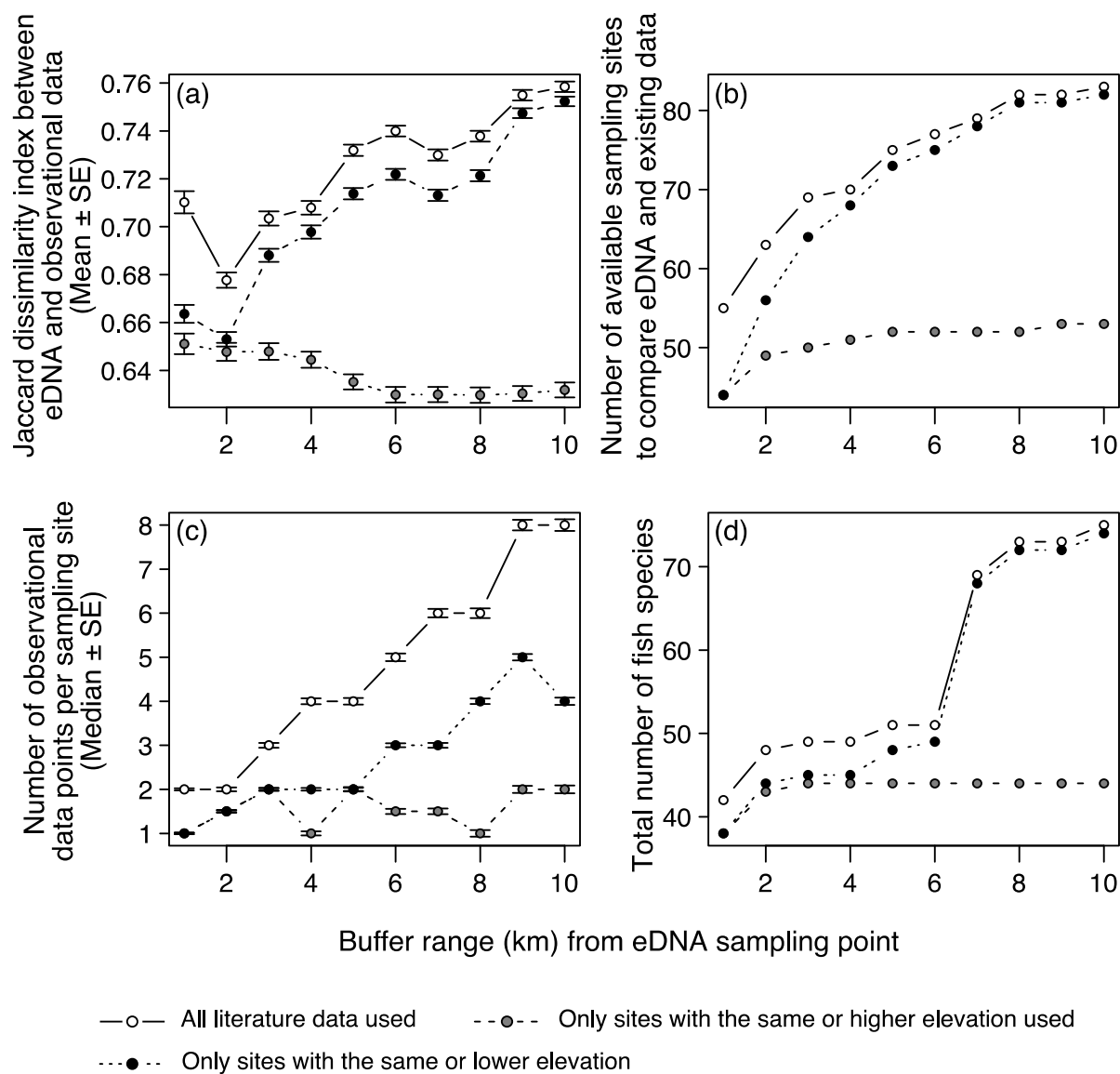
694

695 Figure 2.



696

697 Figure 3.

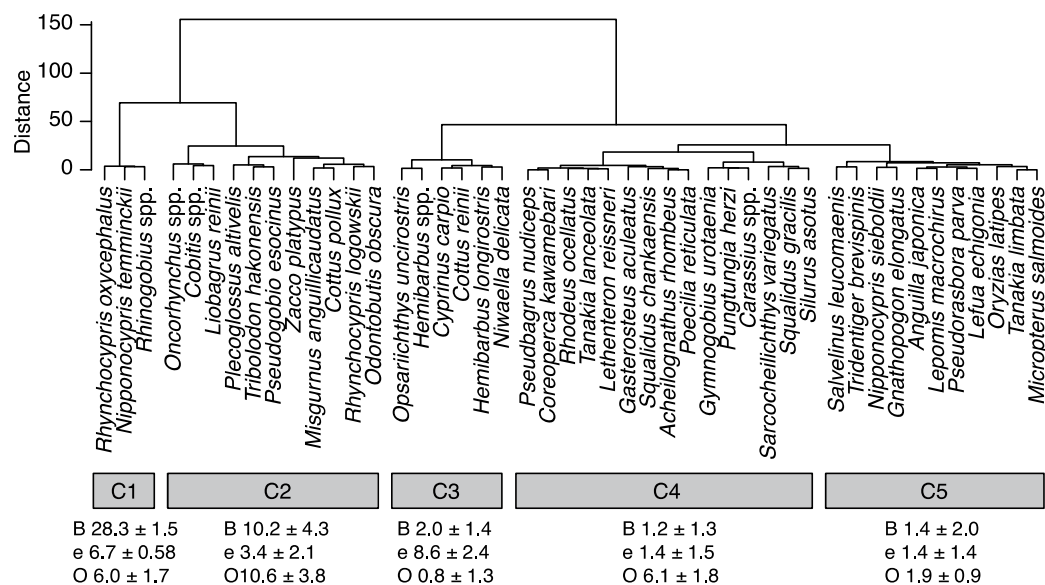


698

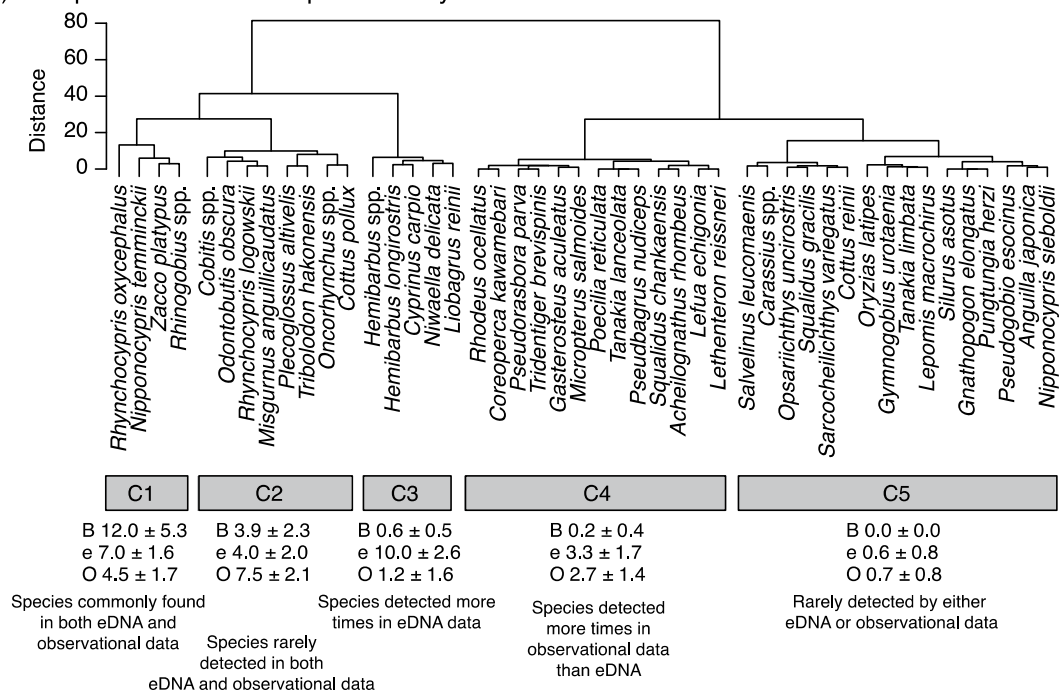
699

700 Figure 4.

(a) Full data set



(b) Samples with ≥ 1000 sequences only



B: Number of sampling sites where fish species was inferred by both eDNA and the observational data (mean ± SD).

e: Number of sampling sites where fish species was inferred by the observational data alone (mean ± SD).

O: Number of sampling sites where fish species was inferred by eDNA alone (mean ± SD).

702 Figure 5.

