An efficient early-pooling protocol for environmental DNA metabarcoding

AUTHOR(S):

Ushio, Masayuki; Furukawa, Saori; Murakami, Hiroaki; Masuda, Reiji; Nagano, Atsushi J.

CITATION:

Ushio, Masayuki ...[et al]. An efficient early-pooling protocol for environmental DNA metabarcoding. Environmental DNA 2022, 4(6): 1212-1228

ISSUE DATE:

2022-11

URL:

http://hdl.handle.net/2433/278932

Received: 18 February 2022 | Revised: 25 May 2022 | Accepted: 21 June 2022

DOI: 10.1002/edn3.337

Environmental DNA
Open Access
Dedicated to the study and use of environmental DNA for basic and applied sciences
WILEY

METHOD

An efficient early-pooling protocol for environmental DNA metabarcoding

Masayuki Ushio[1,2] | Saori Furukawa[2,3] | Hiroaki Murakami[4,5] | Reiji Masuda[6] | Atsushi J. Nagano[3,7]

[1]Hakubi Center, Kyoto University, Kyoto, Japan

[2]Center for Ecological Research, Kyoto University, Otsu, Japan

[3]Faculty of Agriculture, Ryukoku University, Otsu, Japan

[4]Center for the Promotion of Interdisciplinary Education and Research, Kyoto University, Kyoto, Japan

[5]Graduate School of Agricultural Science, Tohoku University, Sendai, Japan

[6]Maizuru Fisheries Research Station, Field Science Education and Research Center, Kyoto University, Kyoto, Japan

[7]Institute for Advanced Biosciences, Keio University, Tsuruoka, Japan

Correspondence
Masayuki Ushio, Hakubi Center, Kyoto University, Kyoto 606-8501, Japan.
Email: ong8181@gmail.com

Atsushi J. Nagano, Faculty of Agriculture, Ryukoku University, Otsu 520-2194, Japan/Institute for Advanced Biosciences, Keio University, Tsuruoka 997-0017, Japan.
Email: anagano1234@gmail.com

Funding information
Hakubi Project, Kyoto University; Japan Society for the Promotion of Science (JSPS), Grant/Award Number: KAKENHI (B) 20H03323

Abstract

Environmental DNA (eDNA) metabarcoding, a method that applies high-throughput sequencing and universal primer sets to eDNA analysis, has been a promising approach for efficient, comprehensive biodiversity monitoring. However, significant money-, labor-, and time-costs are still required for performing eDNA metabarcoding. In this study, we assessed the performance of an "early-pooling" protocol (a protocol based on 1st PCR tagging) to reduce the experimental costs of library preparation for eDNA metabarcoding. Specifically, we performed three experiments to investigate the effects of 1st PCR-tagging and 2nd PCR-indexing protocols on the community composition revealed by eDNA metabarcoding, the effects of post-1st PCR exonuclease purification on tag jumping (corresponds to index hopping in 2nd PCR indexing), and the effects of the number of PCR replicates and the eDNA template volume on the number of detected OTUs. Analyses of 204 eDNA libraries from three natural aquatic ecosystems and one mock eDNA sample showed that (i) 1st PCR tagging does not cause clear biases in the outcomes of eDNA metabarcoding, (ii) post-1st PCR exonuclease purification reduces the risk of tag jumping, and (iii) increasing the eDNA template volume may increase the number of detected OTUs and reduce variations in the detected community compositions, similar to increasing the number of 1st PCR replicates. Our results show that an early-pooling protocol with post-1st PCR exonuclease purification and an increased amount of the DNA template reduces the risk of tag jumping, the costs for consumables and reagents (except for many tagged 1st PCR primers), and the handling time in library preparation, and produces similar results to a 2nd PCR-indexing protocol. Therefore, once a target metabarcoding region is selected and a set of tagged-1st PCR primers is prepared, the early-pooling protocol provides a cost, labor, and time-efficient approach for processing a large number of samples.

KEYWORDS
amplicon sequencing, biodiversity, early pooling, environmental DNA, exonuclease, index hopping, library preparation, metabarcoding, tag jumping

Masayuki Ushio and Saori Furukawa contributed equally to this work

1 | INTRODUCTION

Frequent and comprehensive ecosystem monitoring is a basis for effective biodiversity conservation, resource management, and near-future forecasting. Direct visual census, direct capture (e.g., fishing and insect collection), and camera/video trapping (e.g., for forest mammals) have traditionally been used as tools for biodiversity monitoring (Masuda et al., 2010; Nakagawa, 2019; Samejima et al., 2012). These data are invaluable and contribute to better understanding, conservation, and management of ecosystems under intense human pressures. However, these methods are usually time-consuming and require professional expertise such as taxonomic identification skill, which prevents their application to a frequent, large spatial-scale biodiversity monitoring.

Environmental DNA (eDNA), namely DNA isolated from environmental samples without capturing target organisms, has been used to detect the presence of macro-organisms (e.g., Miya et al., 2015; Taberlet et al., 2012, 2018; Yamamoto et al., 2017). In the case of macro-organisms, eDNA originates from various sources such as metabolic waste or damaged tissue (Kelly, Port, Yamahara, Martone, et al., 2014b), and the eDNA contains information about the species identity of organisms that produced it. Since the first application of eDNA analysis to natural ecosystems (Ficetola et al., 2008), eDNA has been used in many studies as a tool for investigation of the distributions of fish species in ponds, rivers, and seawater (Jerde et al., 2011; Minamoto et al., 2011; Sigsgaard et al., 2015; Ushio et al., 2018), as well as the distributions of other aquatic/semi-aquatic/terrestrial organisms (Bista et al., 2017; Deiner et al., 2016; Ishige et al., 2017; Ushio et al., 2017; Yonezawa et al., 2020).

eDNA metabarcoding, a method that applies universal primer sets and high-throughput sequencing to eDNA analyses, has now been widely used for comprehensive sequencing of target metabarcoding regions in an eDNA sample (Kelly, Port, Yamahara, & Crowder, 2014a; Miya et al., 2015; Taberlet et al., 2012; Ushio, 2022; Ushio et al., 2017; Yamamoto et al., 2017). For example, a previous study demonstrated that an eDNA metabarcoding using fish-targeting universal primers (MiFish primers) enabled the detection of more than 230 fish species from seawater in a single study (Miya et al., 2015). Another study demonstrated that eDNA metabarcoding can detect nearly 300 families from river eDNA samples and that the families contain both aquatic and terrestrial organisms, which suggested that a river collects and transports eDNA from surrounding area and that river eDNA can provide a large amount of information about the biodiversity in the watershed (Deiner et al., 2016).

Although eDNA metabarcoding is a powerful and promising method for comprehensive and efficient biodiversity monitoring, it has significant costs. While sequencing costs have been continuously declining over the past few decades (for example, see https://www.genome.gov/about-genomics/fact-sheets/DNA-Sequencing-Costs-Data), sample collection, DNA extraction, and library preparation are still laborious and time-consuming. To reduce the costs of eDNA metabarcoding, researchers have been developing technologies that enable automated sampling and rapid DNA extraction (Fukuzawa et al., 2022; Yamahara et al., 2019). However, the commonly used 2-step PCR protocol for eDNA library preparation remains laborious (e.g., Minamoto et al., 2021), and researchers have begun to develop an efficient eDNA library preparation protocol (Buchner et al., 2021).

The 2-step PCR protocol includes the first-round PCR (1st PCR) that amplifies a target metabarcoding region using a universal primer (e.g., mitochondrial 12S region by MiFish primers), followed by the second-round PCR (2nd PCR) that appends sample-specific sequences to amplicons. After the 2nd PCR, multiple samples may be combined (i.e., multiplexing or "pooling") because samples may subsequently be distinguished by the sample-specific sequences. However, until sample-specific sequences are appended, each sample needs to be separately processed, and careful operations are required to prevent cross-contamination. Alternatively, a sample-specific sequence may be appended at the 1st PCR. The sample-specific sequence is termed "index" when it is appended just outside sequence primers, or "tag" when it is appended just outside a universal primer, of which terminology we employed in this study according to Bohmann et al. (2022). The approach that appends the tag/index at the 1st PCR has been used in metabarcoding studies targeting prokaryotes (Caporaso et al., 2011; Smets et al., 2016) and eukaryotes (Bohmann et al., 2022; Leray & Knowlton, 2017; Minardi et al., 2022; Zizka et al., 2019). Although the preparation of many 1st PCR primers with sample-specific sequences is costly, this protocol enables multiplexing after the 1st PCR (i.e., "early-pooling") and may be efficient and suitable for large-scale metabarcoding. However, such an approach may introduce tag-/index-specific amplification biases (Berry et al., 2011; O'Donnell et al., 2016). In addition, if samples are pooled without the inactivation of primers and enzymes in the reaction, early pooling may increase the risk of tag jumping or index hopping because inter-sample chimeras may form due to the amplification of pooled amplicons with primers with different tags/indices (see Bohmann et al., 2022; Snyder & Stepien, 2020).

Furthermore, because the concentration of target eDNA is often low, especially for macro-organisms (e.g., fish), researchers have recommended preparing multiple technical replicates at the 1st PCR to maximize the species detection probabilities (e.g., >4–8 replicates; Doi et al., 2019), which inevitably increases the handling time and costs for plastic consumables and reagents. Increasing the number of the 1st PCR replicates, however, increases the total volume of template eDNA used in the PCR. For example, eight replicates of a 1st PCR which uses 1 μl of template eDNA per replicate uses 8 μl of template eDNA in total. Therefore, although increasing the number of the 1st PCR replicates is currently recommended, it is still unclear whether the thus-far observed patterns that the species detection probability increased with the number of replicates employed in the 1st PCR is due to the increased number of replicates or the increased volume of template DNA.

Because techniques in molecular biology have been continuously improving, whether common recommendations still hold for the current situation and how outcomes change using updated reagents and protocols are often unclear. Therefore, we herein assessed the

performance of an "early-pooling" protocol (i.e., a protocol based on tagging at the 1st PCR) to reduce the experimental costs of library preparation for eDNA metabarcoding. Using tagged 1st PCR primers, updated reagents, and several customized protocols, we performed fish eDNA metabarcoding for natural and mock samples and compared the results obtained with those from a common 2nd PCR-indexing protocol. Specifically, we conducted three experiments to test the effects of (i) tagging/indexing methods on the community composition revealed by eDNA metabarcoding (Experiment I), (ii) post-PCR exonuclease purification on tag jumping (Experiment II), and (iii) the number of PCR replicates and volumes of the eDNA template on the number of detected OTUs (Experiment III). Based on the experiments, we discuss the advantages and disadvantage of the early-pooling protocol and propose an efficient protocol for eDNA metabarcoding library preparation which is applicable to a large number of eDNA samples.

2 | MATERIALS AND METHODS

2.1 | Safeguarding against potential contaminations during the sampling and experiments

Prior to DNA extraction and library preparation, the work-space was sterilized using RNAase Quiet (Nacalai Tesque, Kyoto, Japan) and 70% EtOH and equipment was sterilized with a UV light or 10% commercial bleach. We used filtered pipette tips and conducted pre- and post-PCR manipulations in separate rooms to safeguard against cross-contamination.

2.2 | Preparation of eDNA samples and standard fish DNAs

Water samples for the method testing were collected from three natural environments: two marine and one freshwater environments in Japan. Seawater samples were collected at Nagahama, Maizuru in Kyoto prefecture (35°29′24″N, 135°22′6″E) on May 13, 2021, and Otomi, Takahama in Fukui prefecture (35°32′24″N, 135°30′3″E) on May 14, 2021. Seawater was collected by throwing a bucket tied with a rope from a pier 10 times. Three Sterivex filter cartridges (SVHV010RS, Merck Millipore, Darmstadt, Germany) were used at each site, and 1000 ml of seawater was filtered using a disposal syringe for each cartridge, followed by the addition of RNAlater (ThermoFisher Scientific, Waltham, Massachusetts, USA) to prevent DNA degradation. The cartridges were transferred to the laboratory within 30 min and stored at −20°C.

Freshwater samples were collected from the Seta River in Otsu, Shiga prefecture, Japan (34°57′39″N, 135°54′32″E) on March 30, 2020. We collected 1600 ml of river water using plastic bottles, and filtered the water sample using eight Sterivex filter cartridges (each Sterivex filtered 200 ml of water sample). The cartridges were brought back to the laboratory within 1 h, and stored at −20°C. The samples were kept at 4°C during transport.

In our experiments, we did not analyze field negative controls (e.g., samples that filtered sterilized distilled H_2O instead of field water) because the primary objective of this study was to investigate whether and how fish community compositions change when identical eDNA samples are analyzed using different experimental protocols, and not to elucidate fish community compositions under field conditions. In eDNA studies in which the primary objective is to reveal ecological community compositions, the inclusion of multiple field negative controls is necessary.

Detailed protocols for DNA extraction are described in the Appendix S1. Briefly, DNA was extracted using a DNeasy Blood & Tissue kit (Qiagen, Hilden, Germany) following a protocol described previously (Miya et al., 2016). After the cell lysis and purification, DNA was eluted using 100 µl of the elution buffer. Extracted DNAs from the same study site were combined and treated as one composite sample for the method testing. Eluted DNA samples were stored at −20°C until further processing.

In addition to the three samples obtained from natural ecosystems, we prepared 10 fish-like standard DNAs (i.e., artificial DNA fragments with MiFish primer regions and conservative regions that are the same as existing fish species; see *Preparations of standard fish DNAs* in the Appendix S1). These 10 fish-like standard DNAs were synthesized using gBlocks Gene Fragments service by Integrated DNA Technologies, Inc. (Coralville, IA, USA). Equal amounts of the 10 standard DNAs were mixed and treated identically with the eDNA samples from the three study sites. More information about the standard fish DNAs is available in the Appendix S1 and Table S1.

2.3 | Quantification and normalization of fish eDNA concentrations

Total DNA concentrations varied among samples (Nagahama, 44.8 ng/µl; Otomi, 9.1 ng/µl; Seta River, 4.8 ng/µl; quantified by NanoDrop One, ThermoFisher Scientific, Waltham, MA, USA). Before we started the experiments, the concentrations of fish eDNA and the standard DNAs were normalized so that similar numbers of sequence reads could be generated for each site and each replicate. Fish eDNA concentrations were measured using quantitative PCR (qPCR) as described in the Appendix S1. The minimum concentration of fish eDNA was found in the river eDNA sample (43.2 copies/µl), and thus, the other eDNA samples were diluted so that their fish eDNA concentrations are 43.2 copies/µl.

2.4 | Experiment I: Effects of tagging/indexing methods on the community composition revealed by eDNA metabarcoding

2.4.1 | Tagging/indexing methods and two DNA polymerases

In Experiment I, the primary objective was to test the effects of two tagging/indexing methods, 1st PCR tagging and 2nd PCR

indexing, on the detected fish eDNA compositions. First PCR tagging is a method to append a sample-specific sequence in the 1st PCR (Figure 1a). This method is referred to as "1st PCR tagging," not "1st PCR indexing" (Bohmann et al., 2022), and similar approaches were previously applied in eukaryotic eDNA metabarcoding studies (Bohmann et al., 2022; Leray & Knowlton, 2017; Minardi et al., 2022; O'Donnell et al., 2016; Zizka et al., 2019). The 1st PCR primer set for 1st PCR tagging (hereafter referred to as "the 1st tagging-1st PCR primer") is composed of an Illumina sequencing primer, 8-base sample-specific sequences, and a MiFish-U primer (Table 1). In the 1st PCR-tagging protocol, PCR products are pooled and purified after the 1st PCR, and Illumina P5/P7 adapters are appended for the pooled sample at the 2nd PCR using "the 1st tagging-2nd PCR primer" (Table 1). On the contrary, the 2nd PCR-indexing protocol does not append a sample-specific sequence at the 1st PCR and uses "the 2nd indexing-1st PCR primer" to amplify a target region (Figure 1a; e.g., Miya et al., 2015). "The 2nd indexing-1st PCR primer" is identical to MiFish-U primers, except that it does not include six ambiguous sequences (Miya et al., 2015). After the 1st PCR, each PCR product is separately purified, and is then used as a template for the 2nd PCR. At the 2nd PCR, sample-specific index sequences and the Illumina P5/P7 adapters are appended to the amplicon.

As a secondary objective, we evaluated effects of two DNA polymerases on the outcomes of eDNA metabarcoding: KAPA HiFi HotStart ReadyMix (KAPA Biosystems, Wilmington, WA, USA; hereafter, "KAPA") and Platinum SuperFi II PCR Master Mix (ThermoFisher Scientific, Waltham, MA, USA; hereafter, "Platinum"). The former is a commonly used enzyme in eDNA metabarcoding, while the latter is a newer enzyme with higher fidelity according to the manufacturer's information.

2.4.2 | Experimental design

In Experiment I, we tested three library preparation protocols: 2nd PCR indexing with KAPA, 2nd PCR indexing with Platinum, and 1st PCR tagging with Platinum. The first two were used to compare the effects of enzymes, and the last two were to compare the effects of tagging/indexing methods. In each library preparation protocol, eDNA samples originated from three study sites (i.e., Nagahama, Otomi, and Seta-river) and one standard DNA mix were analyzed. In each study site, five technical replicates and one negative control (H_2O) were included. Thus, we had 72 eDNA libraries in total, that is, (5 technical replicates + 1 negative control) × (3 study sites and 1 standard DNA) × (3 protocols) = 72.

2.4.3 | Library preparations and iSeq sequencing

The thermal cycle profiles of the 1st and 2nd PCR are described in Table S2. We used dual-unique combinations of the forward and reverse tags/indices for different templates (samples) for Illumina sequencing, which reduced the risk of tag jumping (Esling et al., 2015). All combinations of tags/indices are available at https://github.com/ong8181/eDNA-early-pooling/tree/main/sampledata.

For all treatments, we performed the 1st and 2nd PCR following the manufacturers' protocols as described in the Appendix S1. Briefly, for "the 2nd PCR indexing with KAPA" treatment (i.e., a common 2-step PCR protocol), KAPA HiFi HotStart ReadyMix was used for the 1st and 2nd PCR. Sample-specific index sequences were appended in the 2nd PCR. For "the 2nd PCR indexing with Platinum" treatment, Platinum SuperFi II PCR Master Mix was used instead of KAPA, and sample-specific index sequences were appended in the 2nd PCR. In the "1st PCR tagging with Platinum" treatment, sample-specific sequences were appended in the 1st PCR. After the 2nd PCR, the pooled 2nd PCR products from each treatment were purified, target-sized DNA was excised, the double-stranded DNA concentrations of the libraries were quantified, and the libraries from the three treatments were combined as one sample. The double-stranded DNA concentration of the combined library was then adjusted to 50 pM using 10 mM Tris–HCl (pH 8.5), and DNA was then sequenced by the iSeq 100 system (Illumina, San Diego, CA, USA) using iSeq 100 Reagent v2 (2 × 150 bp PE). For the sequencing, 30% PhiX was spiked-in to improve the sequencing quality.

2.5 | Experiment II: Effects of post-1st PCR purification on tag-jumping events

The early-pooling protocol may cause tag-jumping events because the enzyme and tagged primers were not inactivated before sample pooling. Therefore, in Experiment II, we measured the frequency of tag-jumping events in the 1st PCR-tagging treatment and investigated the effects of exonuclease purification, temperature, and time after sample pooling on the frequency of tag-jumping events (Figure 1b).

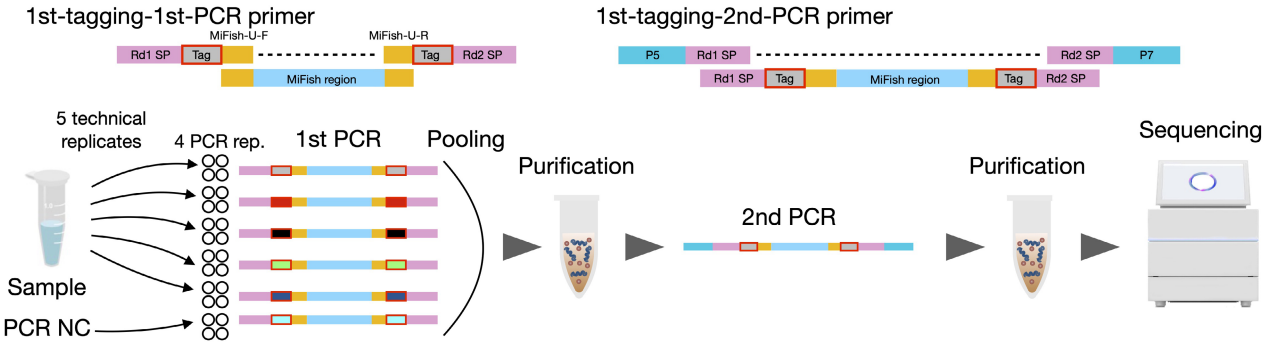
2.5.1 | Experimental design

In Experiment II, the frequency of tag-jumping events was examined by preparing one positive sample and two negative samples for

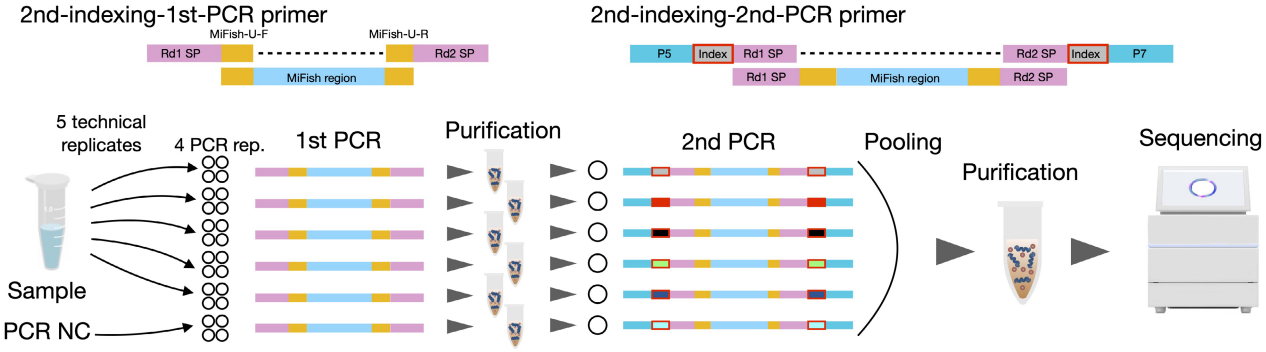
FIGURE 1 Experimental designs of this study. (a) Experiment I. the effects of the tagging/indexing method were examined. The 1st PCR-tagging protocol appends sample-specific sequences in the 1st PCR using "1st tagging-1st PCR primers," while the 2nd PCR-indexing protocol appends index sequences in the 2nd PCR using "2nd indexing-2nd PCR primers." "purification" indicates magnetic bead purification in experiment I. (b) Experiment II. The effects of exonuclease purification on tag-jumping events were examined. Exonuclease purification was performed for each sample. Then, samples were combined and incubated to test the effects of incubation time and temperature. (c) Experiment III. The effects of the 1st PCR replicates and template DNA volume were tested. Other processes were identical with that of experiment II with exonuclease purification

(a) Experiment I: Tagging/indexing method test

1st PCR tagging protocol

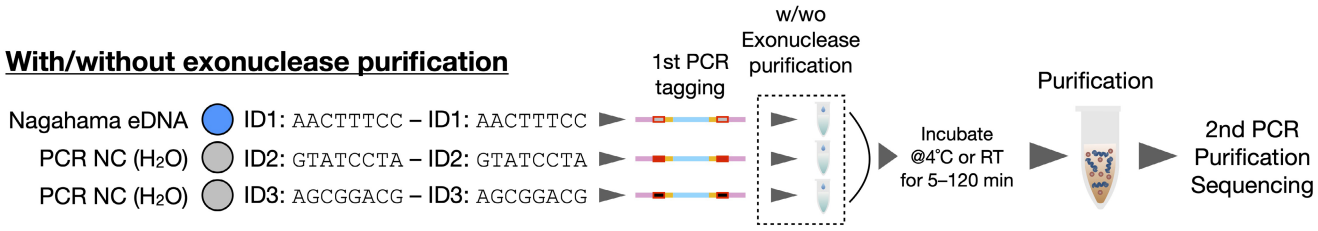


2nd PCR indexing protocol



(b) Experiment II: Tag-jumping test

With/without exonuclease purification



(c) Experiment III: The number of replicates and template volume test

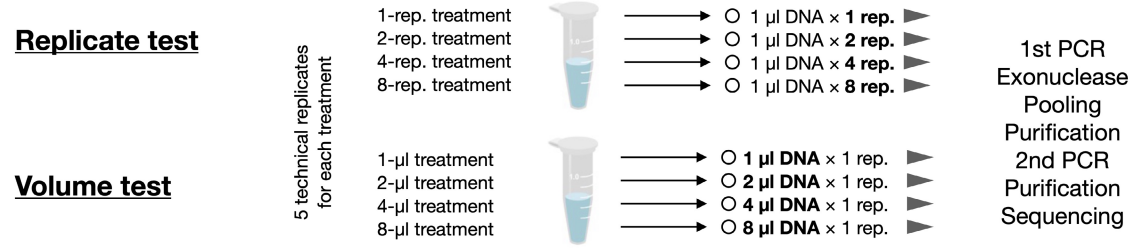


TABLE 1 Primer sequences used in the present study

Primer information	Primer sequence[a,b,c,d]	Length
The 1st PCR-tagging method		
"1st tagging-1st PCR primers"		
Forward primer for the 1st PCR	*ACACTCTTTCCCTACACGACGCTCTTCCGATCT* **XXXXXXXX** GTCGGTAAAACTCGTGCCAGC	62
Reverse primer for the 1st PCR	*GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT* **XXXXXXXX** CATAGTGGGGTATCTAATCCCAGTTTG	69
"1st tagging-2nd PCR primers"		
Forward primer for the 2nd PCR	AATGATACGGCGACCACCGAGATCTACAC *ACACTCTTTCCCTACACGACGCTCTTCCGATCT*	62
Reverse primer for the 2nd PCR	CAAGCAGAAGACGGCATACGAGAT *GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT*	58
The 2nd PCR-indexing method		
"2nd indexing-1st PCR primers"		
Forward primer for the 1st PCR	*ACACTCTTTCCCTACACGACGCTCTTCCGATCT* GTCGGTAAAACTCGTGCCAGC	54
Reverse primer for the 1st PCR	*GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT* CATAGTGGGGTATCTAATCCCAGTTTG	61
"2nd indexing-2nd PCR primers"		
Forward primer for the 2nd PCR[e]	AATGATACGGCGACCACCGAGATCTACAC **XXXXXXXX** *ACACTCTTTCCCTACACGA*	56
Reverse primer for the 2nd PCR[e]	CAAGCAGAAGACGGCATACGAGAT **XXXXXXXX** *GTGACTGGAGTTCAGACGTGT*	53

[a]Normal characters indicate target-specific universal primers (MiFish-U-F/R in the present study).
[b]Italic characters indicate the Illumina sequencing primers.
[c]**X** indicates tag/index sequences to identify each sample.
[d]Underlined characters indicate P5/P7 adapter sequences for Illumina sequencing.
[e]Several bases of the 3′ end of the 2nd PCR primers for the 2nd PCR indexing were deleted to reduce the cost of purchasing the primers.

each treatment. Two negative controls were prepared to increase the number of combinations of unused tags. eDNA extracted from Nagahama seawater was used as the positive sample, while H_2O was used as negative samples. These three samples were distinguished by dual-unique tags, for example, the combinations of ID1–ID1, ID2–ID2, and ID3–ID3 (Figure 1b). In the 1st PCR-tagging protocol in Experiment I, 1st PCR products were immediately combined after the 1st PCR and the polymerase and primers were not inactivated when the samples were pooled. Therefore, we may find "unused" tag combinations, such as ID1–ID2, ID2–ID1, ID1–ID3, and ID3–ID1, which is a signature of tag-jumping events. Two-time tag jumping may generate the combinations of ID2–ID2 and ID3–ID3; however, we disregarded this because it should occur less frequently than one-time tag jumping.

We evaluated the effects of the three variables on the frequency of tag-jumping events: exonuclease purification for each 1st PCR product (hereafter referred to as the "without exonuclease" or "with exonuclease" treatment) and the incubation temperature and time after sample pooling. We examined the effects of temperature and time because we expected these two variables to affect the performance of DNA polymerase and, hence, the amplification of the pooled 1st PCR products with primers with different tags. The purification of each 1st PCR product was performed by incubating PCR products using ExoSAP-IT Express (ThermoFisher Scientific, Waltham, MA, USA) only in the "with exonuclease" treatment, which is less labor-intensive than purification using magnetic beads. The three 1st PCR products for each treatment were then pooled and incubated at two temperatures (on ice [ca. 4°C] or at room temperature [ca. 22°C]) for three durations (5, 30, and 120 min). Thus, we had 36 eDNA libraries in total, that is, (1 positive sample + 2 negative controls) × (with/without exonuclease) × (2 temperatures) × (3 durations) = 36.

2.5.2 | Library preparations and iSeq sequencing

Detailed protocols for Experiment II are described in the Appendix S1. The 1st PCR was performed on 36 samples using the 1st PCR-tagging protocol with Platinum as in Experiment I with some modifications. Briefly, for the "with exonuclease" treatment only, 4 μl of ExoSAP-IT Express was added to each 10-μl PCR product. The mixture was incubated for 4 min at 37°C, followed by the incubation for 1 min at 80°C. In addition, we performed the 2nd PCR using "2nd indexing-2nd PCR primers" to append additional indices to the library (i.e., tag-index combined approach). This approach was adopted to ensure that the three tags used to distinguish samples in

each treatment were identical among the treatments. The double-stranded DNA concentration of the combined library was then adjusted and DNA was sequenced by the iSeq 100 system using iSeq 100 Reagent v2 (2×150 bp PE).

2.6 | Experiment III: Effects of the number of PCR replicates and volume of eDNA template on the number of OTUs detected

As the results of Experiments I and II suggested that fish community compositions revealed by 1st PCR tagging with exonuclease purification were similar to those revealed by the 2nd PCR-indexing protocol, in Experiment III, we evaluated the effects of replications and the template volume in the 1st PCR to further reduce potential experimental and labor costs associated with library preparation.

2.6.1 | Experimental design

In Experiment III, we prepared eight treatments, that is 1, 2, 4, and 8 replicate treatments ("replicate" treatment) and 1, 2, 4, and 8-μl treatment ("volume" treatment), for two sample types (Nagahama seawater and standard DNA samples; Figure 1c). In each treatment, five technical replicates and one negative control were included. We chose the Nagahama sample because its eDNA diversity was highest among the four sample types (ca. 20–30 fish species were detected in Experiment I and II). A standard DNA sample consisting of 10 fish-like DNAs was included to evaluate species detection probabilities. We had 96 eDNA libraries in total, that is, (5 technical replicates + 1 negative control) × (4 replicate treatments and 4 volume treatments) × (2 sample types) = 96.

2.6.2 | Library preparations and iSeq sequencing

Detailed protocols are described in the Appendix S1. DNA libraries were prepared using the 1st PCR-tagging protocol with Platinum, as described in Experiment I with the following modifications. Briefly, for all treatments, exonuclease purification for each 1st PCR product was included. In the "replicate" treatment, the number of 1st PCR replicates was changed for each treatment. The number of 1st PCR replicates was 1, 2, 4, or 8, and replicates were combined after the 1st PCR. In the "volume" treatment, the total reaction volume was increased from 10 to 20 μl to include up to 8 μl of the eDNA template. The number of replicates was 1 for all "volume" treatments. As in Experiment II, "2nd indexing-2nd PCR primers" were used to append additional indices to the library (i.e., tag-index combined approach). This approach was adopted to distinguish libraries with identical 1st PCR tags. The double-stranded DNA concentration of the combined library was then adjusted and DNA was sequenced by the iSeq 100 system using iSeq 100 Reagent v2 (2×150 bp PE).

2.7 | Sequence data processing

Detailed procedures for our sequence data processing are described in the Appendix S1. Our sequence data included sample-specific tag sequences inside the sequencing primers and we prepared a custom shell script using seqkit 2.1.0 (Shen et al., 2016) to demultiplex our samples (see https://github.com/ong8181/eDNA-early-pooling/tree/main/01_Demultiplex and https://doi.org/10.5281/zenodo.6045851). Raw sequences were demultiplexed and MiFish primer regions were trimmed by cutadapt 2.10 (Martin, 2011). The quality of our sequence data was high, and after primer trimming, 8,948,113 sequence reads (% > Q30 = 93.9; average 43,863 reads per sample) remained for the three experiments.

The demultiplexed, primer-trimmed sequences were processed using DADA2 (Callahan et al., 2016), an amplicon sequence variant (ASV) approach, for each iSeq run. First, at the quality filtering process, low quality and unexpectedly short reads were removed. Error rates were learned, and sequences were dereplicated, error-corrected, and merged to produce an ASV-sample matrix. Then, chimeric sequences were removed. ASVs detected in the three experiments were merged and clustered into OTU at 97% similarity using DECIPHER package of R (Wright, 2016), which converted the ASV-sample matrix into the OTU-sample matrix. We converted ASVs to OTUs because ASVs often detected intraspecific variations and the focus of this experiment was fish community compositions at the species level (not at the intra-species level). Taxonomic identification was performed for OTUs based on the query-centric auto-*k*-nearest-neighbor (QCauto) method (Tanabe & Toju, 2013) implemented in Claident v0.9.2021.10.22 (https://www.claident.org/). The QCauto method is a conservative method, and all of the nearest neighbors of a query sequence need to have the same taxa information for taxa assignment. For example, if the QCauto method assigns a genus name to an OTU, all of the nearest neighbors surrounding the OTU have the same genus name. This method increases the number of OTUs without species names, but decreases the potential for misassignments in species information more than other common methods (Sato et al., 2018; Wang et al., 2007). As the QCauto method requires at least two sequences from a single taxon, standard DNAs were separately identified using BLAST (Camacho et al., 2009).

2.8 | Statistical analyses and data visualization

All statistical analyses and visualizations were performed using a free statistical environment, R 4.1.2 (R Core Team, 2021). The sample metadata, taxa information assigned to each OTU, and OTU-sample matrix were imported as an object using phyloseq package of R (McMurdie & Holmes, 2013). The merged phyloseq object was then divided into the three experiments and analyzed separately. In this study, we describe the results of statistical tests using the term "statistical clarity" rather than "statistical significance" to avoid misinterpretations, according to the recommendations by Dushoff et al. (2019).

2.8.1 | Statistical analyses of experiment I

In Experiment I, the data set was first divided into each sample type (i.e., Nagahama, Otomi, Seta-river, or standard DNA; Figure S1) after a brief inspection of the sequence reads distribution. The sequence reads generated in the negative controls were mostly minor compared with those generated in the samples (Figure S1). However, in 4 of the 12 negative controls, we detected non-negligible sequence reads from one or two OTUs (see Results). Therefore, we subtracted the sequence reads per OTU in each negative control in each treatment from true sample reads. Then, the sequence reads were rarefied by the minimum number of sequence reads detected in each sample type to remove the effects of sequence depth on results. Rarefying sequence reads by the minimum number of sequence reads is usually not recommended for ecological studies (McMurdie & Holmes, 2014). However, in our case, this approach was suitable because replicates in each sample type were technical replicates, and because we wanted to focus on the effects of the library preparation protocols on the outcomes by removing the effects of sequence depth. After rarefactions, we removed rare OTUs with <0.01% relative abundance from each sample to further mitigate the effects of contaminated sequence reads.

After the rarefaction within each sample type, the fish community compositions were first visualized using bar plots. The effects of the library preparation protocols on the numbers of detected OTUs were tested using the generalized linear model (GLM with Poisson error distribution). The dependence of fish community compositions on the library preparation protocols was assessed by nonmetric dimensional scaling (NMDS) based on the Bray–Curtis dissimilarity, and the differences in the detected community compositions among the library preparation protocols were tested by the analysis of similarities (ANOSIM) using vegan package of R (Oksanen et al., 2008). Also, the effects of the library preparation protocols on the Bray–Curtis dissimilarities were tested by bootstrap analysis (see https://github.com/ong8181/eDNA-early-pooling/tree/main/08_Exp1_1st2nd for detail).

2.8.2 | Statistical analyses of experiment II

In Experiment II, there were one positive sample and two negative samples in each treatment. As each treatment has three dual-unique tags (ID1, ID2, and ID3), there were nine possible combinations of tags (3×3 combinations), among which four "unused" combinations indicated possible tag-jumping events (i.e., ID1-ID2, ID2-ID1, ID1-ID3, and ID3-ID1). Sequence reads assigned to OTUs with these four tag combinations were standardized by the sequence reads of the positive sample. For example, in the case that OTU001 has 10,000 reads in the positive sample (ID1-ID1) and 10 reads are detected as tag jumps (e.g., the ID1-ID2 combination), the proportion of tag-jumped reads is calculated as 0.1%. The pattern was visualized and the statistical clarity of three variables, that is, purification process, temperature, and time, was tested by GLM.

2.8.3 | Statistical analyses of experiment III

In Experiment III, the data set was first divided into each sample type (i.e., Nagahama, or standard DNA). Similar to Experiment I, we subtracted sequence reads per OTU in each negative control in each treatment from true sample reads. Sequence reads were then rarefied to the minimum number of sequence reads in each sample type to remove the effects of sequence depth. After rarefactions, we removed rare OTUs with <0.01% relative abundance from each sample to further mitigate the effects of contaminated sequence reads. The effects of the reaction scale (i.e., the number of 1st PCR replicates or the volume of template DNA) were then statistically tested using GLM. In addition, how the number of 1st PCR replicates and template DNA volume influence the detection of rare OTUs was evaluated (see Results). The effects of the number of replicates and total template volume on the Bray–Curtis dissimilarities were tested by bootstrap analysis as described for Experiment I.

2.8.4 | Tag- and protocol-specific biases in the detected community composition

We performed additional analyses to investigate tag- and protocol-specific biases in the detected community composition using data from Experiments I, II, and III. In the tag-specific bias test, samples sequenced using the identical tag were grouped, and variations in the relative abundance of dominant OTUs among the sequences were tested. We performed this test because previous studies reported that a 1st PCR-tagging protocol introduced tag-specific biases in eDNA-based community composition detections (Berry et al., 2011; O'Donnell et al., 2016). In the protocol-specific bias test, we visualized how the protocols changed the fish eDNA compositions of the Nagahama samples because these samples were analyzed using various library preparation protocols (e.g., the 1st PCR-tagging method, the 2nd PCR-indexing method, and with/without exonuclease purification).

2.9 | Code and data availability

Sequence data are deposited in DDBJ Sequence Read Archives (DRA) (DRA accession number = DRA013399). Analysis scripts and the information about R packages are archived at Zenodo (https://doi.org/10.5281/zenodo.6679138).

3 | RESULTS

In total, 8,948,113 sequence reads were generated from 204 samples by three iSeq runs, of which 8,530,488 reads remained after DADA2 sequence processing. Of the sequences that remained, 8,518,552 reads were assigned to 113 fish OTUs, including the

standard DNAs. A small number of sequence reads (11,936 reads; 0.13%) was assigned to 21 non-fish OTUs such as mouse (*Mus musculus*), pig (*Sus scrofa*), and common pochard (*Aythya ferina*). In the subsequent analysis, we focused on the 113 fish OTUs.

3.1 | Experiment I: Effects of tagging/indexing methods on the community composition revealed by eDNA metabarcoding

In total, 1,910,880 reads were generated from 72 samples in Experiment I. The sequence reads generated in the negative controls were mostly minor compared with those generated in the samples (Figure S1; 0–1.2% of true sample reads for 8 of the 12 negative controls). However, in 4 of the 12 negative controls, we detected non-negligible sequence reads from one or two OTUs (e.g., 1705–3647 *Acanthopagrus* sp. reads were detected in the negative control samples in "2nd PCR indexing with Platinum"). Therefore, we subtracted the sequence reads per OTU found in each negative control in each treatment from true sample reads. To archive fair comparisons between the protocols, we rarefied sequence reads to the minimum number of sequence reads within each sample type (21,452 reads for Nagahama; 8351 reads for Otomi; 10,151 reads for Seta river; 9574 reads for standard DNA mix). After rarefactions, to further mitigate the effects of contaminated sequence reads, we removed rare OTUs with <0.01% relative abundance from each sample.

Overall, we detected very similar OTU richness and community compositions regardless of the library preparation protocols (Figure 2 and Figure S2). The effect of the protocols on fish OTU richness was statistically unclear (Figure S2a; $p > 0.05$). However, in standard DNA mix samples, one or two standard DNAs were not detected in four of five replicates in the 1st PCR-tagging protocol, while all 10 standard DNAs were detected in most replicates in the 2nd PCR-indexing protocol. *Acanthopagrus* sp (OTU002), *Hexagrammos* sp. (OTU092), and Cyprinidae (OTU004) were the most-dominant fish OTUs in Nagahama, Otomi, and Seta river, respectively, and similar community compositions were detected by the three protocols (Figure S2b). Similarly, the NMDS plot shows that the samples are largely overlapped among the protocols (Figure 2a–d). Difference in the fish community composition between "the 2nd PCR indexing with Platinum" and "the 2nd PCR indexing with KAPA" was statistically unclear (ANOSIM; $p > 0.05$). Similarly, difference between "1st PCR tagging with Platinum" and "2nd PCR indexing with Platinum" was statistically unclear (ANOSIM; $p > 0.05$). We further investigated the reproducibility in each protocol (i.e., variations in the technical replicates within a protocol) in the fish community compositions by calculating Bray–Curtis dissimilarity. We found that 1st PCR tagging increased dissimilarities among technical replicates in standard DNA samples (Figure 2e; $p < 0.001$). On the contrary, for the other samples, differences in the dissimilarities among technical replicates were statistically unclear (Figure 2e; $p > 0.05$). In addition, the relative abundance of the dominant fish OTUs were not different among the protocols (Figure S3; $p > 0.05$) except that of the second most-dominant OTU (*Parablennis*) in the Nagahama samples.

3.2 | Experiment II: Effects of post-PCR clean-up on tag jumping

In total, 3,060,823 reads were generated from 36 samples in Experiment II. We detected tag-jumping events and the proportion of tag-jumped sequence reads per sample ranged between 0% and 0.486% (mean ± SD = 0.05 ± 0.08%). Although the proportion of sequence reads for each OTU generated by tag jumping was generally low, "unused" combinations of tag sequences were found in many OTUs in the without-exonuclease treatments (red and dark red points in Figure 3; mean and maximum tag-jumping rates per OTU were 0.0509%–0.0761% and 0.249%–0.475%, respectively). On the contrary, the frequency of tag-jumping events markedly decreased when 1st PCR products were purified using exonuclease before sample pooling (blue and light blue points in Figure 3; mean and maximum tag-jumping rates per OTU were 0%–0.0395% and 0%–0.855%, respectively). No tag-jumping events were observed with the exonuclease treatment within 5 min after sample pooling at 4°C and room temperature. Some tag-jumping events were observed with the exonuclease treatment 30 and 120 min after sample pooling at 4°C and room temperature (0.0007%–0.0395% of reads in each OTU had tag jumps on average); however, the number of OTUs with tags that were jumped was markedly lower than that with non-purified treatments (the number of tag-jumped OTUs = 4–7 OTUs with the exonuclease treatment vs. 39–59 OTUs without the exonuclease treatment). The effect of exonuclease was statistically clear (binominal GLM; $p < 0.0001$), while those of time and temperature were not ($p > 0.05$).

3.3 | Experiment III: Effects of the number of 1st PCR replicates and volume of eDNA template on the number of OTUs detected

In total, 3,546,849 reads were generated from 96 samples in Experiment III. The sequence reads generated in the negative controls were negligible (0 reads were detected for 14 of the 16 negative controls, while only 12 or 63 reads were detected from two out of the 16 negative samples). To be consistent with Experiment I, we subtracted the sequence reads per OTU found in each negative control in each treatment from true sample reads. We then rarefied sequence reads to the minimum number of sequence reads within the sample type (15,394 reads for Nagahama; 3025 reads for standard DNA mix). After rarefactions, we removed rare OTUs with <0.01% relative abundance from each sample to further mitigate the effects of contaminated sequence reads. One of the five replicates in the 1-rep. Treatment was excluded from the analysis because only 304 reads were detected from the replicate (the 1-, 2-, 4-, and, or

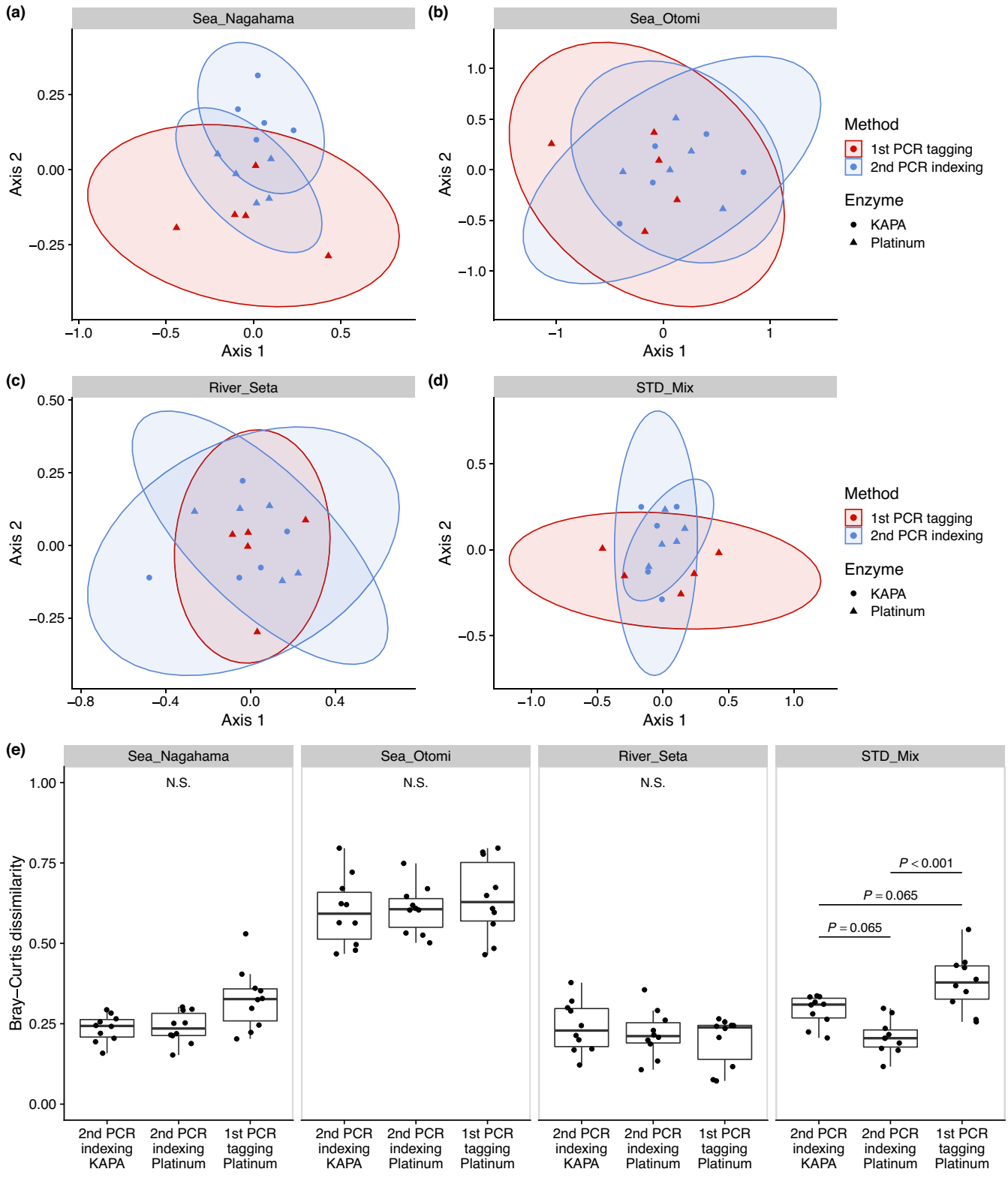


FIGURE 2 Nonmetric dimensional scaling (NMDS) and Bray–Curtis dissimilarities of fish eDNA composition detected in experiment I. fish eDNA compositions of (a; Sea_Nagahama) seawater samples collected in Nagahama, Kyoto, Japan, (b; Sea_Otomi) seawater samples collected in Otomi, Fukui, Japan, (c; River_Seta) freshwater samples collected in Seta River, Otsu, Japan, and (d; STD_mix) a mixture of 10 standard fish DNAs. Filled circles and triangles indicate that KAPA HiFi HotStart ReadyMix and platinum SuperFi II PCR master Mix was used for PCR, respectively. Red and blue circles indicate that sample-specific sequences were appended by the 1st PCR-tagging method and the 2nd PCR-indexing method, respectively. (e) each panel represents eDNA samples collected in one study site. Points in each panel represent Bray–Curtis dissimilarities of fish eDNA compositions between two eDNA samples. The thick bar indicates the median value of the Bray–Curtis dissimilarities in each treatment. Statistical clarity was tested by bootstrap test (see Materials and Methods)

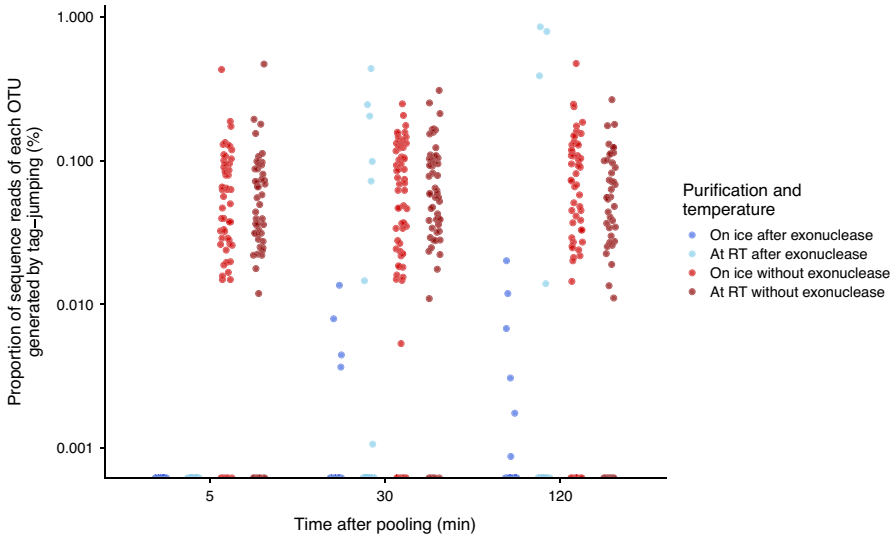


FIGURE 3 Estimation of the tag-jumping probability of each library preparation method. The *y*-axis indicates the proportion of sequence reads for each OTU generated by tag jumping. We used three tag combinations in experiment II, that is, ID1_F (AACTTTCC)-ID1_R (AACTTTCC) (positive sample), ID2_F (GTATCCTA)-ID2_R (GTATCCTA) (H_2O), and ID3_F (AGCGGACG)-ID3_R (AGCGGACG) (H_2O). Tag jumping is defined as the occurrence of any of the four "unused" tag combinations in this study: ID1_F-ID2_R, ID1_F-ID3_R, ID2_F-ID1_R, and ID3_F-ID1_R. We ignored the other possible combinations as described in the main text. Each point represents the tag-jumped sequence reads for each OTU divided by the sequence reads of the corresponding OTU detected in the positive sample. Blue and light blue points indicate that exonuclease-purified 1st PCR products were pooled and incubated on ice and at room temperature (22°C), respectively. Red and dark red points indicate not-purified 1st PCR products were pooled and incubated on ice and at room temperature (22°C), respectively

8-replicate treatments are hereafter referred to the 1-rep., 2-rep., 4-rep., and 8-rep. Treatments, respectively). Overall, the community compositions detected in the treatments were similar (Figure S4), and the numbers of OTUs in Nagahama seawater and standard DNA samples increased with elevations in the number of 1st PCR replicates and volumes of the eDNA template (Figure 4a,b).

Regarding Nagahama seawater samples, in the replicate test, 11.0±1.2 (mean±standard deviation), 16.0±2.6, 21.8±1.8, and 28.0±3.2 fish OTUs were detected in the 1-rep., 2-rep., 4-rep., and 8-rep. Treatments, respectively (Figure 4a). In the volume test, 12.4±0.9, 13.6±2.9, 18.8±1.9, and 25.2±1.3 fish OTUs were detected for Nagahama samples in the 1-, 2-, 4-, and 8-μl treatments, respectively (Figure 4a). In Nagahama samples, the effect of the reaction scale (an increasing number of replicates or an increasing template volume) on the number of fish OTUs was statistically clear (Poisson GLM; $p<0.0001$), while that of the test category ("replicate" test or "volume" test) was not.

Regarding standard DNA samples, in the replicate test, 6.4±1.7 and 9.0±1.0 fish OTUs were detected in 1-rep. and 2-rep. treatments, respectively (Figure 4b). In the volume test, 8.4±1.1 and 9.2±0.4 fish OTUs were detected in 1- and 2-μl treatments, respectively (Figure 4b). Concerning the other treatments, all 10 standard fish DNAs were detected in all replicates. In standard DNA samples, the effect of the reaction scale was statistically unclear (Poisson GLM; $p = 0.103$), which may have been due to the small number of fish standard DNAs included in the samples.

As for the reproducibility of each treatment (i.e., within-treatment variations in the detected community compositions), increasing the number of the 1st PCR replicates or template volumes statistically clearly reduced the Bray–Curtis dissimilarities (Figure 4c,d; $p<0.05$). For both Nagahama and standard DNA samples, 8-rep. and 8-μl treatments showed more consistent community compositions among the technical replicates than 1-rep. and 1-μl treatments did (Figure 4c,d).

We further investigated how the number of 1st PCR replicates and template DNA volumes influenced OTU detection using Nagahama samples (for standard DNA samples, all 10 standard DNAs were detected in the 1-μl or 1-rep. Treatments when OTUs detected in the 5 replicates were merged). We did this by identifying OTUs that were newly detected in a more-replicated treatment or a larger template-volume treatment (Figure 4e,f). For example, "Detected OTUs in 2-rep." indicates that the OTUs were not detected in the 1-rep. treatment but were newly detected in the 2-rep. treatment (Figure 4e). Newly detected OTUs in more-replicated treatments generally had lower abundance than OTUs detected in 1-rep. treatment (Figure 4e), which indicates that rare OTUs can be detected more easily by increasing the number of 1st PCR replicates. These patterns were almost the same for the template volume test (Figure 4f). Newly detected OTUs in larger template-volume treatments generally had lower abundance than OTUs detected in 1-μl treatment (Figure 4f).

3.4 | Testing tag- and protocol-specific biases

Using the data from Experiments I, II, and III, we tested tag- and protocol-specific biases in eDNA metabarcoding. Tag-specific biases were tested by quantifying the relative abundance of dominant OTUs detected in the Nagahama seawater samples (Figure 5a).

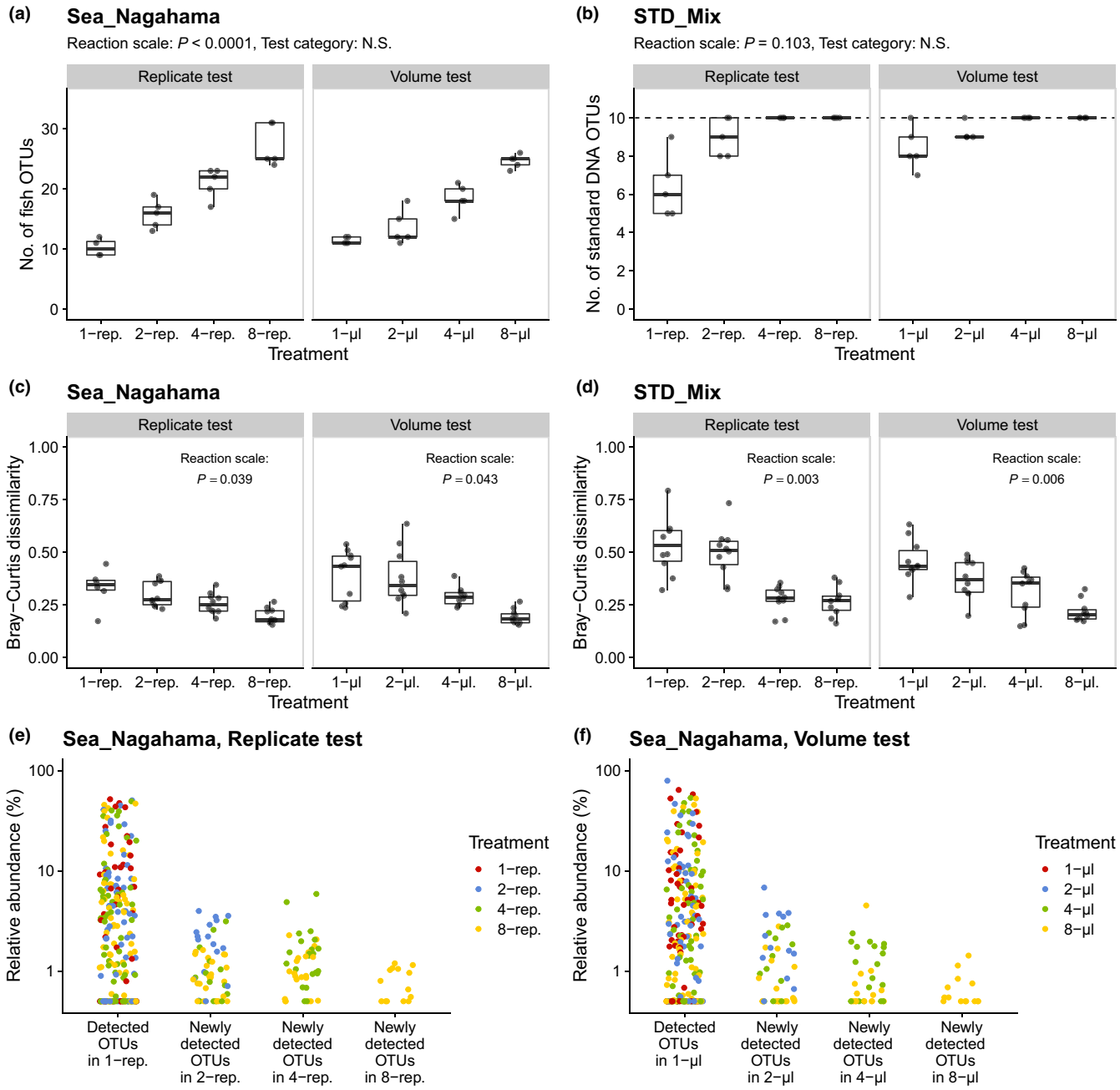


FIGURE 4 Effects of the number of replicates and the volume of template DNA in the 1st PCR reaction on community compositions and OTU richness. (a, b) effects of the number of replicates and the template DNA volume on the number of fish OTUs in the Nagahama samples (a; Sea_Nagahama) and standard DNA samples (b; STD_Mix). In b, dashed horizontal line indicates the number of standard DNAs included in the reactions. Statistical clarity was tested by GLM. (c, d) effects of the number of replicates and the template DNA volume on Bray–Curtis dissimilarities in the Nagahama samples (c) and standard DNA samples (d). Statistical clarity was tested by the bootstrap analysis. (e, f) the relationship between the relative abundance of OTUs and detected treatments. In e, each point indicates that each OTU is newly detected in the replicate treatment described on *x*-axis. Points at "Detected OTUs in 1-rep." indicate all OTUs detected in the 1-rep. Treatment. Colors indicate the treatment in which OTUs are detected. (f) Results for the volume test of the Nagahama samples

OTU002, OTU007, and OTU056 were the three most abundant taxa and were assigned as *Acanthopagrus* sp. (most likely, Japanese black seabream, *A. schlegelii*), *Parablennius yatabei* (Yatabe blenny), and *Takifugu* sp. (most likely, pufferfish, *T. alboplumbeus*), respectively, which are all common fish species in the region (Masuda, 2008; Masuda et al., 2010). The relative abundance of the three OTUs estimated by a common method, the 2nd-indexing protocol with KAPA or Platinum, were 39.5%, 18.4%, and 6%, respectively (red dashed horizontal lines in Figure 5a). Technical replicates were grouped into each tag sequence in Experiments I, II, and III, ignoring detailed experimental steps in each Experiment (e.g., with/without exonuclease, the number of 1st PCR replicates, and the volume of template DNA). We found that the relative abundance of the top three OTUs were similar to those using the common 2nd-indexing protocol

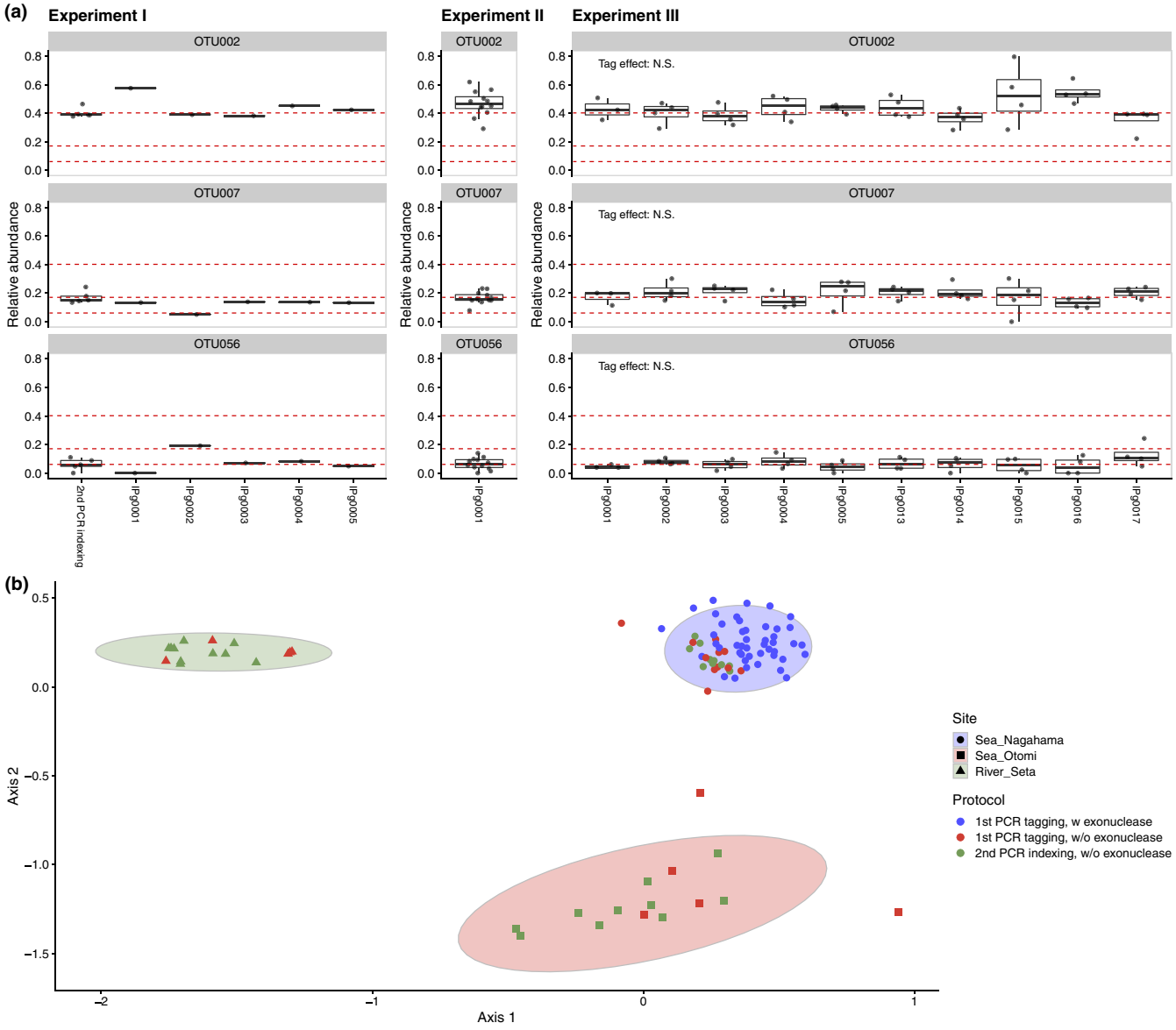


FIGURE 5 Effects of tag sequences and protocols on the relative abundance and community compositions of fish detected in Nagahama samples. (a) the *y*-axis shows the relative abundance of the three most-dominant OTUs. The *x*-axis shows the library preparation method (= "2nd PCR indexing") or the names of the tag sequences used in the 1st PCR. Red dashed horizontal lines indicate the mean relative abundance of each OTU detected by the 2nd PCR-indexing method. In experiment I, only one library was sequenced using each tag. In experiments II and III, multiple libraries were sequenced using the same tag. Note that different experimental treatments (e.g., incubation time and temperature in experiment II) are grouped in this figure. In experiment III, statistical clarity was tested by GLM. (b) the effects of library preparation protocols on the fish community compositions of three study sites visualized by nonmetric dimensional scaling (NMDS). All Nagahama samples are clearly distinguished from natural eDNA samples from the other study sites. Symbols and colors indicate the library preparation protocols and study sites, respectively. Ellipses indicate 95% confidential intervals for each study site. An NMDS plot with more detailed sample information is shown in Figure S5b

regardless of the tag sequences, and the effect of the tag sequence on the relative abundance of the three OTUs was statistically unclear (binomial GLM; $p > 0.05$; tested only for Experiment III).

Protocol-specific community compositions were tested by leveraging the results from Experiment I, II, and III. Fish community compositions of the Nagahama seawater samples were grouped into each protocol and visualized in Figure 5b and Figure S5. In general, the fish community compositions were similar regardless of the protocols used (Figure S5a). These patterns were also evident in NMDS (Figure 5b and Figure S5b). Exonuclease purification after the 1st PCR may slightly affect community compositions; however, fish community compositions were clearly distinguished from those detected in different sample types, such as river and other seawater samples (Figure 5b and Figure S5b).

4 | DISCUSSION

4.1 | Effects of the tagging/indexing method on fish community compositions

In Experiment I, we showed that 1st PCR tagging and early pooling did not result in statistically clear protocol-specific biases in the community compositions or relative abundance of dominant OTUs (Figures 2 and 5, and Figures S1–S3), which is in contrast to previous findings (Berry et al., 2011; O'Donnell et al., 2016). In the study by O'Donnell et al. (2016), biases could be caused by tag-specific mismatches with template DNAs and subsequent variations in primer-template binding efficiencies between taxa. O'Donnell et al. (2016) employed shorter primers (31 bases including three ambiguous bases and six unique bases) for 1st PCR tagging than those used in the present study (62–69 bases including eight unique bases; Table 1), which caused larger variations in T_m among the primers. As differences in T_m affect primer-template binding efficiency (Wu et al., 2009), this difference may be a reason for the discrepancy. Previous studies also reported little differences in outcomes between the 1st PCR-tagging and 2nd PCR-indexing protocols (Leray & Knowlton, 2017; Zizka et al., 2019), which is consistent with the present study. Although difficulties are associated with elucidating the mechanisms responsible for the differences observed between studies, 1st PCR tagging generated qualitatively and quantitatively similar results to those by 2nd PCR indexing, at least in our protocol.

4.2 | Effects of post-1st PCR purification on tag-jumping events

Tag jumping, or index hopping, is an event that prevents accurate DNA-based evaluations of ecological community compositions (Esling et al., 2015), and the development of strategies to reduce the risk of tag-jumping/index-hopping events is an important issue. This is particularly the case in an early-pooling protocol because tagged/indexed primers and enzymes are still "active" at the time of pooling. In Experiment II, tag-jumping events were confirmed by the detection of "unused" combinations of tag sequences (Figure 3). The proportion of tag-jumped sequence reads without the exonuclease treatment ranged between 0% and 0.486%, which is consistent with previous findings (0.18%–0.42%; Snyder & Stepien, 2020). Post-1st PCR exonuclease purification reduced this risk (Figure 3; no tag-jumped sequences were detected in the 5-min incubation after the exonuclease treatment). Although adding the purification step increases the cost and handling time for the library preparation, the exonuclease-based purification is much easier and more rapid compared with magnetic beads-based purification. Also, it does not require an expensive instrument (e.g., a microplate washer such as Hydrospeed, TECAN) even for a large number of samples. Furthermore, pooling steps may be performed at room temperature, which did not increase the risk of tag-jumping events.

4.3 | Effects of the number of 1st PCR replicates and template volume on OTU richness

In Experiment III, we showed that increasing the number of the 1st PCR replicates and template DNA volume had similar effects on the detected OTU richness (Figure 4a,b). Our analysis found that rarer OTUs tended to be detected more easily in the more-replicated and larger template-volume treatments (Figure 4e,f), suggesting that the positive effects of increasing the number of the 1st PCR replicates and template DNA volume are simply due to the increased probability of the occurrence of rare OTUs in the reaction. In addition, increasing the number of the 1st PCR replicates and template DNA reduces variations among technical replicates (Figure 4c,d), which could be due to the mitigation of random sampling effects of rare DNAs. In the previous study, increasing the number of the 1st PCR replicates was recommended to increase the eDNA-based detection probability of fish species (Doi et al., 2019). However, this inevitably increases the pipetting frequency and the amount of consumables used in the experiment (e.g., pipette tips and PCR tubes), which will increase the cost of consumables and reagents and the handling time of the library preparations. On the contrary, increasing template volume does not increase the cost or handling time, and thus would be preferable to increasing the number of replicates, especially when a large number of samples are processed.

4.4 | A recommended protocol

We showed that (i) 1st PCR tagging did not cause clear biases on outcomes, at least in our experimental settings (i.e., with our reagents and primer structures), (ii) post-1st PCR exonuclease purification reduced the risk of tag-jumping events, and (iii) increasing the template DNA volume enhanced the OTU richness detected and reduced variations in the detected community composition among technical replicates, similar to a higher number of 1st PCR replicates. Based on these results, we recommend the following protocol for eDNA metabarcoding library preparation:

1. Prepare tagged 1st PCR primers (e.g., 48 or 96 tagged primers).
2. Append tag sequences in the 1st PCR (the 1st PCR tagging). In the 1st PCR, the number of 1st PCR replicate(s) may be one if the template DNA volume is 4–8 μl or greater.
3. Perform post-1st PCR purification by exonuclease to reduce the risk of tag jumping.
4. Pool purified 1st PCR products.
5. If the concentration of 1st PCR products is low, concentrate DNA using magnetic beads. Also, if primer dimers exist, bead purification can effectively remove them.
6. Perform the 2nd PCR to append the sequencing adapter for one composite sample (2nd indices may be appended to increase the maximum number of multiplexed samples; the "tag-index combined approach").
7. Purify PCR products, adjust concentrations, and sequence DNA.

This "early-pooling" protocol reduces the costs for consumables, reagents (except for the tagged 1st PCR primers), and handling times (see Table S3 for cost and time estimations), and does not require expensive automated liquid handlers (Buchner et al., 2021). In addition, the "tag-index combined approach" enables the easy multiplexing of a large number of samples. For example, the combination of 96 1st PCR tags and 96 2nd PCR indices enables the differentiation of nearly 10,000 samples (96×96 = 9216) in a single sequencing run, which will contribute to large-scale eDNA monitoring.

4.5 | Potential limitations of the early-pooling protocol

Despite several advantages of the early-pooling eDNA metabarcoding protocol, there are several potential limitations/disadvantages. The early-pooling protocol requires tagged 1st PCR primers for each metabarcoding region, which will potentially increase the associated costs for primers for the 1st PCR. Therefore, researchers attempting to test many different metabarcoding regions for a relatively small number of samples may first consider using the common 2nd-PCR-indexing protocol, which uses the same indexed primer set for any metabarcoding region. Second, increasing the template DNA volume will simultaneously increase the concentrations of PCR inhibitors such as humic substances (Schrader et al., 2012). For samples with a high concentration of PCR inhibitor, one may consider increasing the number of the 1st PCR replicates or using a PCR-inhibitor-tolerant DNA polymerase.

5 | CONCLUSIONS

In the present study, we showed that an early-pooling protocol with post-1st PCR exonuclease purification and an increased amount of the DNA template will reduce the risk of tag jumping and costs for consumables, reagents, and handling times in library preparation, and that it produces similar results as those with a common 2nd PCR-indexing protocol. Therefore, once a target metabarcoding region is selected and a set of tagged 1st PCR primers is prepared, the early-pooling protocol provides a cost, labor, and time-efficient approach for processing a large number of samples, which will contribute to a more detailed understanding and near-future forecasting of ecosystem dynamics.

AUTHOR CONTRIBUTIONS

MU and AJN conceived and designed research; MU, SF, HM, and RM collected test samples; SF and MU performed experiments; MU analyzed data; MU, SF, and AJN wrote the draft; All authors discussed the results and completed the manuscript.

ACKNOWLEDGEMENTS

We thank Misaki Shiomi in Maizuru Fisheries Research Station for assisting with water collection and eDNA extraction. This research was supported by JSPS KAKENHI (B) 20H03323 (to MU) and the Hakubi Project in Kyoto University (to MU).

CONFLICT OF INTEREST

SF is employed by Clockmics, which is undertaking eDNA library preparations and sequencing. The commercial affiliation of the author (SF) does not alter our adherence to the journal policies on sharing data and materials. None of the authors will benefit directly financially from the publication of this paper.

DATA AVAILABILITY STATEMENT

All scripts used in the present study are archived in Zenodo (https://doi.org/10.5281/zenodo.6785073). Sequence data are deposited in DDBJ Sequence Read Archives (DRA) (DRA accession number = DRA013399).

ORCID

Masayuki Ushio https://orcid.org/0000-0003-4831-7181
Atsushi J. Nagano https://orcid.org/0000-0001-7891-5049

REFERENCES

Berry, D., Ben Mahfoudh, K., Wagner, M., & Loy, A. (2011). Barcoded primers used in multiplex amplicon pyrosequencing bias amplification. *Applied and Environmental Microbiology*, *77*(21), 7846–7849. https://doi.org/10.1128/AEM.05220-11

Bista, I., Carvalho, G. R., Walsh, K., Seymour, M., Hajibabaei, M., Lallias, D., Christmas, M., & Creer, S. (2017). Annual time-series analysis of aqueous eDNA reveals ecologically relevant dynamics of lake ecosystem biodiversity. *Nature Communications*, *8*, 14087. https://doi.org/10.1038/ncomms14087

Bohmann, K., Elbrecht, V., Carøe, C., Bista, I., Leese, F., Bunce, M., Yu, D. W., Seymour, M., Dumbrell, A. J., & Creer, S. (2022). Strategies for sample labelling and library preparation in DNA metabarcoding studies. *Molecular Ecology Resources*, *22*(4), 1231–1246. https://doi.org/10.1111/1755-0998.13512

Buchner, D., Macher, T.-H., Beermann, A. J., Werner, M.-T., & Leese, F. (2021). Standardized high-throughput biomonitoring using DNA metabarcoding: Strategies for the adoption of automated liquid handlers. *Environmental Science and Ecotechnology*, *8*, 100122. https://doi.org/10.1016/j.ese.2021.100122

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, *13*(7), 581–583.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics*, *10*, 421. https://doi.org/10.1186/1471-2105-10-421

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Lozupone, C. A., Turnbaugh, P. J., Fierer, N., & Knight, R. (2011). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(Supplement 1), 4516–4522. https://doi.org/10.1073/pnas.1000080107

Deiner, K., Fronhofer, E. A., Mächler, E., Walser, J.-C., & Altermatt, F. (2016). Environmental DNA reveals that rivers are conveyer belts of biodiversity information. *Nature Communications*, *7*, 12544.

Doi, H., Fukaya, K., Oka, S., Sato, K., Kondoh, M., & Miya, M. (2019). Evaluation of detection probabilities at the water-filtering and initial PCR steps in environmental DNA metabarcoding using a

multispecies site occupancy model. *Scientific Reports*, *9*(1), 3581. https://doi.org/10.1038/s41598-019-40233-1

Dushoff, J., Kain, M. P., & Bolker, B. M. (2019). I can see clearly now: Reinterpreting statistical significance. *Methods in Ecology and Evolution*, *10*(6), 756–759. https://doi.org/10.1111/2041-210X.13159

Esling, P., Lejzerowicz, F., & Pawlowski, J. (2015). Accurate multiplexing and filtering for high-throughput amplicon-sequencing. *Nucleic Acids Research*, *43*(5), 2513–2524. https://doi.org/10.1093/nar/gkv107

Ficetola, G. F., Miaud, C., Pompanon, F., & Taberlet, P. (2008). Species detection using environmental DNA from water samples. *Biology Letters*, *4*(4), 423–425. https://doi.org/10.1098/rsbl.2008.0118

Fukuzawa, T., Kameda, Y., Nagata, H., Nishizawa, N., & Doi, H. (2022). Filtration extraction method using a microfluidic channel for measuring environmental DNA. *Molecular Ecology Resources*, 1–11. https://doi.org/10.1111/1755-0998.13657

Ishige, T., Miya, M., Ushio, M., Sado, T., Ushioda, M., Maebashi, K., Yonechi, R., Lagan, P., & Matsubayashi, H. (2017). Tropical-forest mammals as detected by environmental DNA at natural saltlicks in Borneo. *Biological Conservation*, *210*, 281–285. https://doi.org/10.1016/j.biocon.2017.04.023

Jerde, C. L., Mahon, A. R., Chadderton, W. L., & Lodge, D. M. (2011). "Sight-unseen" detection of rare aquatic species using environmental DNA. *Conservation Letters*, *4*(2), 150–157. https://doi.org/10.1111/j.1755-263X.2010.00158.x

Kelly, R. P., Port, J. A., Yamahara, K. M., & Crowder, L. B. (2014a). Using environmental DNA to census marine fishes in a large mesocosm. *PLoS ONE*, *9*(1), e86175. https://doi.org/10.1371/journal.pone.0086175

Kelly, R. P., Port, J. A., Yamahara, K. M., Martone, R. G., Lowell, N., Thomsen, P. F., Mach, M. E., Bennett, M., Prahler, E., Caldwell, M. R., & Crowder, L. B. (2014b). Harnessing DNA to improve environmental management. *Science*, *344*(6191), 1455–1456. https://doi.org/10.1126/science.1251156

Leray, M., & Knowlton, N. (2017). Random sampling causes the low reproducibility of rare eukaryotic OTUs in Illumina COI metabarcoding. *PeerJ*, *5*, e3006. https://doi.org/10.7717/peerj.3006

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet Journal*, *17*(1), 10–12. https://doi.org/10.14806/ej.17.1.200

Masuda, R. (2008). Seasonal and interannual variation of subtidal fish assemblages in Wakasa Bay with reference to the warming trend in the sea of Japan. *Environmental Biology of Fishes*, *82*, 387–399.

Masuda, R., Shiba, M., Yamashita, Y., Ueno, M., Kai, Y., Nakanishi, A., Torikoshi, M., & Tanaka, M. (2010). Fish assemblages associated with three types of artificial reefs: Density of assemblages and possible impacts on adjacent fish abundance. *Fishery Bulletin*, *108*(2), 162–173.

McMurdie, P. J., & Holmes, S. (2013). Phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE*, *8*(4), e61217.

McMurdie, P. J., & Holmes, S. (2014). Waste not, want not: Why rarefying microbiome data is inadmissible. *PLoS Computational Biology*, *10*(4), e1003531.

Minamoto, T., Miya, M., Sado, T., Seino, S., Doi, H., Kondoh, M., Nakamura, K., Takahara, T., Yamamoto, S., Yamanaka, H., Araki, H., Iwasaki, W., Kasai, A., Masuda, R., & Uchii, K. (2021). An illustrated manual for environmental DNA research: Water sampling guidelines and experimental protocols. *Environmental DNA*, *3*(1), 8–13. https://doi.org/10.1002/edn3.121

Minamoto, T., Yamanaka, H., Takahara, T., Honjo, M. N., & Kawabata, Z. (2011). Surveillance of fish species composition using environmental DNA. *Limnology*, *13*(2), 193–197. https://doi.org/10.1007/s10201-011-0362-4

Minardi, D., Ryder, D., del Campo, J., Garcia Fonseca, V., Kerr, R., Mortensen, S., Pallavicini, A., & Bass, D. (2022). Improved high throughput protocol for targeting eukaryotic symbionts in metazoan and eDNA samples. *Molecular Ecology Resources*, *22*(2), 664–678. https://doi.org/10.1111/1755-0998.13509

Miya, M., Minamoto, T., Yamanaka, H., Oka, S., Sato, K., Yamamoto, S., Sado, T., & Doi, H. (2016). Use of a filter cartridge for filtration of water samples and extraction of environmental DNA. *Journal of Visualized Experiments*, *117*, e54741. https://doi.org/10.3791/54741

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 more than 230 subtropical marine species. *Royal Society Open Science*, *2*(7), 150088. https://doi.org/10.1098/rsos.150088

Nakagawa, H. (2019). A data set of long-term observation of aquatic vertebrates and associate environmental factors in a temperate stream, the Yura River, Kyoto, Japan. *Ecological Research*, *34*(1), 8. https://doi.org/10.1111/1440-1703.1059

O'Donnell, J. L., Kelly, R. P., Lowell, N. C., & Port, J. A. (2016). Indexed PCR primers induce template-specific bias in large-scale DNA sequencing studies. *PLoS ONE*, *11*(3), e0148698.

Oksanen, J., Kindt, R., Legendre, P., O'Hara, B., Simpson, G. L., Solymos, P., Stevens, M. H. H., & Wagner, H. (2008). *Vegan: Community ecology package (R package)*. http://cran.r-project.org/, http://vegan.r-forge.r-project.org/

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.r-project.org/

Samejima, H., Ong, R., Lagan, P., & Kitayama, K. (2012). Camera-trapping rates of mammals and birds in a Bornean tropical rainforest under sustainable forest management. *Forest Ecology and Management*, *270*, 248–256. https://doi.org/10.1016/j.foreco.2012.01.013

Sato, Y., Miya, M., Fukunaga, T., Sado, T., & Iwasaki, W. (2018). MitoFish and MiFish pipeline: A mitochondrial genome database of fish with an analysis pipeline for environmental DNA metabarcoding. *Molecular Biology and Evolution*, *35*(6), 1553–1555. https://doi.org/10.1093/molbev/msy074

Schrader, C., Schielke, A., Ellerbroek, L., & Johne, R. (2012). PCR inhibitors – Occurrence, properties and removal. *Journal of Applied Microbiology*, *113*(5), 1014–1026. https://doi.org/10.1111/j.1365-2672.2012.05384.x

Shen, W., Le, S., Li, Y., & Hu, F. (2016). SeqKit: A cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLoS ONE*, *11*(10), e0163962. https://doi.org/10.1371/journal.pone.0163962

Sigsgaard, E. E., Carl, H., Møller, P. R., & Thomsen, P. F. (2015). Monitoring the near-extinct European weather loach in Denmark based on environmental DNA from water samples. *Biological Conservation*, *183*, 46–52. https://doi.org/10.1016/j.biocon.2014.11.023

Smets, W., Leff, J. W., Bradford, M. A., McCulley, R. L., Lebeer, S., & Fierer, N. (2016). A method for simultaneous measurement of soil bacterial abundances and community composition via 16S rRNA gene sequencing. *Soil Biology and Biochemistry*, *96*, 145–151. https://doi.org/10.1016/j.soilbio.2016.02.003

Snyder, M. R., & Stepien, C. A. (2020). Increasing confidence for discerning species and population compositions from metabarcoding assays of environmental samples: Case studies of fishes in the Laurentian Great Lakes and Wabash River. *Metabarcoding and Metagenomics*, *4*, e53455. https://doi.org/10.3897/mbmg.4.53455

Taberlet, P., Bonin, A., Zinger, L., & Coissac, E. (2018). *Environmental DNA: For biodiversity research and monitoring*. OUP Oxford.

Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C., & Willerslev, E. (2012). Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology*, *21*(8), 2045–2050. https://doi.org/10.1111/j.1365-294X.2012.05470.x

Tanabe, A. S., & Toju, H. (2013). Two new computational methods for universal DNA barcoding: A benchmark using barcode sequences of bacteria, archaea, animals, fungi, and land plants. *PLoS ONE*, *8*(10), e76910. https://doi.org/10.1371/journal.pone.0076910

Ushio, M. (2022). Interaction capacity as a potential driver of community diversity. *Proceedings of the Royal Society B: Biological Sciences*, *289*(1969), 20212690. https://doi.org/10.1098/rspb.2021.2690

Ushio, M., Fukuda, H., Inoue, T., Makoto, K., Kishida, O., Sato, K., Murata, K., Nikaido, M., Sado, T., Sato, Y., Takeshita, M., Iwasaki, W., Yamanaka, H., Kondoh, M., & Miya, M. (2017). Environmental DNA enables detection of terrestrial mammals from forest pond water. *Molecular Ecology Resources*., *17*, e63–e75. https://doi.org/10.1111/1755-0998.12690

Ushio, M., Murakami, H., Masuda, R., Sado, T., Miya, M., Sakurai, S., Yamanaka, H., Minamoto, T., & Kondoh, M. (2018). Quantitative monitoring of multispecies fish environmental DNA using high-throughput sequencing. *Metabarcoding and Metagenomics*, *2*, e23297. https://doi.org/10.3897/mbmg.2.23297

Wang, Q., Garrity, G. M., Tiedje, J. M., & Cole, J. R. (2007). Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology*, *73*(16), 5261–5267. https://doi.org/10.1128/AEM.00062-07

Wright, E. S. (2016). Using DECIPHER v2.0 to analyze big biological sequence data in R. *The R Journal*, *8*(1), 352–359.

Wu, J.-H., Hong, P.-Y., & Liu, W.-T. (2009). Quantitative effects of position and type of single mismatch on single base primer extension. *Journal of Microbiological Methods*, *77*(3), 267–275. https://doi.org/10.1016/j.mimet.2009.03.001

Yamahara, K. M., Preston, C. M., Birch, J., Walz, K., Marin, R. I., Jensen, S., Pargett, D., Roman, B., Ussler, W. I., Zhang, Y., Ryan, J., Hobson, B., Kieft, B., Raanan, B., Goodwin, K. D., Chavez, F. P., & Scholin, C. (2019). In situ autonomous acquisition and preservation of marine environmental DNA using an autonomous underwater vehicle. *Frontiers in Marine Science*, *6*, 373. https://doi.org/10.3389/fmars.2019.00373

Yamamoto, S., Masuda, R., Sato, Y., Sado, T., Araki, H., Kondoh, M., Minamoto, T., & Miya, M. (2017). Environmental DNA metabarcoding reveals local fish communities in a species-rich coastal sea. *Scientific Reports*, *7*, 40368. https://doi.org/10.1038/srep40368

Yonezawa, S., Ushio, M., Yamanaka, H., Miya, M., Takayanagi, A., & Isagi, Y. (2020). Environmental DNA metabarcoding reveals the presence of a small, quick-moving, nocturnal water shrew in a forest stream. *Conservation Genetics*, *21*(6), 1079–1084. https://doi.org/10.1007/s10592-020-01310-5

Zizka, V. M. A., Elbrecht, V., Macher, J.-N., & Leese, F. (2019). Assessing the influence of sample tagging and library preparation on DNA metabarcoding. *Molecular Ecology Resources*, *19*(4), 893–899. https://doi.org/10.1111/1755-0998.13018

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Ushio, M., Furukawa, S., Murakami, H., Masuda, R., & Nagano, A. J. (2022). An efficient early-pooling protocol for environmental DNA metabarcoding. *Environmental DNA*, *4*, 1212–1228. https://doi.org/10.1002/edn3.337