Evaluation of biodiversity metrics through environmental DNA metabarcoding compared with visual and capture surveys in river fish community

Hideyuki Doi1,*,†, Ryutei Inui2,3,*,†, Shunsuke Matsuoka1,†, Yoshihisa Akamatsu2, Masuji Goto2, and Takanori Kono2

1 Graduate School of Simulation Studies, University of Hyogo, 7-1-28 Minatojima Minami-machi, Chuo-ku, Kobe, 650-0047, Japan
2 Graduate School of Science and Engineering, Yamaguchi University, 2-16-1 Tokiwadai, Ube, Yamaguchi, 755-8611, Japan
3 Faculty of Socio-Environmental Studies, Fukuoka Institute of Technology, 3-30-1 Wajiro-higashi, Higashi-ku, Fukuoka, 811-0295, Japan

†These authors equally contributed to this study.

*Corresponding authors:
Hideyuki Doi (hideyuki.doi@icloud.com)
Ryutei Inui (inuiryutei@gmail.com)

Running head: eDNA metabarcoding for fish diversity
Abstract

1. Information on alpha (local), beta (between habitats), and gamma (regional) diversity is fundamental to understanding biodiversity as well as the function and stability of community dynamics. The methods like environmental DNA (eDNA) metabarcoding are currently considered useful to investigate biodiversity.

2. We compared the performance of eDNA metabarcoding with visual and capture surveys in estimating alpha/gamma diversity and the variation of the community assemblages of river fish communities, particularly considering community nestedness and turnover.

3. In five rivers across west Japan, with comparing to visual/capture surveys, eDNA metabarcoding detected more species in the study sites, consequently the overall number of species in the region (i.e., gamma diversity) was higher. In particular, the species found by visual/capture surveys were encompassed by those by eDNA metabarcoding.

4. With analyzing the community assemblages between the rivers, we showed the different results between the both methods. While, in the same river, the nestedness and species turnover changing from upstream to downstream did not significantly differ between the both methods. Our results suggest that eDNA metabarcoding may be
suitable method, especially for understanding regional community patterns, for fish monitoring in rivers.

Key words: eDNA, community, river, alpha and gamma diversity, nestedness
Introduction

The maintenance of biodiversity underpins the stability of ecosystem processes in constantly changing environments (Primack, 1993; Margules & Pressey, 2000; Pecl et al., 2017). Moreover, biodiversity loss affects ecosystem functions and services and, consequently, human society (Primack 1993; Margules & Pressey, 2000; Pecl et al., 2017). Ecologists have made efforts to conserve biodiversity based on essential biodiversity survey methods, e.g., species richness and distribution (Primack, 1993; Margules & Pressey, 2000, Doi & Takahara, 2016, Pecl et al., 2017). Biodiversity can be evaluated in different levels: e.g., by estimating alpha (local), beta (between habitats), and gamma (regional) diversity and the variation of the community assemblages. To conserve local communities, ecologists incorporated these diversity measurements into management decision-making (Primack 1993; Margules & Pressey, 2000, Socolar et al., 2016). For example, the variation of the community assemblages can quantify biodiversity loss and inform the placement of protected areas and the management of biological invasions and landscapes (Socolar et al., 2016). Thus, robust methods for monitoring biodiversity are fundamental for biodiversity and environmental management.

Environmental DNA (eDNA) analysis is considered a useful tool to investigate the distribution and richness of aquatic and terrestrial organisms (Takahara et al., 2012, 2013; Rees et al., 2014; Goldberg et al., 2015; Miya et al., 2015; Thomsen & Willerslev, 2015; Doi et al., 2017; Doi et al., 2019; Fujii et al., 2019). High-throughput sequencing derived from eDNA, called “eDNA metabarcoding”, is an
exceptionally useful and powerful tool for community biodiversity surveys (Taberlet et al., 2012; Deiner et al., 2016, 2017; Sato et al., 2017; Bylemans et al., 2018; Fujii et al., 2019). eDNA metabarcoding has recently been applied in fish community surveys, e.g. Miya et al. (2015) designed and applied universal PCR primers (the MiFish primers) to survey marine fish communities. To confirm the usefulness of eDNA metabarcoding for community assessment, many studies compared eDNA metabarcoding to a species list generated by traditional surveys including visual and capture methods (Deiner et al., 2016; Bylemans et al., 2018; Nakagawa et al., 2018; Yamamoto et al., 2018; Fujii et al., 2019). However, evaluating the performance of eDNA metabarcoding estimating alpha diversity was still limited quantitively and statistically (but Drummond et al., 2015; Deiner et al., 2016; Staehr et al. 2016, Maechler et al. 2019), especially when gamma diversity and the variation of the community assemblages evaluating and comparing to traditional surveys.

Variation of the community assemblages is a fundamental aspect for communities, and it is important for evaluating community responses to environmental gradients and the spatial locations of community across different habitats and ecosystems (Chase et al., 2004; Chase, 2010; McCune & Vellend, 2013; Socolar et al., 2016). The variation itself and trends along environmental gradients, such as productivity and disturbance, can be influenced by alpha diversity (Chase, 2010). Furthermore, variation of the community assemblages can be influenced by alpha and gamma diversity (Olden & Poff, 2003; Van Calster et al., 2007; Keith et al., 2009). Therefore, the variation evaluation of community assemblages can vary considerably depending on differences in alpha and gamma diversity evaluation by different survey methods, i.e. eDNA metabarcoding vs. traditional surveys.
Variation of the community assemblages is considered to reflect two different components: nestedness and species turnover (Harrison et al., 1992; Baselga et al., 2007; Baselga, 2010). Nestedness occurs when the community at the sites with less species are subsets of the community at the sites with higher species richness (Wright & Reeves, 1992; Ulrich & Gotelli, 2007) and generally reflects a non-random process of species loss (Gaston & Blackburn, 2000). Contrastingly, species turnover implies the replacement of some species by others because of environmental sorting or spatial/historical constraints (Baselga, 2010). Statistical separation methods for nestedness and species turnover were applied for evaluating variation of the community assemblages in various systems (Baselga, 2010; Baselga et al., 2012). However, the method has never been applied to evaluate the performance of eDNA metabarcoding estimating community assemblages of fish species. Moreover, Baselga's (2010) framework can be applied to compare the performance among methods when evaluating alpha diversity via nestedness and species turnover.

Using statistical methods, we can quantitatively compare the performance of eDNA metabarcoding and traditional surveys for alpha/gamma diversity evaluation of biological communities and the variation of the community assemblages among the study sites. Here, we tested the performance of eDNA metabarcoding in five river systems in different regions with various fish species. We conducted eDNA metabarcoding using universal MiFish primers that target fish and identified the fish by visual snorkeling and hand-net capture surveys. We evaluated the performance of eDNA metabarcoding by comparing the obtained fish community structure to that evaluated by visual/capture survey with special regard to nestedness and species turnover.
Methods

Site description

In 2016, we conducted field surveys in five river systems across Japan (river map in Fig. S1): the Kyuragi River on October 10, the Koishiwara River on October 21, the Yato River on October 25, the Hazuki River on November 2, and the Oze River on November 6. The survey sites were set at a site at each of three river segments (upstream, mid-stream, and downstream, the internal distances ranged from 4.5 to 25.8 km, Fig. S1) for each river. Each site was set so that the length in the up-down direction was approximately 100 m with a riffle at the downstream end (e.g., Fig. S2).

Water collection for eDNA survey

In each site, we collected 1 L of surface water in bleached bottles at two points, the stream near the downstream end and the shore with static or semi-static water (Fig. S2) immediately before visual and capture surveys. eDNA was removed from the bottles and filtering equipment using 10% commercial bleach (ca. 0.6% hypochlorous acid) and washing with DNA-free distilled water. One milliliter of benzalkonium chloride (BAC, 10% w/v) was added per liter of water sample to avoid a decrease in eDNA concentration in the samples (Yamanaka et al., 2016). During transport, samples were stored in a cooler with ice packs. The ‘cooler blank’ contained 1 L DNA-free water, which we brought to the field and treated identically to the other water samples, except that it was not opened at the field sites.
Visual observation and capture methods

After water sampling, the fish fauna survey was conducted by visual observation with snorkel and collection with hand net. For visual observation, we observed and recorded fish species by snorkeling in a 100-m transect (snorkeling by 1 person for 1 h, Fig. S2). We observed at various micro habitats, including the riffle, pool, and shore bank from the downstream end to upstream end. We also conducted a hand-net capture survey (1 person × 1 h) using a D-frame net (2 mm mesh, net opening: 0.16 m²) in the various habitats in the river, including the riffle, pool, and shore bank. Fishes were identified according to Nakabo et al. (2013) at the survey site. We used the combined taxa list from both traditional surveys to compare to that of eDNA metabarcoding. In order to prevent contamination of eDNA samples, the investigator who collected and identified the fish and the investigator who sampled the water were different.

eDNA collection, extraction and measurements

Collected water samples were vacuum-filtered into GF/F glass filters (47 mm diameter, pore size: 0.7 µm, GE Healthcare, Little Chalfont, UK) in the laboratory within 24 h of sampling. After filtration, all filters were stored at -20 °C before eDNA extraction. The cooler blank was also processed in the same manner. A liter of Milli-Q water was used as the filtering blank to monitor contamination during filtering in each site and during subsequent DNA extraction.

To extract the DNA from the filters, we followed the methods described in Uchii, Doi, & Minamoto (2016). We incubated the filter by submerging the mixed buffer of 400 µL of Buffer AL in DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany) and 40 µL of Proteinase K (Qiagen, Hilden, Germany), using a Salivette tube.
Sarstedt, Nümbrecht, Germany) at 56 °C for 30 min. The Salivette tube with filters was centrifuged at 5000 × g for 5 min. Then, we added 220 µL of TE buffer (pH: 8.0) onto the filter and again centrifuged at 5000 × g for 5 min. The DNA was purified using a DNeasy Blood & Tissue Kit with extracted the DNA in 200 µL in Buffer AE. Samples were stored at -20 °C until the 1st-PCR assay.

Library preparation and MiSeq sequencing

The detailed molecular methods are described in Fujii et al. (2019) with a two-step PCR-procedure for Illumina MiSeq sequencing. Briefly, we performed 1st-PCR with MiFish-U-F and MiFish-U-R primers (Miya et al., 2015), which were designed to contain Illumina sequencing primer regions and 6-mer Ns;

Forward: 5′-ACACTCTTTCCCTACGACGCTCTTCCGATCTNNNNNN
GTCGGTAAAAACTCGCCAGC-3′,
Reverse: 5′-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNN
CATAGTGAGGTATCTAATCCCAGTTTG-3′

The italicized and non-italicized letters represent MiSeq sequencing primers and MiFish primers, respectively, and the six random bases (N) were used to enhance cluster separation on the flow cells during initial base call calibrations on the MiSeq (Miya et al. 2015, Doi et al. 2019).

We performed the 1st-PCR with a 12 µL reaction volume containing 1× PCR Buffer for KOD FX Neo (Toyobo, Osaka, Japan), 0.4 mM dNTP mix, 0.24 U KOD FX Neo polymerase, 0.3 µM of each primer, and 2 µL template. The thermocycling conditions for this step were as follows: initial denaturation at 94 °C for 2 min, followed by 35 cycles of denaturation at 98 °C for 10 s, annealing at 65 °C for 30 s, and...
elongation at 68 °C for 30 s, followed by final elongation at 68 °C for 5 min. The first
PCRs were performed using eight replicates (Doi et al. 2019) and individual first PCR
replicates were pooled and purified using AMPure XP (Beckman Coulter, Brea CA,
USA) as templates for the 2nd-PCR. The Illumina sequencing adaptors and the eight bp
identifier indices (XXXXXXXX) were added to the subsequent PCR process using a
forward and reverse fusion primer:

| Forward: 5′-AATGATACGGCGACCACCGAGATCTACA XXXXXXXX |
| ACACTCTTTCCCTACACGACGCTCTTCCGATCT-3′ |
| Reverse: 5′-CAAGCAGAAGACGGCATACGAGAT XXXXXXXX |
| GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT-3′ |

The italicized and non-italicized letters represent MiSeq P5/P7 adapter and
sequencing primers, respectively. The 8X bases represent dual-index sequences inserted
to identify different samples (Hamady 2008). We performed the 2nd-PCR with 12
cycles of a 12 µL reaction volume containing 1× KAPA HiFi HotStart ReadyMix, 0.3
µM of each primer, and 1.0 µL of the first PCR production. The thermocycling
conditions profile after an initial 3 min denaturation at 95 °C was as follows:
denaturation at 98 °C for 20 s, annealing, and extension combined at 72 °C (shuttle
PCR) for 15 s, with the final extension at the same temperature for 5 min. We
confirmed the positive bands of the targeted 1st-PCR amplicons by electrophoresis. The
2nd-PCR products were pooled in equal volumes and purified using AMPure XP.
The purified PCR products were loaded on a 2% E-Gel SizeSelect (Thermo
Fisher Scientific, Waltham, MA, USA) and the target size of the libraries
(approximately 370 bp) was collected. The samples concentration and quality were
estimated by a Qubit dsDNA HS assay kit and a Qubit 2.0 (Thermo Fisher Scientific).
The amplicon libraries were sequenced by 2 × 250 bp paired-end sequencing on the MiSeq platform using the MiSeq v2 Reagent Kit. Note that the sequencing run contained a total of 339 libraries including 40 of our library (30 samples plus five cooler and five filter negative controls) and 299 libraries from other research projects. The MiSeq sequencing was conducted in the Department of Environmental Solution Technology, Faculty of Science and Technology, Ryukoku University. All sequence data were deposited in DNA Data Bank of Japan (DRA, Accession number: DRA008090).

Bioinformatic analysis for MiSeq sequencing

The detailed procedures used for bioinformatics analysis are described in Fujii et al. (2019). First, low-quality tails were trimmed from each read and paired-end reads were then merged. For the obtained 1,823,446 reads, primer sequences were removed and identical sequences (i.e., 100% sequence similarity) were merged using UCLUST (usearch 7.0.1001, Edgar, 2010). The sequences with 10 or more identical reads were subjected to the downstream processes. To annotate the taxonomy, local BLASTN search using BLAST 2.2.29 was conducted with the reference database of fish species for processed reads (Miya et al., 2015). The top BLAST hit with a sequence identity ≥ 97% was applied to species detection of each sequence. Note that the species were mostly identified with a ≥ 99% match. From the BLAST results, we identified the species using methods previously described (Sato et al., 2017). Also, we detected the most of fish species inhabiting the rives in the regions with regarding to the known distributions (Kawanabe et al. 2001).
Statistical analyses

All statistical analyses and graphics were conducted in R ver. 3.4.4 (R Core Team, 2018). All statistics were set at the significance level $\alpha = 0.05$. To compare between eDNA metabarcoding and visual survey data, the taxonomic levels in the species list from visual survey were compared to the lists from eDNA metabarcoding (Table S1, S2) in reference to previous studies using the MiFish primer (Sato et al., 2017; Fujii et al., 2019). Before statistical analysis, we confirmed that the sequencing depth was sufficient to detect alpha diversity in the samples by “iNEXT” and “ggiNEXT” functions in the “iNEXT” ver. 2.0.19 package (Chao et al. 2014, Fig. S3). We merged the community data from two points, the stream near the downstream end and the shore, to compare with visual/capture surveys.

We tested the differences in fish richness of sites, segments, and rivers between both methods using generalized linear mixed models (GLMMs) with the “lmer” function in the “lme4” ver. 1.1-pacakge (Bates et al. 2015). In the GLMM models, the method was treated as a fixed effect with Poisson distribution, and the rivers and segments were treated as random effects.

The differences in community compositions were visualized using nonmetric multidimensional scaling (NMDS) with 500 separate runs of real data. For NMDS, the community dissimilarity was calculated based on incidence-based Jaccard indices. We evaluated the differences in community structures between methods and sites using permutational multivariate analysis of variance (PERMANOVA). For PERMANOVA, we used Jaccard and Raup-Crick similarity matrix and calculated the statistical values with 999 permutations. Raup-Crick index is the probability that compared the sampling sites have non-identical species composition (i.e., considering the alpha diversity of
sites). We used “metaMDS” and “adonis” functions in the “vegan” ver. 2.5-6 package (https://github.com/vegandevs/vegan) for NMDS ordination and PERMANOVA, respectively.

For the communities evaluated by both methods, indicator taxa analysis (Cáceres & Legendre, 2009) was performed to determine which taxa had significantly different frequencies between both methods. The analysis was performed using the “signassoc” function in the “indicspecies” ver. 1.7.8 package on the present/absence data for the testing with regarding the package description and Cáceres & Legendre (2009). The “signassoc” function can calculate the index with both present/absence and abundance data. We used mode = 1 (group-based) and calculated the P-values with 999 permutations after Sidak’s correction of the multiple testing.

To compare community composition (i.e., the variation of the community assemblages) focusing on nested and turnover structures between eDNA metabarcoding and visual/capture survey, we calculated the pairwise indices for nestedness and species turnover (Baselga, 2010; Baselga, 2012) using the “beta.pair” function in the “betapart” ver. 1.5.1 package (Baselga and Orme 2012). Standardized effect sizes (SESs) of nestedness and turnover indices were calculated to show the degree of nestedness and turnover structure. The significance was defined by deviation from zero, and the expectation of random assembly (a null model) was estimated with 999 random sampling replicates. The SES was defined as follows: ($\beta_{\text{obs}} - \beta_{\text{null}}$)/$\beta_{sd}$, where $\beta_{\text{obs}}$ is the observed beta diversity (here, the variation of the community assemblages among the sites), $\beta_{null}$ is the mean of the null distribution of beta diversity, and $\beta_{sd}$ is the standard deviation of the null distribution. SES values greater than zero indicate statistically stronger nestedness or turnover structure than expected under a random model of
community assembly, while negative values indicate weaker nestedness or turnover than expected. The randomized community data were generated with independent swap algorithm (Gotelli, 2000) using “randomizeMatrix” function in the “picante” ver. 1.8 package (Kembel et al. 2010). First, to evaluate the differences in nestedness and species turnover between the survey methods (eDNA metabarcoding vs. visual/capture survey) at the same segments, the SES of pairwise nestedness and turnover were calculated for each sample pair within the river. Then, the fish community longitudinal nestedness and turnover structure along with river flow (i.e., upstream to downstream) by each method was evaluated with NODF and pairwise indices of nestedness. First, a nestedness metric (NODF, Almeida-Neto et al., 2008) and their SES value were calculated with 999 randomizations using the "nestednodf" and the “oecosimu” function in the “vegan” package. Then, we tested the differences in SES of pairwise indices between the survey methods (eDNA metabarcoding vs. visual/capture survey) by generalized linear mixed model (GLMM, with Gaussian distribution) with the “lmer” function in the “lme4” package. In the GLMM models, the SES of pairwise indices was treated as a fixed effect, and the rivers-pairs and segment-pairs (i.e., three pairs in each river) were treated as random effects. For the map graphics in Supplemental Materials, we used the stream and coast line data from the National Land Information Division, Ministry of Land, Infrastructure, Transport and Tourism of Japan (http://nlftp.mlit.go.jp/ksj/).

**Results**

**Overview**
We detected 53 fish taxa, almost all identified to the species or genus level, by eDNA metabarcoding in five rivers (Table S1, and S2) and visually observed 38 fish taxa in total. MiSeq paired-end sequencing for the library (30 samples plus five cooler and five filter negative controls) yielded a total of 1,601,816 reads (53,351 ± 17,639; mean ± S. D. for each sample, Table S2). We confirmed very low reads from negative controls (Table S2) with only detecting a fish species, Tridentiger sp. in the blank of the Yato River, probably because of the cross-contamination among the samples. The read of Tridentiger sp. in the samples (4783-9818 reads) were high than the detected sites (1286 reads), therefore we used the species for the analysis.

Diversity indices between methods

We found significant differences in fish local richness (alpha diversity) between eDNA metabarcoding and visual/capture surveys (Fig. 1, GLMM with river as a random effect, t = –5.45, P = 0.000018). Richness was significantly different among river segments (t = –5.85, P = 0.000004), indicating higher alpha diversity estimated by eDNA metabarcoding than by visual/capture surveys. While, richness was not significantly different among rivers (t = 1.737, P = 0.0942).

We found differences in community structure between the two methods by NMDS ordination by Jaccard index (Fig. 2) as well as Raup-Crick (Fig. S4), especially among the study rivers. The patterns of differences in the both Jaccard and Raup-Crick indices were similar.

The PERMANOVA results with Jaccard and Raup-Crick indices for the ordination suggested there were differences in community composition evaluated by each method, eDNA metabarcoding and visual/capture survey (P < 0.012, Table S3).
Moreover, communities from the combined results of eDNA metabarcoding and visual/capture survey were significantly different among rivers and segments ($P < 0.029$, Table S3). We found different patterns in ordered river sites for each method (Fig. 2 for Jaccard index, Fig. S4 for Raup-Crick index).

To test the community differences among the segments and rivers, the PERMANOVA results with Jaccard index determined that communities were significantly different among the rivers by eDNA metabarcoding ($P = 0.001$) but not by visual/capture survey ($P = 0.12$, Table S4). Conversely, the communities were significantly different among river segments by visual/capture survey ($P = 0.011$) but not significantly different (albeit marginally) by eDNA metabarcoding ($P = 0.061$, Table S4). The differences in PERMANOVA results suggest that differences in the variation of the community assemblages among rivers across regions can be detected by eDNA metabarcoding but not by visual/capture survey. While, the PERMANOVA results with Raup-Crick index determined that communities were significantly different among the rivers by the both eDNA metabarcoding ($P = 0.001$) and visual/capture survey ($P = 0.018$, Table S4).

Indicator taxa analysis comparing the communities estimated by both methods, eDNA metabarcoding and visual/capture survey, detected capture bias by both methods for several taxa, including Japanese eel (*Anguilla japonica*), salmon (e.g., *Oncorhynchus masou*), and Amur catfish (*Silurus asotus*) ($P < 0.05$, Table 1 for statistically significant taxa, Table S5 for all taxa).

Nestedness and species turnover
We compared the pairwise standardized effect size (pSES) between eDNA metabarcoding and visual/capture survey in nestedness and species turnover (Fig. 3). Nestedness pSES was significantly positive without overlapping the zero-pSES 95% confidence interval, indicating that the visual/capture survey community was nested in that detected by eDNA metabarcoding. The significantly negative pSES in species turnover indicated that the taxa did not significantly turnover in the community evaluated by eDNA metabarcoding and visual/capture survey compared to that in random communities in the segment-scale comparisons.

We compared the longitudinal nested structure (upstream to downstream) of the fish community in the study rivers using nested pSES (Fig. 4a). The longitudinal nested structure was not significantly different between both methods (GLMM, $P = 0.302$). The positive pSES indicated that downstream communities were significantly nested in the upstream ones. The species per segment were nested as downstream > mid-stream > upstream both methods (Fig. S5 and S6, NOFD, $P < 0.001$). We also compared the longitudinal species turnover of fish communities in the study rivers using turnover pSES (Fig. 4b). The longitudinal species turnover was not significantly different between both methods (GLMM, $P = 0.280$) and the negative turnover values indicated no turnover observed.

**Discussion**

We found that river fish communities estimated by eDNA metabarcoding were significantly nested in the communities estimated by visual/capture survey. Furthermore, the species turnover in the communities was very weak between methods.
Nestedness of species assemblages occurs when the communities obtained by the method estimating lower number of species are subsets of the communities estimated by other methods with higher species richness (Baselga 2010). The eDNA metabarcoding estimated higher alpha diversity than visual/capture survey for river fish communities and provide the similar gamma diversity and species composition with visual/capture survey. In fact, the fish local and regional richness (alpha and gamma diversity) evaluated by eDNA metabarcoding was significantly higher, including almost all taxa evaluated by visual/capture survey.

eDNA metabarcoding has been reported to perform better than traditional methods in evaluating species richness (Deiner et al., 2016; Sato et al., 2017; Bylemans et al., 2018; Nakagawa et al., 2018; Yamamoto et al., 2018; Fujii et al., 2019).

Nakagawa et al. (2018) investigated freshwater fish communities in 100 rivers and confirmed that the community detected by eDNA metabarcoding were similar to the species lists observed in government-authorized monitoring. Furthermore, several eDNA metabarcoding studies on fish communities have been performed in other river systems (Bylemans et al., 2018), marine habitats (Yamamoto et al., 2018), and freshwater lakes (Sato et al., 2017; Fujii et al., 2019). Deiner et al. (2016) showed that river eDNA metabarcoding can reflect the community in a watershed, indicating that eDNA metabarcoding has high performance for gamma diversity evaluation. These studies indicated the great potential of eDNA metabarcoding as a useful tool for alpha and gamma diversity assessment by simply comparing the community data obtained from eDNA metabarcoding and traditional surveys. However, previous studies did not evaluate performance in terms of nestedness and species turnover between eDNA metabarcoding and other community data. Here, we support the previous literature.
(Maechler et al. 2019) by showing patterns observed in alpha and gamma diversity are also observed in nestedness, species turnover, and the capture bias of communities detected by eDNA metabarcoding and traditional methods in segment scale. With comparing to the same segment of the same river we found that the community detected by visual/capture survey was nested with that by eDNA metabarcoding, with scarce species turnover in the community.

We especially focused on the variation of the community assemblages evaluated by eDNA metabarcoding compared to visual/capture survey. For evaluating the fish communities, the community compositions were different between eDNA metabarcoding and visual/capture survey. That is, eDNA metabarcoding detected significant differences between the rivers, while visual/capture survey detected significant differences between the segments rather than between the rivers. On the other hand, with focusing on a single river, the patterns of nestedness and species turnover were not significantly different between eDNA metabarcoding and visual/capture survey.

Using Jaccard index, a higher variation of the river fish communities was statistically detected by eDNA metabarcoding than by visual/capturing survey and the variation of the community assemblages between segments could be significantly detected by visual/capture surveys but not by eDNA metabarcoding. While the results of Raup-Crick index, considering the alpha diversity of sites, showed the same results of the both methods. This might suggest that Jaccard dissimilarity detect significant differences between the segments for visual/capturing survey due to the low number of species detected in the upstream. These differences in the variation of the community assemblage evaluation may lead us to interpret the variation of the community
assemblages using the results from both survey methods, eDNA and visual/capture survey.

We further compared the indicator taxa for the communities obtained from both eDNA metabarcoding and visual/capture survey and concluded that several taxa, including eel, salmon, and catfish, were significantly better detected by eDNA metabarcoding, whereas non-indicator taxa were detected by visual/capture surveys. These results indicated that eDNA metabarcoding had higher detection frequency of visual/capture surveys in fish taxa detection. The community structures estimated by eDNA metabarcoding and visual/capture survey were slightly different, as reported in previous studies (e.g., Sato et al., 2017; Fujii et al., 2019), probably because of the differences in taxa-detection performances. Discriminated taxa in this analysis included eel, salmon, and catfish, which mostly had larger body size and lower abundances in these rivers (Kawanabe, 2001; Nakabo, 2013). In fact, the Japanese eel *Anguilla japonica*, was difficult to find by visual observation, probably due to its hiding behavior (Itakura et al., 2019). Such endangered species would be important as top predators (Nakabo, 2013). eDNA metabarcoding can evaluate the distribution of such rare and important taxa in fish communities better than traditional surveys. While we did not detect any indicator taxa by visual/capture surveys, we found couple species detected only by visual/capture surveys, for example, *Lepomis macrochirus* and *Biwia zezera* (see Table S1). *Lepomis macrochirus* usually inhabit in lotic systems and rarely observed in rivers (Nakabo, 2013). *Biwia zezera* distributed in this region (Hosokawa et al., 2007). We only observed the both species in a downstream segment of Kyuragi and Oze river, respectively, thus, this species was rarely observed in this study. Also, the BLAST identification for *L. macrochirus* and *B. zezera* were preliminary confirmed
Therefore, we speculated that we could not capture the eDNA of *L. macrochirus* and *B. zezera*, due to the sampling timing or points. Further study needs to the species detection which we captured, especially rarely-observed species.

To evaluate the comparison among the local sites and rivers, we examined the performances of eDNA metabarcoding at three sites of five rivers with eDNA sampling from only two habitats. Thus, our understanding of some aspects of the fish community spatial structure in the rivers and the performance of community evaluation in local habitats, such as backwater, was still limited. In fact, Bylemans et al. (2018) found that river morphology in these habitats influenced the optimal sampling strategy for eDNA metabarcoding. Moreover, in backwater lakes, the performance of eDNA metabarcoding varied with different lake morphologies (Fujii et al. 2019). However, testing the usefulness of the eDNA metabarcoding for assessing river fish community biodiversity has been limited. Further research is needed to evaluate fish community spatial structure in rivers. In addition, we should consider that eDNA recovered from a water sample came from an individual in the survey area. The previous studies suggested the eDNA came from the upstream (Deiner, K., & Altermatt 2014; Deiner et al. 2016). Therefore, the comparisons between eDNA metabarcoding and visual/capture methods were using the community data with different spatial scales, for example, community data from a 100-m reach by visual/capture methods, but eDNA potentially detected the community in larger area than surveyed. We should carefully consider the phenomenon, especially for alpha diversity of community.
In conclusion, eDNA metabarcoding have similar species detection performance with the visual/capture survey in evaluating the community among the rivers. The eDNA metabarcoding was much less effort in the field and detected the community in broader area than visual/capture survey, therefore, eDNA metabarcoding may be suitable method, especially for regional community patterns. Biodiversity testing using statistical frameworks, especially community nestedness and turnover, provided the quantitative evidence to compare the performance of eDNA metabarcoding and traditional surveys. eDNA methods for biodiversity assessment may provide more information to us, as shown here, but we should also pay attention to the unknown characteristics of eDNA, such as the origins, degradation, and transport of eDNA in water which are still unknown in various habitats (Barnes & Turner, 2016; Seymour et al., 2018). To routinely use eDNA methods to assess biodiversity, we encourage testing the effect of eDNA degradation and transport on the performance of eDNA methods for biodiversity evaluation.

Acknowledgments

MiSeq sequencing was conducted in the Department of Environmental Solution Technology, Faculty of Science and Technology, Ryukoku University, and we thank to H. Yamanaka and H. Sato for supporting the experiments involving MiSeq sequencing.

This study was supported by the Environment Research and Technology Development Fund (4-1602) of Environmental Restoration and Conservation Agency, Japan and JST-CREST (JPMJCR13A2).

Data availability
All data of the MiSeq sequencing was shared in DRA (Accession number: DRA008090), and the observed species data was shared in Table S2 in Supplemental Materials.

**Author contributions**

HD, RI, and YA designed the study, RI, MG, TK, and YA contributed to field survey and sampling. SM, HD, RI, and MG contributed to molecular experiments. SM and HD analyzed the data and interpreted the results. HD, SM, and RI wrote the initial draft of the manuscript. All other authors critically reviewed the manuscript.

**References**


Kawanabe, H., Mizuno, N. & K. Hosoya (eds.) (2001) Freshwater Fishes of Japan
Yama-Kei, Tokyo (in Japanese).

Taxonomic homogenization of woodland plant communities over 70 years.
Proceedings of the Royal Society of London. Series B, Biological science, 276,
3539–3544.

Kembel, S. W., Cowan, P. D., Helmus, M. R., Cornwell, W. K., Morlon, H., Ackerly,
Bioinformatics, 26, 1463-1464.

405, 243.

Maechler, E., Little, C.J., Wuethrich, R., Alther, R., Fronhofer, E.A., Gounand, I.,
Assessing different components of diversity across a river network using eDNA.
Environmental DNA, 1, 290-301.


Miya, M., Sato, Y., Fukunaga, T., Sado, T., Poulsen, J. Y., Sato, K., ... Kondoh, M.


Yamanaka, H., Motozawa, H., Tsuji, S., Miyazawa, R. C., Takahara, T., & Minamoto, T. (2016). On-site filtration of water samples for environmental DNA analysis to avoid DNA degradation during transportation. Ecological Research, 31(6), 963–967.
Table 1 Indicator taxa analysis for the taxa had significantly different frequency between eDNA metabarcoding and visual/capture methods (P < 0.05). Best means preferred methods. P-values was calculated with 999 permutations after Sidak’s correction of the multiple testing.

<table>
<thead>
<tr>
<th>Taxa</th>
<th>P-value for eDNA metabarcoding</th>
<th>P-value for visual and capturing survey</th>
<th>best</th>
<th>P-value for multiple testing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anguilla japonica</td>
<td>0.001</td>
<td>1.000</td>
<td>eDNA metabarcoding</td>
<td>0.002</td>
</tr>
<tr>
<td>Oncorhynchus masou</td>
<td>0.001</td>
<td>1.000</td>
<td>eDNA metabarcoding</td>
<td>0.002</td>
</tr>
<tr>
<td>Oncorhynchus mykiss</td>
<td>0.016</td>
<td>1.000</td>
<td>eDNA metabarcoding</td>
<td>0.032</td>
</tr>
<tr>
<td>Cyprinus carpio</td>
<td>0.001</td>
<td>1.000</td>
<td>eDNA metabarcoding</td>
<td>0.002</td>
</tr>
<tr>
<td>Silurus asotus</td>
<td>0.003</td>
<td>1.000</td>
<td>eDNA metabarcoding</td>
<td>0.006</td>
</tr>
<tr>
<td>Misgurnus anguillicaudatus</td>
<td>0.009</td>
<td>1.000</td>
<td>eDNA metabarcoding</td>
<td>0.018</td>
</tr>
</tbody>
</table>
**Figure legends**

Figure 1. Venn diagrams for the number of detected taxa. eDNA metabarcoding (eDNA, blue) and visual/capture methods (V/C, red) in each site of the five study river systems. The bottom diagrams (All) showed the number of detected taxa in the river systems. The numbers represent the number of taxa.

Figure 2. Nonmetric multidimensional scaling (NMDS) ordination with Jaccard index (with Raup-Crick index in Fig. S4). Fish communities evaluated by the study rivers (shape) and each segment (colored) of the river. MDS stress was 0.158.

Figure 3. Pairwise effect size (SES, standard effect size). Comparison between community nestedness and species turnover between the communities detected by eDNA metabarcoding and visual/capture methods. Error bars indicate 95% confidence interval. The horizontal dotted line represents SES = 0, indicating a non-significant effect.

Figure 4. Effect size (SES, standard effect size). a) community nestedness and b) species turnover SES among the segment communities evaluated by eDNA metabarcoding and visual/capture methods. The error bars indicate 95% confidence interval. The horizontal dotted line represents SES = 0, indicating non-significant effect.
Figure 1
Figure 2

eDNA metabarcoding vs. Visual/capture survey:

- **eDNA metabarcoding**
  - NMDS1 vs. NMDS2 plot

- **Visual/capture survey**
  - Site markers:
    - Hazuki
    - Koishiwara
    - Kyuragi
    - Oze
    - Yato
  - Segment markers:
    - Downstream
    - Midstream
    - Upstream

The diagram illustrates the comparison between eDNA metabarcoding and visual/capture survey methods, showing the distribution of sites and segments in a multidimensional scaling (NMDS) plot.
Figure 3

Pairwise standardized effect size (pSES)
Figure 4