bioRxiv preprint doi: https://doi.org/10.1101/113472; this version posted March 5, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Quantitative monitoring of multispecies fish environmental DNA using high-throughput sequencing

Masayuki Ushio*,[1,2] Hiroaki Murakami,[3] Reiji Masuda,[3] Tetsuya Sado,[4] Masaki Miya,[4] Sho Sakurai,[5] Hiroki Yamanaka,[1,6] Toshifumi Minamoto,[7] and Michio Kondoh[1]

[1] *Department of Environmental Solution Technology, Faculty of Science and Technology, Ryukoku University, 1-5 Yokotani, Seta Oe-cho, Otsu, Shiga 520-2194, Japan*

[2] *Joint Research Center for Science and Technology, Ryukoku University, Otsu, Shiga 520-2194, Japan**

[3] *Maizuru Fisheries Research Station, Kyoto University, Maizuru, Kyoto 625-0086, Japan*

[4] *Department of Ecology and Environmental Sciences, Natural History Museum and Institute, Chiba 260-8682, Japan*

[5] *Graduate School of Science and Technology, Ryukoku University, Otsu 520-2194, Japan*

[6] *The Research Center for Satoyama Studies, Ryukoku University, 1-5 Yokotani, Seta Oe-cho, Otsu, Shiga 520-2194, Japan*

[7] *Graduate School of Human Development and Environment, Kobe University, Hyogo 657-8501, Japan*

In the present study, we added internal standard DNAs (i.e., quantified short DNA fragments from fish species that have never been observed in a sampling area) to environmental DNA (eDNA) samples, which were collected weekly from a coastal marine ecosystem in Maizuru-Bay, Japan (from April 2015 to March 2016), and performed metabarcoding analysis to identify fish species and quantify fish eDNA copy number simultaneously. A standard curve was drawn for each sample using the number of reads and the added amount of the standard DNA, which was used to convert the reads of eDNA to the copy numbers. The converted copy numbers showed significant positive correlation with those determined by quantitative PCR, suggesting that eDNA metabarcoding with standard DNA enabled the quantification of eDNA as accurately as quantitative PCR. Furthermore, for samples that show a high level of PCR inhibition, eDNA metabarcoding with internal standard DNAs might allow more accurate quantification than qPCR because the standard curves drawn for internal standard DNAs would include the effect of PCR inhibition. A single run of Illumina MiSeq produced 70 quantitative fish eDNA time series in our study, showing that our method would contribute to more efficient and quantitative monitoring of biodiversity.

Keywords: biodiversity monitoring; environmental DNA; quantification; internal standard DNA; qPCR; metabarcoding

I. INTRODUCTION

Effective ecosystem conservation and resource management require quantitative monitoring of biodiversity, including accurate descriptions of species composition and temporal variations of species abundance. Accordingly, quantitative monitoring of biodiversity has often been performed for many ecosystems. For example, fishing (in aquatic ecosystems), camera/video trap method (in terrestrial ecosystems) and direct visual census (in aquatic and terrestrial ecosystems) have traditionally been used as tools for biodiversity monitoring [1, 2]. These data are invaluable in conservation ecology, but at the same time, the traditional approaches are usually time-, effort- and cost-consuming. In addition, most of the traditional methods require professional expertise such as taxonomic identification skill in the field. These difficulties prevent the collection of quantitative, comprehensive (i.e., multispecies and frequent) and long-term monitoring data of biodiversity.

Researchers have recently been using environmental DNA (eDNA), which is a genetic material that is derived from organisms living in that habitat (and found in an environment such as water), to detect the presence of macro-organisms, particularly those living in an aquatic environment. In the case of macro-organisms, eDNA originates from various sources such as metabolic waste, or damaged tissue [3], and the eDNA contains information about the species identity of organisms that produced it. Since the first application of eDNA analysis to natural ecosystems [4], eDNA in aquatic ecosystems has been used in many studies as a monitoring tool for

* Present address: Center for Ecological Research, Kyoto University, 2-509-3 Hirano, Otsu, 520-2113 Japan/PRESTO, Japan Science and Technology Agency; Correspondence to ong8181@gmail.com

bioRxiv preprint doi: https://doi.org/10.1101/113472; this version posted March 5, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

investigations of the distributions of fish species in ponds, rivers and seawater [5–9] as well as the distributions of other aquatic/semiaquatic/terrestrial vertebrates [10, 11]. Furthermore, researchers have begun to apply high-throughput sequencing technology (e.g., Illumina MiSeq) and universal primer sets to eDNA studies [11–15]. A previous study demonstrated that an eDNA metabarcoding approach using fish-targeting universal primers (MiFish primers) enabled the detection of more than 230 fish species from seawater in a single study [14]. Accordingly, the eDNA metabarcoding approach has become a cost- and labor-effective approach for capturing aquatic biodiversity.

Though the eDNA metabarcoding approach greatly improved the efficiency of biodiversity monitoring, several potential limitations prevent its use as a tool for quantitative monitoring of biodiversity. First, whether the quantity of eDNA is a reliable index of the abundance (or biomass) of macro-organisms is still controversial. Second, the number of eDNA sequence reads obtained by high-throughput sequencing is not an index of the quantity of eDNA, and thus, we cannot estimate the quantity of eDNA in an environment by the eDNA metabarcoding approach. Regarding the first issue, recent studies showed that eDNA quantity is a good indicator of the abundance or biomass of macro-organisms, at least under a particular condition [7, 16]. Therefore, although still controversial, we may use the quantity of eDNA as an index of abundance/biomass with careful interpretations. Regarding the second issue, some potential approaches to solve this problem were reported in the field of microbial ecology. For example, Smets et al. [17] added an internal standard DNA (DNA of a microbial species that had never been found in a sample) of known quantity to a soil sample. They used the number of sequence reads of the internal standard DNA to estimate the sequence reads per the number of DNA copies, and the sequence reads of DNAs from unknown microbial species (i.e., non-standard microbial species) were converted to the number of microbial DNA copies. The total number of microbial DNA copies estimated was significantly positively correlated with other reliable and quantitative indices of soil microbial abundance. This approach is potentially useful to solve the second issue, and promising for performing quantitative and effective biodiversity monitoring.

In the present study, we applied the internal standard DNA method to the eDNA metabarcoding approach in order to enable quantitative biodiversity monitoring of a natural fish community in a coastal marine ecosystem (i.e., identification of fish species and quantification of the number of fish eDNA copies simultaneously). Water samples were collected weekly from a sampling station in Maizuru Bay, along the Japan Sea coast of central Japan, and eDNAs were extracted from the samples. We added amplified, purified and quantified short DNA fragments derived from five fish species that have never been observed in the sampling region (freshwater fish species in Southeast Asia or Africa) as internal standard DNAs to eDNA samples. Then, a mitochondrial 12S region was amplified using fish-targeting universal primers [14] and sequenced on the Illumina MiSeq platform. A standard curve was drawn between the number of sequence reads obtained and the added amount (copy number) of the standard DNA, and this curve was used to convert number of sequence reads of eDNA samples to the number of eDNA copies. This converted eDNA copy number was compared with the number of eDNA copies estimated by the quantitative PCR method. The results showed that the eDNA metabarcoding with internal standard DNAs enabled the quantification of multi-species ($>$ 70 species) fish eDNAs by a single run, and we propose that our approach is promising for quantitative and effective monitoring of fish eDNA dynamics, which is potentially related to the abundance/biomass of fish species, in an aquatic ecosystem.

II. METHODS

A. *Study site*

Water samples were collected at a floating pier in the Maizuru Fishery Research Station of Kyoto University (Nagahama, Maizuru, Kyoto, Japan: 35°28'N, 135°22'E; Fig.1). The sampling point was located 11 m from the shore, with a bottom depth of 4 m. The adjacent area included a rocky reef, brown algae macrophyte and filamentous epiphyte vegetation, live oysters Crassostrea gigas and their shells, a sandy or muddy silt bottom and an artificial vertical structure that functioned as a fish reef. The surface water temperature and salinity in the area ranged from 1.2 to 30.8°C and from 4.14 to 34.09‰, respectively. The mean ($\pm SD$) surface salinity was 30.0 $\pm$2.9‰(n = 1,753) and did not show clear seasonality. Further information on the study area is available in Masuda et al. [1, 18].

B. *Water sampling and DNA extraction*

All sampling and filtering equipment was washed with a 10% commercial bleach solution before use. We collected 1,000 ml of seawater once a week from

bioRxiv preprint doi: https://doi.org/10.1101/113472; this version posted March 5, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

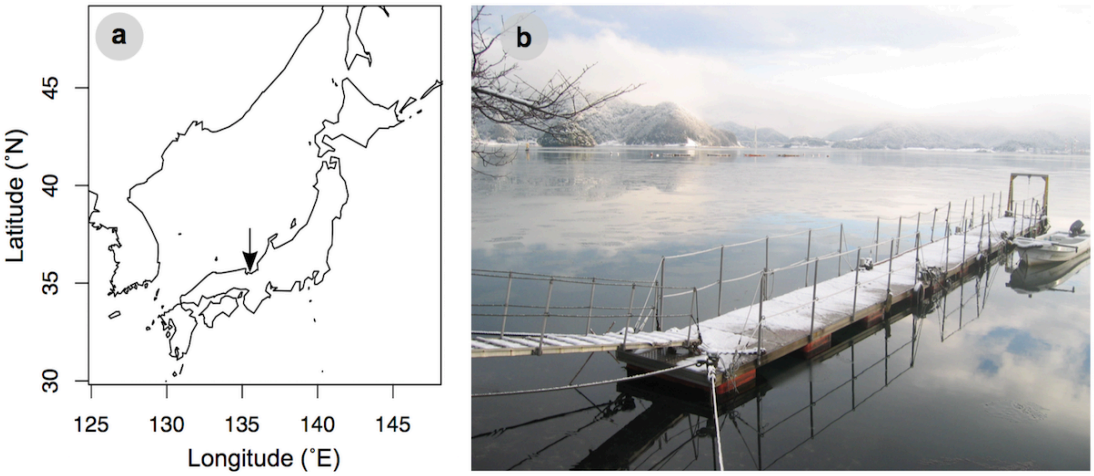

Figure 1. Location of the research site (**a**). The arrow indicates our research site. A floating pier in the Maizuru Fishery Research Station of Kyoto University, Maizuru, Kyoto, Japan, where the weekly water sampling was performed (**b**). Photo taken in winter season by R. Masuda.

a pier (Fig.1b) in the study area using a polyethylene bottle from 7th April 2015 to 29th March 2016. Thus, the number of total eDNA samples (excluding artificial seawater samples as negative controls) was 52. The collected water samples were immediately taken back to the laboratory and filtered using 47-mm diameter glass-fibre filters (nominal pore size, 0.7 μ m; Whatman, Maidstone, UK). The sampling bottles were gently shaken before the filtration. After the filtration, each filter was wrapped in commercially available aluminium foil and stored at -20°C before eDNA extraction. Artificial seawater (1,000 ml) was used as the negative control, and sampling bottles filled with artificial seawater were treated identically to the eDNA samples in order to monitor contamination during the bottle handling, water filtering and subsequent DNA extraction. Negative controls were taken once a month (total 12 negative controls), and all negative controls produced a negligible number of sequences (i.e., the average number of sequence reads is 2,305 for environmental samples, while that is 21 for negative controls; see "Read w/o standard" column in Table S1).

DNA was extracted from the filters using a DNeasy Blood and Tissue Kit (Qiagen, Hilden, Germany) in combination with a spin column (EZ-10; Bio Basic, Markham, Ontario, Canada). After removal of the attached membrane from the spin column (EZ-10), the filter was tightly folded into a small cylindrical shape and placed in the spin column. The spin column was centrifuged at 6,000 g for 1 min to remove excess water from the filter. The column was then placed in the same 2-ml tube and subjected to lysis using proteinase K. For the lysis, sterilized H_2O (200 μl), proteinase K (10 μl) and buffer AL (100 μl) were mixed, and the mixed solution was gently pipetted onto the folded filter in the spin column. The column was then placed on a 56°C preheated aluminium heat block and incubated for 30 min. After the incubation, the spin column was centrifuged at 6,000 g for 1 min to collect DNA. In order to increase DNA yields from the filter, 200 μl of sterilized TE buffer was gently pipetted onto the folded filter and the spin column was again centrifuged at 6,000 g for 1 min. The collected DNA solution (*ca.* 100 μl) was purified using a DNeasy Blood and Tissue Kit following the manufacturer ' s protocol.

C. *Preparation of standard fish DNAs*

Extracted DNAs of five fish species (*Saurogobio immaculatus*, *Elopichthys bambusa*, *Carassioides acuminatus*, *Labeo coubie*, and *Acanthopsoides gracilentus*) that are all freshwater fishes from Southeast Asia or Africa and have never occurred in the sampling region were used as internal standard DNAs. A target region (mitochondrial 12S rRNA) of the extracted DNA was amplified using MiFish primers (without MiSeq adaptors), and the amplified and purified target DNA (*ca.* 220 bp) was excised using E-Gel SizeSelect (ThermoFisher Scientific, Waltham, MA, USA). The DNA size distribution of the library was estimated using an Agilent 2100 BioAnalyzer (Agilent, Santa Clara, CA, USA), and the concentration of double-stranded DNA of the library was quantified using a Qubit dsDNA HS assay kit and a Qubit fluorometer (ThermoFisher Scientific,

bioRxiv preprint doi: https://doi.org/10.1101/113472; this version posted March 5, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Waltham, MA, USA). Based on the quantification values obtained using the Qubit fluorometer, we adjusted the copy number of the standard DNAs and added these DNAs as follows: *S. immaculatus* (500 copies/μl), *E. bambusa* (250 copies/μl), *C. acuminatus* (100 copies/μl), *L. coubie* (50 copies/μl), and *A. gracilentus* (25 copies/μl). The numbers of internal standard DNA copies added to samples were determined by quantification of the number of total fish eDNA copies (i.e., MiFish primer target region) using the SYBR-GREEN quantitative PCR method (see below for the detailed method).

D. *Paired-end library preparation*

Work-spaces and equipment were sterilized prior to the library preparation, filtered pipet tips were used, and separation of rooms for pre- and post-PCR was carried out to safeguard against cross-contamination. We also employed negative controls to monitor contamination during the experiments.

The first-round PCR (1st PCR) was carried out with a 12-μl reaction volume containing 6.0 μl of 2 × KAPA HiFi HotStart ReadyMix (KAPA Biosystems, Wilmington, WA, USA), 0.7 μl of each primer (5 μM), 0.6 μl of sterilized H_2O, 2 μl of standard DNA and 2.0 μl of template. The final concentration of each primer (MiFish-U-F/R) was 0.3 μM. The sequences of MiFish primers are: GTC GGT AAA ACT CGT GCC AGC (MiFish-U-F) and CAT AGT GGG GTA TCT AAT CCC AGT TTG (MiFish-U-R). MiSeq sequencing primers and six random bases (N) were combined to MiFish-U primers [see 14 for detailed sequences]. The six random bases were used to enhance cluster separation on the flowcells during initial base call calibrations on the MiSeq platform. The thermal cycle profile after an initial 3 min denaturation at 95°C was as follows (35 cycles): denaturation at 98°C for 20 s; annealing at 65°C for 15 s; and extension at 72°C for 15 s, with a final extension at the same temperature for 5 min. We performed triplicate first-PCR, and the replicates were pooled in order to mitigate the PCR dropouts. The pooled first PCR products were purified using Exo-SAPIT (Affymetrix, Santa Clara, CA, USA). The pooled, purified, and 10-fold diluted first PCR products were used as templates for the second PCR.

The second-round PCR (2nd PCR) was carried out with a 24-μl reaction volume containing 12 μl of 2 × KAPA HiFi HotStart ReadyMix, 1.4 μl of each primer (5 μM), 7.2 μl of sterilized H_2O and 2.0 μl of template. Different combinations of forward and reverse indices were used for different templates (samples) for massively parallel sequencing with MiSeq. The thermal cycle profile after an initial 3 min denaturation at 95°C was as follows (12 cycles): denaturation at 98°C for 20 s; annealing and extension combined at 72°C (shuttle PCR) for 15 s, with the final extension at 72°C for 5 min. The products of the second PCR were combined, purified, excised and sequenced on the MiSeq platform using a MiSeq v2 Reagent Nano Kit for 2 × 150 bp PE.

E. *Sequence read processing and taxonomic assignment*

The overall quality of the MiSeq reads was evaluated, and the reads were assembled using the software FLASH with a minimum overlap of 10 bp [19]. The assembled reads were further filtered and cleaned, and the pre-processed reads were subjected to the clustering process and taxonomic assignments. The pre-processed reads from the above custom pipeline were dereplicated using UCLUST [20]. Those sequences represented by at least 10 identical reads were subjected to the downstream analyses, and the remaining under-represented sequences (with less than 10 identical reads) were subjected to pairwise alignment using UCLUST. If the latter sequences (observed from less than 10 reads) showed at least 99% identity with one of the former reads (i.e., no more than one or two nucleotide differences), they were operationally considered as identical (owing to sequencing or PCR errors and/or actual nucleotide variations in the populations).

The processed reads were subjected to local BLASTN searches against a custom-made database [21]. The custom-made database was generated as described in the previous study [14]. The top BLAST hit with a sequence identity of at least 97% and E-value threshold of 10^{-5} was applied to species assignments of each representative sequence. The detailed information for the above bioinformatics pipeline from data pre-processing through taxonomic assignment is available in the supplemental information in Miya et al. [14] and Ushio et al. [11]. Also, online version of this pipeline is available at http://mitofish.aori.u-tokyo.ac.jp/mifish.

F. *Determination of the number of eDNA copies by quantitative PCR*

The copy number of total fish eDNA was quantified using the SYBR-GREEN qPCR method using a StepOne-Plus™ Real-Time PCR system (Applied Biosystems, Foster City, CA, USA). SYBR-GREEN qPCR was conducted in a 10-μL volume and the reaction solution consisted of 5 μl of PowerUpTM SYBR® GREEN Master Mix (Thermo Fisher Scien-

bioRxiv preprint doi: https://doi.org/10.1101/113472; this version posted March 5, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

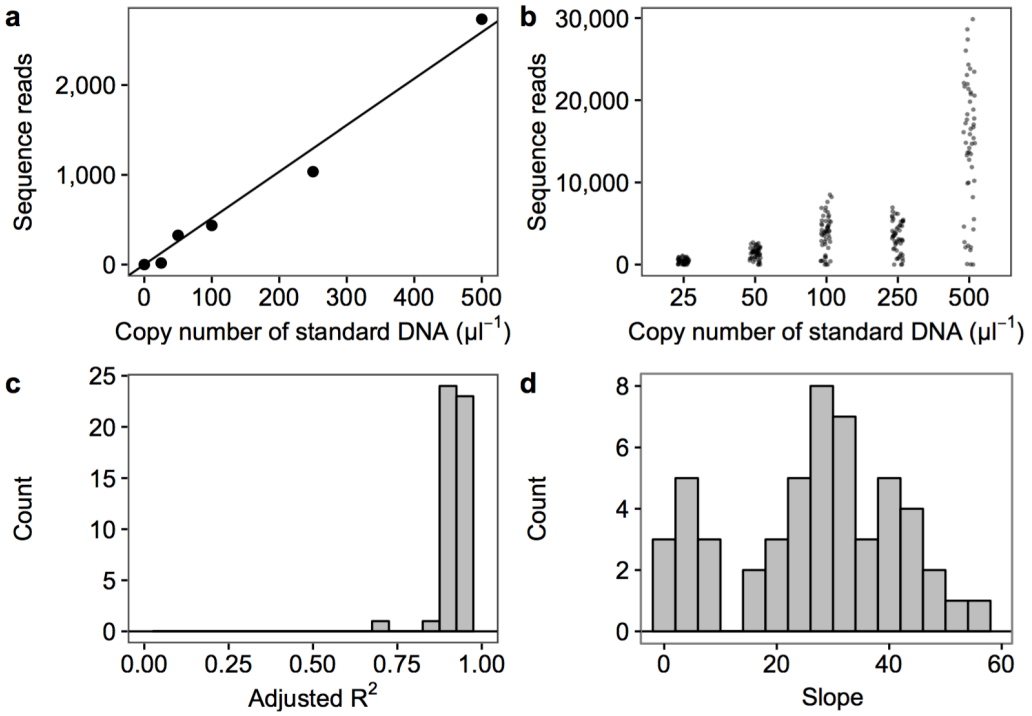

Figure 2. Summary of standard curves constructed using the number of copies added and sequence reads of internal standard DNAs. An example of a standard curve (**a**). The line indicates the regression line between the copy number of standard DNA (copies/μl) and sequence reads. The intercept of the regression line was set as zero. Distributions of sequence reads of internal standard DNAs (**b**). Distribution of adjusted R^2 of the standard curves (**c**). Note that a standard curve was drawn for each eDNA sample and that the number of standard curves is equal to the number of eDNA samples (N = 52). Distribution of slopes of standard curves (**d**).

tific, Wilmington, DE, USA), 0.6 μl of 5 μM MiFish-F/R primers (without adaptor), 10 μl of sterilized H_2O, and 1.0 μl of DNA template. SYBR-GREEN qPCR was performed in triplicate for each eDNA sample, standard dilution series, and PCR negative controls. The standard dilution series was prepared using DNA extracted from *Capoeta capoeta* (a freshwater fish species in Southeast Asia). We selected *C. capoeta* as the standard because the length of the MiFish region of this species is close to the average length in fish species (*C. capoeta* = 174 bp, the average length of the MiFish region = *ca.* 173 bp). The thermal cycle profile after preconditioning for 2 min at 50°C and 2 min at 95°C was as follows (40 cycles): denaturation at 95°C for 3 s; annealing and extension combined at 60°C (shuttle PCR) for 30 s. In all experiments, PCR negative controls showed no detectable amplifications.

In addition, fish-species-specific eDNA was quantified by real-time TaqMan® PCR according to Takahara et al. [5] using a StepOne-Plus™ Real-Time PCR system. The cytochrome b region of mitochondrial DNA was targeted for amplification from eDNA samples for each target species by using the following primer sets and associated probes which were designed and confirmed to be able to amplify each target species-specifically by Minamoto et al. (unpublished). In the TaqMan qPCR analysis, *Engraulis japonicus* and *Trachurus japonicas* were chosen because they are abundant in the study area and standard dilution series were already available. For *Engraulis japonicas*, primers Eja-CytB-Forward (5′-GAA AAA CCC ACC CCC TAC TCA-3′), Eja-CytB-Reverse (5′-GTG GCC AAG CAT AGT CCT AAA AG-3′), and Eja-CytB-Probe (5′-FAM-CGC AGT AGT AGA CCT CCC AGC ACC ATC C-TAMRA-3′) were used. For *Trachurus japonicas*, primers Tja-CytB-Forward (5′-CAG ATA TCG CAA CCG CCT TT-3′), Tja-CytB-Reverse (5′-CCG ATG TGA AGG TAA ATG CAA A-3′), and Tja-CytB-Probe (5′-FAM-TAT GCA CGC CAA CGG CGC CT-TAMRA-3′) were used. The length of the PCR amplicon produced by each primer set was 115 bp and 127 bp for *Engraulis japonicas* and for *Trachurus japonicas*, respectively. PCR was conducted in a 15-μl volume containing each primer at 900 nM,

bioRxiv preprint doi: https://doi.org/10.1101/113472; this version posted March 5, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

TaqMan® probe at 125 nM, and 2 μl of sample DNA in 1 × PCR master mix (TaqMan® gene expression master mix; Life Technologies, Carlsbad, CA, USA). A dilution series of standards was prepared for quantification and analyzed at the concentrations of 3×10^1 to 3×10^4 copies per well in each experiment to obtain standard curves. The standards were pTAKN-2 plasmids containing commercially synthesized artificial DNA that had the same sequence as the amplification region of each species. The thermal cycle profile after preconditioning for 2 min at 50°C and 10 min at 95°C was as follows (55 cycles): denaturation at 95°C for 15 s; combined annealing and extension at 60°C (shuttle PCR) for 60 s. qPCR was performed in triplicate for each eDNA sample, standard dilution series, and PCR negative controls. In all experiments, PCR negative controls showed no detectable amplification.

G. *Statistical analyses*

For all analyses, the free statistical environment R was used [22]. A linear regression was used to draw a standard curve between sequence reads and copy number of the standard DNA. The relationships between the numbers of eDNA copy estimated by qPCR and MiSeq sequencing were also analysed using a linear regression. Linear relationships were considered significant if P values were smaller than 0.05.

III. RESULTS AND DISCUSSION

A. *Relationship between the copy number and sequence reads of the standard DNA*

The sequence reads of the internal standard DNAs were significantly positively correlated with the copy number of those DNAs (Fig.2a,b). A standard curve was drawn for each eDNA sample, and therefore, the number of standard curves equals the number of eDNA samples (= 52). R^2 values of the standard curves ranged from 0.71 to 0.98, and more than 80% of standard curves showed R^2 values higher than 0.9 (Fig.2c), suggesting that sequence reads were proportional to the number of DNA copies in a single sample and that the standard curve can be used to convert sequence reads to the number of DNA copies. Interestingly, the slopes of the standard curves (i.e., sequence reads per DNA copy) were highly variable, ranging from 0 to 54.1 with a median value of 24.6 (Fig.2d). Low slope values (e.g., 0, or close to 0) indicate that internal standard DNAs were not efficiently amplified even if the number of DNA copies added was large, suggesting the presence of PCR inhibitor(s) (e.g., humic substance) in the eDNA samples. Also, these variations in the slope suggested that the degree of PCR inhibition varies depending on the eDNA sample.

B. *Quantification of the copy number using sequence reads and standard curve*

MiSeq sequence reads of each sample were converted using each standard curve (i.e., the number of eDNA copies [copies/μl] = MiSeq sequence reads/estimated regression slope; hereafter referred to as 'Converted MiSeq reads [copies/μl]'). Then, converted MiSeq reads were compared with the number of DNA copies quantified by qPCR (Fig.3, Table S2 and S3). The numbers of eDNA copies estimated by metabarcoding and qPCR were significantly and positively correlated with each other for total fish eDNA (Fig.3a, b). For the total fish eDNA, the number of eDNA copies quantified by qPCR (mean copy number = 683 copies/μl) was higher than that quantified by metabarcoding (mean copy number = 139 copies/μl). This is not surprising because we excised target amplicon fragments (*ca.* 370 bp including MiSeq adaptor), and we discarded non-target amplified fragments (e.g., longer and unknown amplicons), which were quantified by the SYBR-GREEN assay, before MiSeq sequencing. As for eDNA of Engraulis japonicus and Trachurus japonicus, we found that the number of eDNA copies quantified by metabarcoding were similar to that obtained by qPCR (i.e., regression lines were close to the 1:1 line in Fig.3c-f), suggesting that eDNA metabarcoding with the inclusion of internal standard DNAs reliably quantified the number of eDNA copies.

Although the converted MiSeq reads generally corresponded well with the eDNA copy number estimated by qPCR, the converted MiSeq reads of some samples were much higher than the copy number obtained by qPCR (i.e., points close to the x-axis in Fig.3). These samples showed relatively low values of slopes of standard curves (i.e., corresponded to points with darker colour), suggesting that there was an inhibition of PCR in these samples. eDNA metabarcoding with the inclusion of internal standard DNAs can control for PCR-inhibition effects in the estimation of eDNA copy, which may be advantageous compared with qPCR. Conversely, we would suggest that qPCR could not reliably quantify the number of eDNA copies when the influence of PCR inhibitors was strong. These mechanisms would result in some inconsistency between the two methods.

Some samples showed much lower eDNA copy numbers of *Trachurus japonicus* when quantified by

bioRxiv preprint doi: https://doi.org/10.1101/113472; this version posted March 5, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

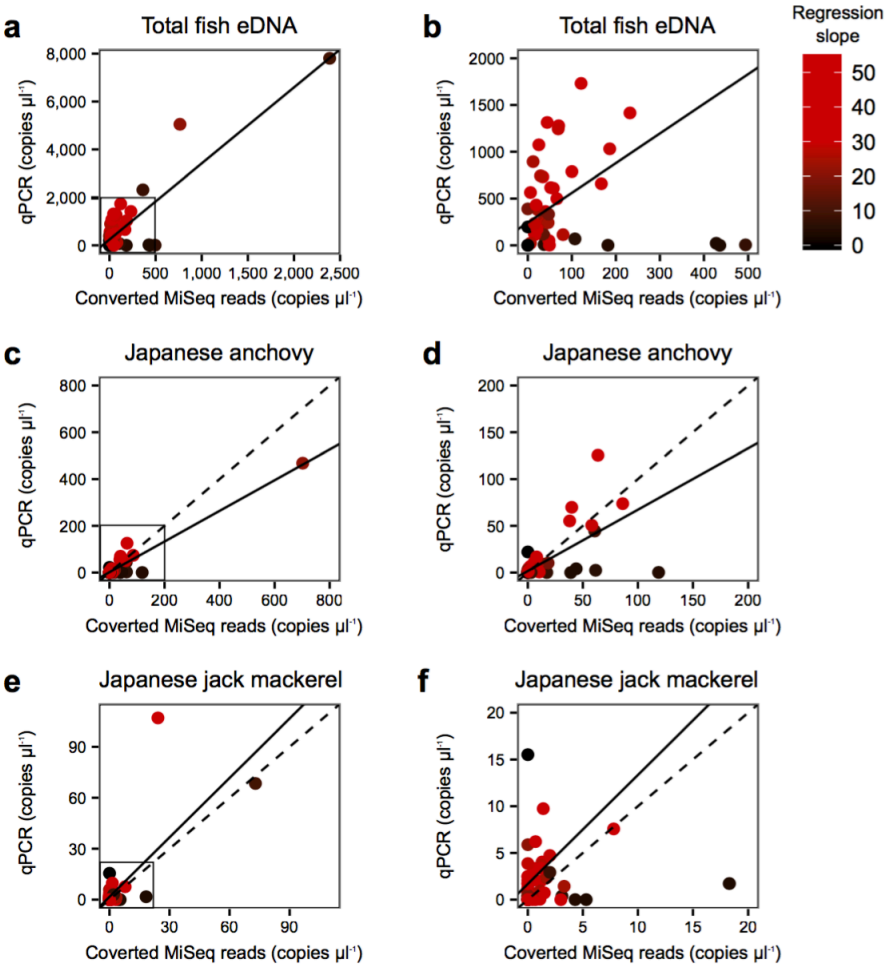

Figure 3. Relationship between the number of eDNA copies quantified by qPCR and that by MiSeq. Correlations for the total fish eDNA (all data, **a**; enlarged figure, **b**), Japanese anchovy (*Engraulis japonicus*; all data, **c**; enlarged figure, **d**), and Japanese jack mackerel (*Trachurus japonicus*; all data, **e**; enlarged figure, **f**). Dashed and solid lines indicate 1:1 line and linear regression line, respectively. Boxed regions in **a**, **c**, and **e** correspond to the range of the graphs in **b**, **d**, and **f**, respectively. The density of red colour indicates the slope of the standard curve used to convert MiSeq reads to the copy number.

eDNA metabarcoding than by qPCR (i.e., points close to the y-axis; Fig.3f). This inconsistency might have been due to the low eDNA copy number of *Trachurus japonicus* (all samples showed less than 100 copies/μl, and most samples showed less than 10 copies/μl). Metabarcoding might not be able to quantify such low numbers of eDNA copy accurately because the lowest copy number of internal standard DNA added was 25 copies/μl. If the copy number of internal standard DNA had been much lower, more accurate quantification would have been archived by eDNA metabarcoding. Furthermore, the difference in cycle number of PCR between qPCR (40-55 cycles) and 1st PCR (35 cycles) could contribute to the different sensitivities (i.e., detection limit) of these methods. Taken together, these results suggested that eDNA metabarcoding with internal standard DNA enabled simultaneous quantification and identification of fish eDNA, but the appropriate range of the copy number of the internal standard DNA should be carefully determined depending on the range of the target eDNA copy number in environmental samples.

C. *Quantitative and multispecies monitoring of fish eDNA using eDNA metabarcoding*

The dynamics of the fish eDNA quantified by converted MiSeq reads generally corresponded well with

bioRxiv preprint doi: https://doi.org/10.1101/113472; this version posted March 5, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

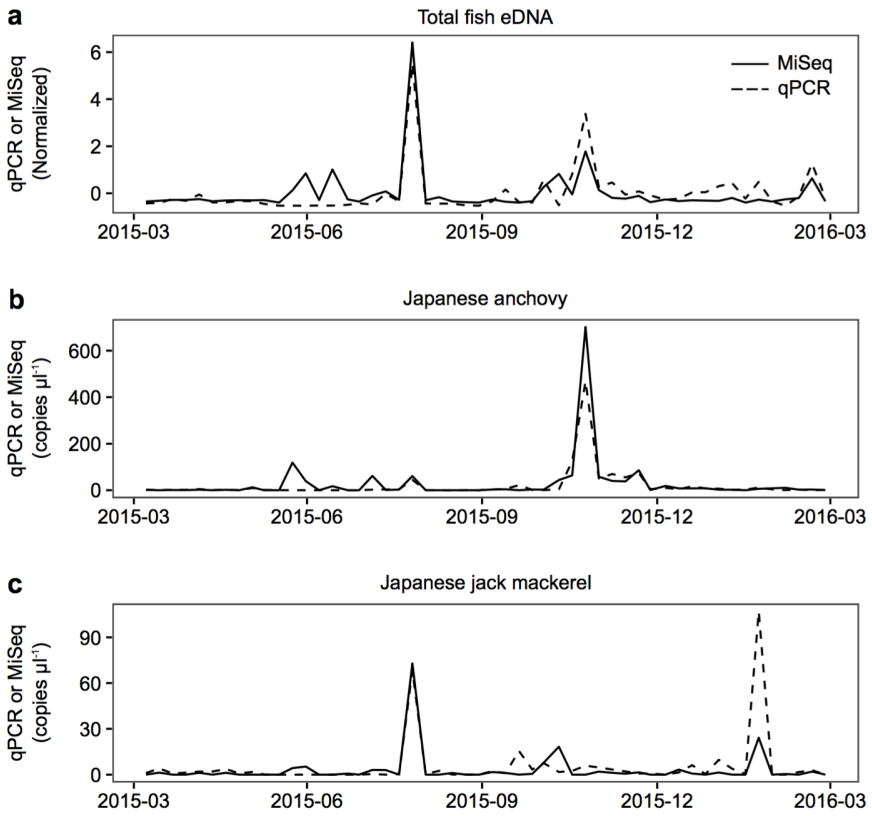

Figure 4. Dynamics of the total fish eDNA (**a**), Japanese anchovy (*Engraulis japonicus*, **b**) and Japanese jack mackerel (*Trachurus japonicus*, **c**) quantified by eDNA metabarcoding and qPCR. Solid and dashed lines indicate the number of eDNA copies quantified by eDNA metabarcoding and qPCR, respectively. Note that the copy number of total fish eDNA is normalized to have zero mean and unit variance.

those quantified by qPCR (Fig.4). For the total fish eDNA, the highest eDNA concentrations were found on 25th August and 24th November by qPCR, and peaks were also detected on those dates by MiSeq (Fig.4a). For *Engraulis japonicus* eDNA, the highest eDNA concentration was found on 24th November by qPCR, and the peak was also detected on this date by MiSeq (Fig.4b). For *Trachurus japonicus* eDNA, one of the highest eDNA concentrations (on 25th August) found by qPCR was also found by MiSeq. On the other hand, another peak found by qPCR (on 23rd February) was not detected by MiSeq (Fig.4c), probably due to the above-mentioned technical issues in eDNA metabarcoding with internal standard DNA.

The results obtained in the present study suggest that eDNA metabarcoding with the inclusion of internal standard DNA can reliably recover the dynamics of fish eDNA. Because eDNA metabarcoding can detect many species (sometimes more than 100 species) in a single run [14], this method enables simultaneous quantifications of eDNA derived from many fish species. In the present study, we detected more than 70 fish species from 52 samples collected from April 2015 to March 2016 in Maizuru Bay, Kyoto, Japan (Supplementary text, Table S2), which is generally consistent with long-term direct visual observations, e.g., a visual census detected a total of 50 fish species [1].

Our method enables the generation of a quantitative time series of eDNA of these fish species by a single MiSeq run, and as an example, an eDNA time series of the 10 most abundant fish species in terms of eDNA concentration is shown in Fig.5. Because eDNA copy numbers may be a proxy for fish biomass/abundance [7, 16], such a multispecies quantitative time series would provide valuable information on the dynamics of fish populations in the sampling area. Nonetheless, it should be mentioned that fish eDNA copy numbers are still "a proxy" or "an index" of fish biomass/abundance (or population size). In other words, evaluating the absolute biomass/abundance (e.g., the number of fish individuals present in a study area) of fish population is still challenging and this issue should be resolved in a future study.

bioRxiv preprint doi: https://doi.org/10.1101/113472; this version posted March 5, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

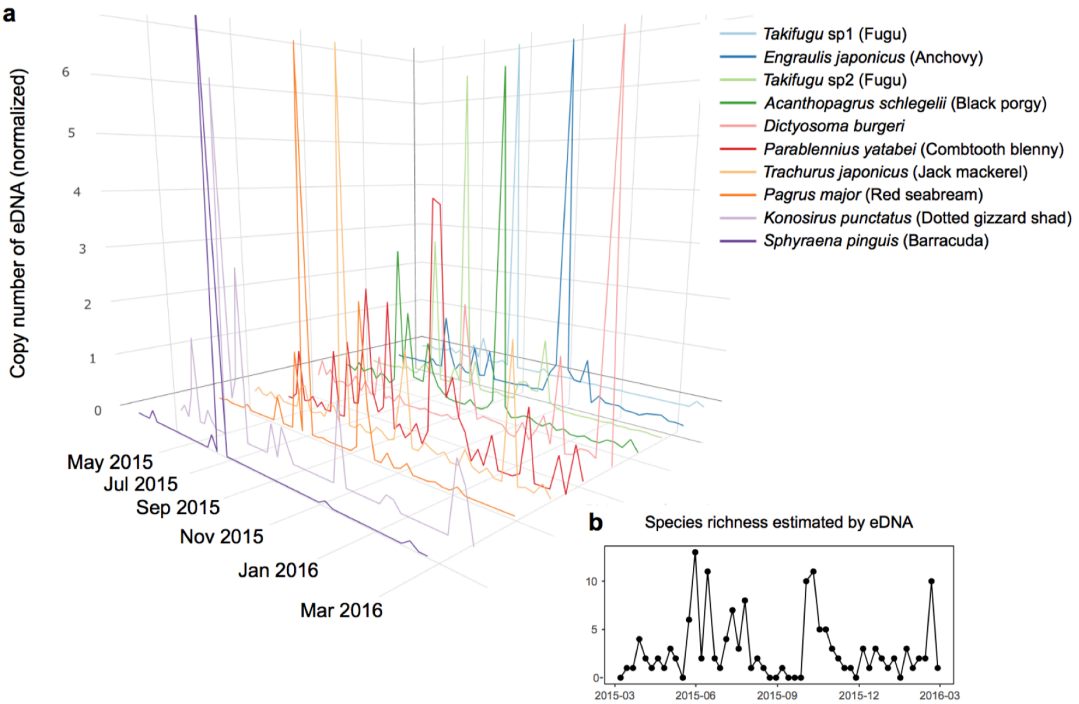

Figure 5. Quantitative and multispecies fish eDNA time series in Maizuru Bay, Kyoto, Japan. Time series of eDNA of 10 dominant fish species (**a**). In the eDNA analysis, two *Takifugu* species were detected as dominant species, and they were designated *Takifugu* sp1 and sp2. Representative sequence of *Takifugu* sp1 is highly similar to that of *T. niphobles*/*T. snyderi* (>99% identity). Representative sequence of *Takifugu* sp2 is identical with that of *T. pardadalis*/*T. xanthopterus*/*T. poecilonotus* (100% identity). Different colours indicate different fish species. The numbers of eDNA copies are normalized to have zero mean and unit variance. Time series of fish species richness detected by eDNA (**b**). Note that only fish species that showed a copy number of eDNA > 5 (copies/μl) were included to avoid potential minor contaminations in eDNA samples.

Although translating quantitative fish eDNA time series into fish biomass/abundance is still challenging, the eDNA time series measured here by eDNA metabarcoding were ecologically interpretable, suggesting that eDNA monitoring using our method would provide ecologically meaningful information on the dynamics of natural fish community at least in our case. For example, twice-a-month visual census detects generally high abundance and species richness in the summer, except for the highly abundant *Engraulis japonicus* in autumn [18] (Table S4); this corresponds well to the general trend of eDNA detection seen here (i.e., a high copy number and a high diversity of eDNA in June to August 2015; Fig.5). In addition, some species, such as Fugu species (*Takifugu* sp.1 [either species of *T. niphobles*/*T. snyderi*] and *Takifugu* sp.2 [either species of *T. pardadalis*/*T. xanthopterus*/*T. poecilonotus*]) and *Acanthopagrus schlegelii* were detected the whole year round by eDNA analysis, which is consistent with the detection in the visual census (Table S4). *Dictyosoma burgeri* was found only in winter in the visual census, and its eDNA was also most abundant in winter. Lastly, the dynamics of fish community diversity (i.e., species richness) can also be monitored by eDNA (Fig.5b). Direct visual census showed seasonal fluctuations in fish species richness (i.e., high in summer and low in winter), which were generally well reflected by eDNA analysis.

IV. CONCLUSION

In the present study, we showed that eDNA metabarcoding with the inclusion of internal standard DNA enables simultaneous determination of the quantity and identity of eDNA derived from multiple fish species. Because the traditional qPCR allows quantification of eDNA from only one fish species in a single experiment, our method is much more efficient compared with qPCR. In addition, our method can take effects of PCR inhibition into account. Taken together, our results show that eDNA metabarcoding with the inclusion of internal stan-

bioRxiv preprint doi: https://doi.org/10.1101/113472; this version posted March 5, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

dard DNAs can be an efficient tool to monitor fish biodiversity. Monitoring ecosystems and generating time series are very useful approaches for resource management, but obtaining multispecies, quantitative time series is often time- and effort-consuming. Our method will improve the efficiency of data generation, and would contribute to more effective resource management and ecosystem monitoring.

Ethics: The experiments were conducted in accordance with the guidelines of Regulation on Animal Experimentation at Kyoto University. Water sampling permission in or around the pier was not needed.

Authors' contributions: MU conceived and designed research; HM and RM performed sampling; MU, HM, RM, TS, MM, SS, HY and TM performed experiments; MU analyzed data; MU wrote the early draft and completed it with significant inputs from all authors.

Competing interests: We have no competing interests.

Acknowledgements: We would like to thank Aina Tanimoto for her help with the collection and filtration of water samples.

[1] R. Masuda, M. Shiba, Y. Yamashita, M. Ueno, Y. Kai, A. Nakanishi, M. Torikoshi and M. Tanaka, *Fishery Bulletin*, 2010, **108**, 162–173.

[2] H. Samejima, R. Ong, P. Lagan and K. Kitayama, *Forest Ecology and Management*, 2012, **270**, 248–256.

[3] R. P. Kelly, J. A. Port, K. M. Yamahara, R. G. Martone, N. Lowell, P. F. Thomsen, M. E. Mach, M. Bennett, E. Prahler, M. R. Caldwell and L. B. Crowder, *Science (New York, N.Y.)*, 2014, **344**, 1455–6.

[4] G. F. Ficetola, C. Miaud, F. Pompanon and P. Taberlet, *Biology letters*, 2008, **4**, 423–5.

[5] T. Takahara, T. Minamoto and H. Doi, *PloS one*, 2013, **8**, e56584.

[6] T. Minamoto, H. Yamanaka, T. Takahara, M. N. Honjo and Z. Kawabata, *Limnology*, 2011, **13**, 193–197.

[7] T. Takahara, T. Minamoto, H. Yamanaka, H. Doi and Z. Kawabata, *PloS one*, 2012, **7**, e35868.

[8] C. L. Jerde, A. R. Mahon, W. L. Chadderton and D. M. Lodge, *Conservation Letters*, 2011, **4**, 150–157.

[9] E. E. Sigsgaard, H. Carl, P. R. Møller and P. F. Thomsen, *Biological Conservation*, 2015, **183**, 46–52.

[10] S. Fukumoto, A. Ushimaru and T. Minamoto, *Journal of Applied Ecology*, 2015, **52**, 358–365.

[11] M. Ushio, H. Fukuda, T. Inoue, K. Makoto, O. Kishida, K. Sato, K. Murata, M. Nikaido, T. Sado, Y. Sato, M. Takeshita, W. Iwasaki, H. Yamanaka, M. Kondoh and M. Miya, *bioRxiv*, 2016, 068551.

[12] P. Taberlet, E. Coissac, F. Pompanon, C. Brochmann and E. Willerslev, *Molecular ecology*, 2012, **21**, 2045–50.

[13] R. P. Kelly, J. A. Port, K. M. Yamahara and L. B. Crowder, *PloS one*, 2014, **9**, e86175.

[14] M. Miya, Y. Sato, T. Fukunaga, T. Sado, J. Y. Poulsen, K. Sato, T. Minamoto, S. Yamamoto, H. Yamanaka, H. Araki, M. Kondoh and W. Iwasaki, *Royal Society open science*, 2015, **2**, 150088.

[15] S. Yamamoto, R. Masuda, Y. Sato, T. Sado, H. Araki, M. Kondoh, T. Minamoto and M. Miya, *Scientific Reports*, 2017, **7**, 40368.

[16] S. Yamamoto, K. Minami, K. Fukaya, K. Takahashi, H. Sawada, H. Murakami, S. Tsuji, H. Hashizume, S. Kubonaga, T. Horiuchi, M. Hongo, J. Nishida, Y. Okugawa, A. Fujiwara, M. Fukuda, S. Hidaka, K. W. Suzuki, M. Miya, H. Araki, H. Yamanaka, A. Maruyama, K. Miyashita, R. Masuda, T. Minamoto and M. Kondoh, *PLOS ONE*, 2016, **11**, e0149786.

[17] W. Smets, J. W. Leff, M. A. Bradford, R. L. McCulley, S. Lebeer and N. Fierer, *Soil Biology and Biochemistry*, 2016, **96**, 145–151.

[18] R. Masuda, Y. Yamashita and M. Matsuyama, *Fisheries Science*, 2008, **74**, 276–284.

[19] T. Magoč and S. L. Salzberg, *Bioinformatics (Oxford, England)*, 2011, **27**, 2957–63.

[20] R. C. Edgar, *Bioinformatics*, 2010, **26**, 2460–2461.

[21] C. Camacho, G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer and T. L. Madden, *BMC bioinformatics*, 2009, **10**, 421.

[22] R Core Team, *R: A Language and Environment for Statistical Computing*, 2016, `http://www.r-project.org/`.