Evaluating intraspecific diversity of a fish population using environmental DNA: An approach to distinguish true haplotypes from erroneous sequences

Satsuki Tsuji*[1], Masaki Miya[2], Masayuki Ushio[3,4], Hirotoshi Sato[5], Toshifumi Minamoto[6], Hiroki Yamanaka[7,8,9]

[1] Graduate School of Science and Technology, Ryukoku University, Shiga, Japan
[2] Department Ecology and Environmental Sciences, Natural History Museum and Institute, Chiba, Japan
[3] Center for Ecological Research, Kyoto University, Otsu, Japan
[4] PRESTO, Japan Science and Technology Agency, Kawaguchi, Japan
[5] Graduate School of Human and Environmental Studies, Kyoto University, Kyoto, Japan
[6] Graduate School of Human Development and Environment, Kobe University, Kobe, Japan
[7] Faculty of Science and Technology, Ryukoku University, Shiga, Japan
[8] The Research Center for Satoyama Studies, Ryukoku University, Shiga, Japan
[9] Ryukoku Center for Biodiversity Science, Ryukoku University, Shiga, Japan

*Corresponding author
E-mail: satsuki.may425@gmail.com, Tel&Fax: +81-(0)77-544-7113

Abstract

1. Recent advances in environmental DNA (eDNA) analysis using high-throughput sequencing (HTS) provide a non-invasive way to evaluate the intraspecific diversity of aquatic macro-organisms. However, many erroneous sequences, which has one or a few mismatches with the original sequence are included in HTS data, causing false positive genotypes. Therefore, reliable strategies are necessary to distinguish erroneous sequences from true haplotypes for evaluation of the intraspecific diversity.
2. In this study, we propose an approach to eliminate false positive haplotypes obtained from an eDNA sample. A mixture of rearing water including nine haplotypes of Ayu (*Plecoglossus altivelis altivelis*) mitochondrial D-loop region was used as an eDNA sample to prepare 15 replicates of sequencing libraries. Then, all library replicates were sequenced using HTS. Subsequently, amplicon sequence variant (ASV) methods were applied to process the sequences. The total number of detected true haplotypes and false positive haplotypes and the detection rates of these haplotypes among the 15 library replicates were compared with and without using the ASV method.
3. The use of the ASV method for the sequence analysis reduced the number of false positive haplotypes from 5,692 to 31, and it detected eight out of the nine true haplotypes. In addition, these true haplotypes always were detected in all 15 library replicates, regardless of whether the ASV method was used; however, the detection rates of respective false positive haplotypes varied from 1/15 to 15/15 replicates. Furthermore, combining the ASV method and removal of haplotypes with lower detection rates than 15/15 considerably reduced the number of false positive haplotypes from 31 to seven.
4. The approach proposed in this study eliminated most of erroneous haplotypes in the HTS data obtained from eDNA samples. Moreover, eliminating haplotypes with low detection rates led to further improvement of the detection accuracy. This approach for eDNA analysis will become an important option for evaluating intraspecific diversity because it does not require capturing or sampling of individual fishes. Moreover, this approach will be useful for future studies aimed at genetic resource management, ecosystem monitoring, phylogeographic analyses in broad area, and effective conservation of endangered species.

Key-words

environmental DNA, intraspecific diversity, mitochondrial haplotype, ASV methods, sequencing error

Introduction

Genetic diversity is a key component of biodiversity and is necessary for species' adaptation to changing natural and human-induced selective pressures (Allendorf et al., 2012; Laikre et al., 2016). Intraspecific diversity of a fish population has typically been analyzed by capturing individuals using traditional capture methods, such as baited traps, casting nets, electrofishing among other methods, followed by genetic analysis of each individual. However, these methods have at least two major limitations: 1) use of traditional sampling approaches causes damage to the target organisms, and 2) large sampling efforts are necessary to provide an accurate estimate of intraspecific diversity across an entire population. Traditional capture methods may threaten the persistence of species/population, particularly for rare and endangered species. In addition, insufficient sampling might lead to an underestimation of intraspecific diversity of a population (Xing et al., 2013). These limitations may reduce the feasibility of surveys and increase the uncertainty of results.

Environmental DNA (eDNA) is DNA released from organisms into the environment (e.g. soil, water, air), and it originates from various sources, such as metabolic waste, damaged tissues or sloughed skin cells (Kelly et al., 2014). and Environmental DNA analysis has recently been used to detect the distribution of macro-organisms, particularly those living in aquatic habitats (Ficetola et al., 2008; Lodge et al., 2012; Thomsen et al., 2015). Environmental DNA analysis allows for non-invasive and cost-effective detection of the presence of a species in a habitat because only water samples are needed to analyze instead of capturing and/or observing the target species (Thomsen et al., 2015). Because of this advantage and its high sensitivity, eDNA analysis has frequently been applied for the detection of not only common species but also rare and endangered species (Takahara et al., 2012; Rees et al., 2014; Fukumoto et al., 2015; Ishige et al., 2017; Katano et al., 2017). In addition, an approach involving eDNA metabarcoding using high-throughput sequencing (HTS) can effectively and comprehensively reveal the aquatic community structure, and thus, it has been gaining attention as a powerful tool for biodiversity monitoring (Thomsen et al., 2012; Kelly et al, 2014; Miya et al., 2015; Yamamoto et al., 2017).

Current applications of eDNA analysis have been limited mostly to the detection and identification of species (e.g. Rees et al., 2015; Thomsen et al., 2015). However, eDNA analysis potentially can be extended to evaluate of intraspecific diversity, because eDNA released from multiple individuals coexist in a water sample. Recently, Uchii et al. (2016, 2017) have developed a method using cycling probe technology and real-time PCR to quantify the relative proportion of two different genotypes of common carp (*Cyprinus carpio*) based on a single nucleotide polymorphism (SNP). These studies revealed that the SNP genotypes were 'embedded' in eDNA samples suspended in the field water. Furthermore, Sigsgaard et al. (2016) applied eDNA analysis for estimates of whale shark (*Rhincodon typus*) intraspecific diversity and found multiple haplotypes that had been identified previously from tissue-derived DNA by Sanger sequencing. These findings show the power and effectiveness of eDNA analysis for analyzing intraspecific diversity of target species. However, caution should be exercised during the use of HTS for intraspecific diversity, because HTS data usually include many erroneous sequences that are generated during PCR and sequencing (Schloss et al., 2011; Coissac et al., 2012; Edgar et al., 2016).

Researchers have tried to address the issue of erroneous sequences using multiple approaches, including the use of high-fidelity DNA polymerase in PCR, quality filtering based on base-call scores and/or clustering of sequences into operational taxonomic units (OTUs, OTU methods). The use of high-fidelity DNA polymerase in PCR contributes to decreased sequencing errors in PCR products (Ramachandran et al. 2011), but it is not completely prevented. The OTU methods involve clustering of sequences that are more different from each other than a fixed dissimilarity threshold (typically 3%; Callahan et al., 2016; Hughes et al., 2017). Thus, true haplotypes that are similar to each other are clustered into single OTU, leading to incorrect evaluations of intraspecific diversity. Therefore, OTU methods cannot be applied to analyze intraspecific diversity. To evaluate intraspecific diversity using eDNA samples, it is necessary to develop effective novel approaches to eliminate erroneous sequences inherent in HTS.

Intraspecific diversity of a population might be analyzed more effectively by use of amplicon sequence variant (ASV) methods, which have recently been developed in the fields of microbiology for correcting erroneous sequences derived from HTS data (e.g. Callahan et al., 2017 and references therein). ASV methods infer unique biological variants in the sample without imposing the arbitrary dissimilarity thresholds that define OTUs. As a core process of an ASV method, 'denoising' is performed using an error model that assumes the biological sequences are more likely to be observed than erroneous sequences (e.g., DADA2; Callahan et al. 2016). The sensitivity and accuracy of ASV methods with respect to correcting erroneous sequences have been shown to be better than those of OTU methods (Eren et al., 2013; Eren et al., 2015; Callahan et al., 2016; Edgar et al., 2016; Needham et the al., 2017). The high resolution of biological sequences afforded by ASV methods has the potential to improve the accuracy of evaluating intraspecific diversity inferred from eDNA.

The purpose of this study is to propose an approach for eliminating false positive haplotypes derived from erroneous sequences in HTS data obtained from an eDNA sample and demonstrate the usefulness of eDNA analysis for the evaluation of intraspecific diversity of a fish population. In this study, we examined genetic diversity in the Ayu (*Plecoglossus altivelis altivelis*) fish, an important fisheries target in Japanese inland waters whose genetic diversity has been evaluated in previous studies (e.g. Iguchi et al. 2002, Takeshima et al. 2016). First, we examined whether we could detect the same mitochondrial haplotype from the rearing water and of Ayu individual maintained in that corresponding tank. Second, we examined whether we could correctly detect variation in mitochondrial haplotypes from an eDNA sample containing multiple haplotypes derived from multiple individuals of Ayu. We prepared multiple library replicates for the eDNA sample and sequenced them separately. We compared the number of true haplotypes and false positive haplotypes between the results obtained with and without the use of ASV methods for processing the HTS data. During the analysis, we put special emphasis on the detection rate of each haplotype in library replicates because erroneous sequences are expected to occur randomly during experimental processes (e.g., PCR and MiSeq sequencing), and false positive haplotypes are expected to be detected rarely in multiple library replicates. Here we expected that false positive haplotypes could be eliminated correctly from HTS data of eDNA sample by using ASV methods and/or removing haplotypes with low detection rates among library replicates.

Materials and Methods

Ethics statement

The fish used for our experiments were purchased from a local fisherman. Based on current laws and guidelines of Japan relating to animal experiments of fish, the collection of fish tissue for extracting DNA and the use of DNA samples are allowed without any ethical approvals from any authorities. However, all experiments were performed by paying attention to animal welfare.

Primer design

We targeted the mitochondrial D-loop region, because it has a higher mutation rate compared with the nuclear DNA regions and the other mtDNA regions (Moritz et al., 1987). To amplify the control region of Ayu, two sets of species-specific primers were developed based on the complete mitochondrial DNA sequence of Ayu from the National Center for Biotechnology Information (NCBI; accession numbers of collected sequences were AB047553, EU12467−EU124683). The first primer set, Paa-Dlp-1 primer, was developed for Sanger-sequencing, which can amplify the nearly entire D-loop region (amplicon length, 541-bp). The sequences of the primers are as follows: PaaDlp-1_F (5′-GCTCCGGTTGCATATATGGACC-3′), PaaDlp-1_R (5′-AGGTCCAGTTCAACCTTCAGACA-3′)

The second primer set, PaaDlp-2, was designed for HTS by referring to instructions suggested previously (Miya et al., 2015; Palumbi, 1996). We obtained 232 sequences of the mtDNA control region from Ayu from the MitoFish v.2.80 (Iwasaki et al., 2013; http://mitofish.aori.u-tokyo.ac.jp/) and aligned these sequences. The information for all sequence data used to design primers for HTS (PaaDlp-2 primers) is listed in Table S1. The aligned sequences were imported into MESQUITE v. 2.75 (Maddison et al., 2010), and the search for a short hypervariable region (up to 200-bp for paired-end sequencing using the Illumina MiSeq) flanked by two

PaaDlp-2_F	5'-	C	C	G	G	T	T	G	C	A	T	A	T	A	T	G	G	A	C	C	T	A	T	T	A	C	-3'
Consensus	5'-	C	C	G	G	T	T	G	C	A	T	A	T	A	T	G	G	A	C	C	T	A	T	T	A	C	-3'
	A	0	0	0	0	0	0	0	0	232	0	232	0	232	0	0	0	231	0	0	0	232	0	0	232	0	
	T	0	0	0	0	231	232	0	0	0	232	0	232	0	232	0	0	0	0	0	232	0	232	232	0	0	
	G	0	0	232	232	1	0	232	0	0	0	0	0	0	0	232	232	0	0	0	0	0	0	0	0	0	
	C	232	232	0	0	0	0	0	232	0	0	0	0	0	0	0	0	1	232	232	0	0	0	0	0	232	

PaaDlp-2_R1/R2	3'-	G	A	A	C	G	C	A	A	T	G	G	T	C	T	G	A	T	**R**	T	T	A	T	C	G	-5'
Consensus	3'-	C	T	T	G	C	G	T	T	A	C	C	A	G	A	C	T	A	TorC	A	A	T	A	G	C	-5'
	A	0	0	0	0	0	0	0	0	232	0	0	226	0	209	0	0	232	0	203	232	0	232	0	0	
	T	0	232	232	0	3	0	232	232	0	0	1	0	0	0	0	232	0	223	0	0	232	0	0	1	
	G	0	0	0	232	0	232	0	0	0	0	0	6	232	23	0	0	0	0	29	0	0	0	232	0	
	C	232	0	0	0	229	0	0	0	0	232	231	0	0	0	232	0	0	9	0	0	0	0	0	231	

Fig. 1

The sequences of the PaaDlp-2 primers and sequence variation in the corresponding region from downloaded MitoFish v.2.80 data. The back-ground black colour indicates that the base does not bind despite the T/G bond. Note the presence of nucleotide substitutions only in one sequence out of 232, which was ignored during primer design.

conservative regions (ca 20–30 bp) was performed in the entire region of aligned sequences. For HTS, we designed the PaaDlp-2 primers on the selected positions within the amplification range of PaaDlp-1, considering the unconventional base pairing in the T/G bond to enhance the primer annealing (i.e. the designed primers use G rather than A when the template is variable C or T, and T rather than C when the template is A or G, Fig. 1). Two types of reverse primers, PaaDlp2_1R and PaaDlp2_2R, were designed, because the reverse priming sites has one variable site (the template is A or G) that does not bind despite the T/G bond. The base R indicates A (PaaDlp-2_R1) or G (PaaDlp-2_R2). The primer sequences are as follows: PaaDlp-2_F (5′-CCGGTTGCATATATGGACCTATTAC-3′), PaaDlp-2_R1 and PaaDlp-2_R2 (5′- GCTATTRTAGTCTGG-TAACGCAAG -3′). To check the specificity of the PaaDlp-1(F/R) and PaaDlp-2(F/R1/R2), we performed an *in silico* specificity screen using Primer-BLAST with default settings (http://www.ncbi.nlm.nih.gov/tools/primer-blast/).

Haplotype determination from tank water eDNA and corresponding individual

We obtained twenty juveniles of Ayu (0.92 ± 0.21 g wet weight, mean ± SD) that were caught by a large fixed net in Lake Biwa (35°18'25" N, 136°3'40" E) in Japan on 24 February 2015. Live fish were brought back to the laboratory and then maintained individually in a small tank with 300 mL of aged tap water at room temperature. After 15 min, each fish was removed from the tanks and anaesthetized with an overdose of clove oil. To extract DNA from the tissues, about 0.02 g of skeletal muscle tissues was collected from each individual. DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany) was used following the manufacturer's protocol.

To collect eDNA, 250 mL of rearing water was sampled from each tank and vacuum-filtered using a Whatman GF/F glass fiber filter (GE Healthcare Japan, Tokyo, Japan; diameter 47 mm; nominal pore size of 0.7 µm). All filter disks were folded in half inward with tweezers and wrapped in aluminum foil, then stored at −20°C. eDNA was extracted according to the methods described in the section, 'eDNA extraction from filters'. As a filtration negative control (FNC), the same volume of ultrapure water was filtered in the same manner after the filtration of the all real samples. The FNC was treated alongside the samples in the following experimental steps to confirm no cross contamination. Before use, all sampling and filtration equipment were exposed to a 10% blech solution for 10 min, washed with running tap water and rinsed with ultrapure water. PCR was performed in a 25-µL reaction for each sample using the StepOnePlus Real-Time PCR System. The mixture of the reaction was as follows: 900 nM each of PaaDlp-1(F/R) in 1 × PCR master mix (TaqMan gene Expression Master Mix, Life Technologies, Carlsbad, CA, USA) with 7 µL of sample eDNA or 1 µL of tissue-derived DNA (10 ng/µL). The PCR thermal conditions were 2 min at 50ºC, 10 min at 95ºC, 44 cycles of 15 s at 95ºC, and 60 s at 60ºC. The PCR products were purified using Nucleo Spin® Gel and PCR Clean-up Kits (Code

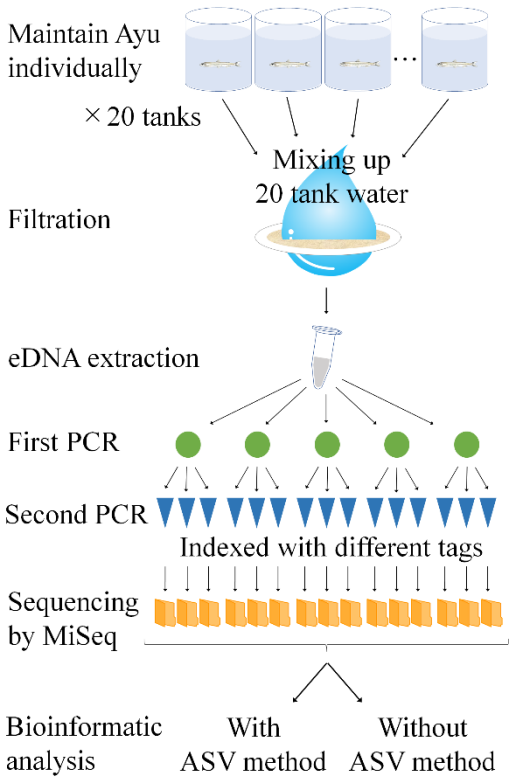

Fig. 2
Experimental design for detecting mitochondrial haplotype diversity from an eDNA sample.

No. 740609.50; TAKARA Bio, Kusatsu, Japan) according to the manufacture's instructions. Sequences were determined by commercial Sanger sequencing service (Takara Bio, Kusatsu, Japan). The sequences which were successfully determined (total 448 bp) were deposited in the DNA database of Japan (DDBJ, https://www.ddbj.nig.ac.jp/dra/index.html; accession numbers, LC406364- LC406383) and are listed in Table S2.

eDNA extraction from filters

The filter samples were subjected to eDNA extraction following the method described in Yamanaka et al. (2017). The filter was rolled into a cylindrical shape using sterile forceps and placed in the upper part of a spin column (EZ-10; Bio Basic, Markham, Ontario, Canada) which was removed the silica membrane before use. Excess water remaining in the filters was removed by centrifugation for 1 min at 6000 *g*, and a mixture of 200 µL of ultrapure water, 100 µL of Buffer AL and 10 µL of proteinase K was dispensed onto the filter in each spin column and incubated for 15 min at 56ºC. The Buffer AL and proteinase K were supplied from the DNeasy Blood & Tissue Kit. After incubation, the spin columns were centrifuged for 1 min at 6000 *g* to elute the eDNA into 2-mL tube. The upper part of the spin column was placed in a new 2-mL tube, and 200 µL of Tris-EDTA buffer (pH 8.0) was added to the filter and incubated for 1 min at room temperature to recover the remaining DNA on the filter. The spin columns were centrifuged for 1 min at 6000 *g* to obtain the second elution and mixed with the first elution. Subsequently, 100 µL Buffer AL and 600 µL ethanol were added to each tube and mixed by pipetting. The eDNA then was collected and purified from each solution using the DNeasy Blood & Tissue Kit following the manufacturer's protocol, with the minor modification that the final elution volume was adjusted to 100 µL of Buffer AE.

Detection of mitochondrial haplotype diversity from an eDNA sample

The experimental design is shown in Fig. 2. We collected 50 mL of rearing water from each tank used in the experiment described above. All collected water was mixed (total volume 1 L) and vacuum-filtered using a Whatman GF/F glass fiber filter. eDNA was extracted according to the methods described in above. After extracting eDNA from the filter sample, we employed a two-step tailed PCR approach to construct paired-end sequencing libraries, according to methods described by Miya et al. (2015). The FNC sample in the first experiment was used again as FNC sample in this experiment, because the filtrations of both experiments were performed at the same time. To avoid the risk of cross-contamination, all sampling and filtering equipment were decontaminated with 10% bleach solution for more than 10 min, carefully washed with tap water, and finally rinsed with ultrapure water. In addition, the PCR set-up was performed in a different room from PCR and HTS. The first PCR was performed in five replicates, each in a 12-µL reaction for a sample. The target region of Ayu was amplified using primers containing adapter sequences and random hexamers (N). The primer sequences are as follows: 5′-ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNN + PaaDlp-2_F (gene-specific sequences) -3′ and 5′-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNN + PaaDlp-2_R1/R2 (gene-specific sequences) -3′. The

mixture of the reaction was as follows: 0.3 µM of PaaDlp-2_F, 0.15 µM each of PaaDlp-2_R1 and R2 in 1 × KAPA HiFi HotStart ReadyMix (KAPA Biosystems, Wilmington, WA, USA) and a 2-µL sample of eDNA. To monitor cross contamination during library preparation, non-template control (NTC) were included in triplicate in the first PCR. The PCR thermal conditions were 3 min at 95ºC, 35 cycles of 20 s at 98ºC 15 s at 60ºC, and 15 s at 72ºC, followed by a final extension for 5 min at 72ºC. The first PCR products were purified twice using Agencourt AMPure XP beads (Beckman Coulter), according to the manufacturer's instructions (reaction ratio; AMPure beads 0.8: PCR product 1, target amplicon length; ca. 290 bp).

The second PCR was performed in three replicates, with each a 12-µL reaction of the first PCR (total 15 replicates per sample). To distinguish library replicates during Illumina MiSeq sequencing, respective library replicates (total 15 replicates) were indexed with different combinations of indexing primers. The primer sequences used in second PCR are listed in Table S3. The mixture of the reaction was as follows: 0.3 µM of each second PCR primer in 1 × KAPA HiFi HotStart ReadyMix and 2 µL of the purified first PCR product from the Agencourt AMPure XP beads. As negative controls in the second PCR, 2 µL of the first PCR product of the NTC was added to each reaction instead of template eDNA. The PCR thermal conditions were 3 min at 95ºC, 12 cycles of 20 s at 98ºC and 15 s at 65ºC, with a final extension for 5 min at 72ºC. The indexed second PCR products were pooled in equal volumes (5 µL each). The target size of the libraries (ca. 370 bp) was collected using 2% E-Gel® SizeSelect™ agarose gels (Thermo Fisher Scientific, Waltham, MA, USA) according to the manufacturer's instructions.

DNA concentrations in the collected libraries were estimated using the Qubit fluorometer (Life Technologies) with the Qubit dsDNA HS assay kit and adjusted to 4 nM (assuming 1 bp equals 660 g mol^{-1}) using ultrapure water. A 5-µL of the 4 nM library was denatured with 5 µL of fresh 0.2 N NaOH, followed by 5 µL of Tris HCl (200 mM, pH 7) and 985 µL of HT1 buffer (including in Miseq Regent Kit) was added to adjust the library concentration to 20 pM. Then, 48 µL of 20 pM PhiX DNA (Illumina, San Diego) and 360 µL of HT1 buffer were added to 192 µL of the 20 pM library to obtain a 8 pM library. The library was sequenced using the MiSeq platform (Illumina, San Diego), with the MiSeq v2 Regent Kit for the 2 × 150 bp PE cartridge (Illumina, San Diego). The sequencing reads obtained in the present study were deposited in the DDBJ Sequence Read Archive (accession number: DRA006638).

The MiSeq paired-end sequencing (2 × 150 bp) of the 21 libraries for this study (including 15 library replicates, three FNC and three NTC), together with 105 libraries from the other study (total number of libraries =126), yielded a total of 15.98 million reads, with 97.5% base calls containing Phred quality scores greater than or equal to 30.0 (Q30; error rate = 0.1% or base call accuracy = 99.9%).

Bioinformatic analysis

The full range of amplicons obtained using the PaaDlp-2 primers were successfully sequenced using the MiSeq platform. However, for the amplicons obtained using the PaaDlp-1 primers, some bases following the forward primer were undetermined by Sanger sequencing of the tissue-derived DNA from 20 individuals of Ayu. The forward primers of PaaDlp-2 and PaaDlp-1 were designed to be close to each other, and thus, three bases after the forward primer of PaaDlp-2 were needed to be omitted to compare the overlapping regions between the two datasets. Therefore, only 163 of the bases successfully determined for the two datasets were used for the subsequent bioinformatic analyses.

To perform a correction for erroneous sequences based on the ASV method, fastq files containing raw reads were processed using the Divisive Amplicon Denoising Algorithm 2 package ver. 1. 6. 0 (DADA2, Callahan et al., 2016) of R. The core algorithm of DADA2 infers unique biological variants using the denoising algorithm that is based on a model of errors in the amplicon sequencing with MiSeq. The detailed algorithm of DADA2 is described in the original paper. Briefly, the adopted error model in DADA2 quantifies the rate λji, at which an amplicon read with sequence i is produced from sample sequence j as a function of sequence composition and quality. Then, the p-value of the null hypothesis that the number of amplicon reads of sequence i is consistent with that of the error model was calculated using a Poisson model for the number of repeated observations of the sequence i, parameterized by the rate λji. Calculated p-values were used as a division criterion for an iterative partitioning algorithm, and sequence reads were divided until all partitions were consistent with being produced from their central sequence. Reads

of sequences inferred as error were replaced with the central sequence of the partition that included its sequence (i.e. error correction).

In this study, reads with one or more expected errors (maxEE = 1) were discarded during quality inspection and trimming of primer sequences. Quality-filtered sequences were dereplicated, and the parameters of the DADA2 error model were trained on a random subset of one million reads. The trained error model was used to identify and correct indel-mutations and substitutions. Denoised forward and reverse reads were merged and read pairs with one or more conflicting bases between the forward and reverse read were removed. DADA2 implements the function 'removeBimeraDenovo' to identify chimeras; however, it was not used in this study because haplotypes of Ayu included in sample water might be incorrectly identified as chimeras due to high sequence similarity. All detected sequences were confirmed to be 100% identical to the Ayu sequences determined in previous studies, using nucleotide BLAST (basic local alignment sequence tool, https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome).

Furthermore, the HTS data also was processed without the ASV method. The base calling errors were eliminated by quality filtering. The data pre-processing and dereplicating were performed using a custom pipeline described by Sato et al. (2018). Briefly, the low-quality tails were trimmed from each read, and the tail-trimmed paired-end reads (reads 1 and 2) were assembled using the software FLASH with a minimum overlap of 10-bp. The primer sequences were then removed with a maximum of three-base mismatches. Only when sequences had 100% identity with each other, they were operationally considered as identical. The fastq format was transformed into fasta, and the pre-processed reads were dereplicated. At this point, the reads were subjected to a local BLASTN search against a custom-made database of the control region of Ayu. The custom-made database was constructed from the 190 haplotypes of mitochondrial D-loop region of Ayu, which represented the sequences downloaded from MitoFish but excluded the 96 individuals (accession number LC406384- LC406403, Table S4) caught at Ado river (35°19'31" N, 136°3'48" E, Japan). The Ado river is connected to Lake Biwa and located close to a large fixed net (ca. 4 km) that was used to catch Ayu juveniles for the present study. The information in all haplotypes included in the custom-made database is listed in Table S4. If the respective sequences obtained in the HTS data had ≥ 99% similarity with the reference haplotype and an E-value < 10^{-5} in the BLAST results, the sequences were identified as those of Ayu.

Statistical analysis

All statistical analyses were performed using R ver. 3. 2. 3 software (R Core Team. 2016) and the minimum level of significance was set at $\alpha = 0.05$. To determine differences in the total number of reads derived from the nine true haplotypes, which were derived from 20 individuals (see "Results"), and false positive haplotypes, a Mann–Whitney *U* test was performed. A generalized linear model (GLM) with Poisson distribution was used to test how the detection rate of false haplotypes in the library replicates affected the total reads of the false haplotypes (glm function in R ver. 3.2.3 software). To visually determine the genetic distances among haplotypes that were detected from 15/15 library replicates (see "Results") and the relative read abundance of these haplotypes, haplotype network was generated using ape v5.1 and pegas v0.11 packages of R (Paradis et al. 2004).

Results

Testing species specificity of the two primer sets

The *in silico* specificity check for PaaDlp-1 and PaaDlp-2 (no adapter sequence) implemented in Primer-BLAST indicated species-specific amplification of Ayu. The direct sequencing of the PCR amplicons corroborated the amplification of the target region of Ayu in section 'Haplotype determination from tank water eDNA and corresponding individual'.

Comparison of detected haplotypes from tissue-derived DNA and corresponding tank eDNA

The sequences from the 20 Ayu individuals that were used for tank experiments were classified into 17 and nine haplotypes based on Sanger sequencing of PCR products amplified using PaaDlp-1 (amplicon length: 448 bp) and PaaDlp-2 (amplicon length: 163 bp), respectively (Fig. S1, Table S2). The detected haplotypes had only one or a few differences from each other, with the maximum pairwise *p*-distances for the two datasets

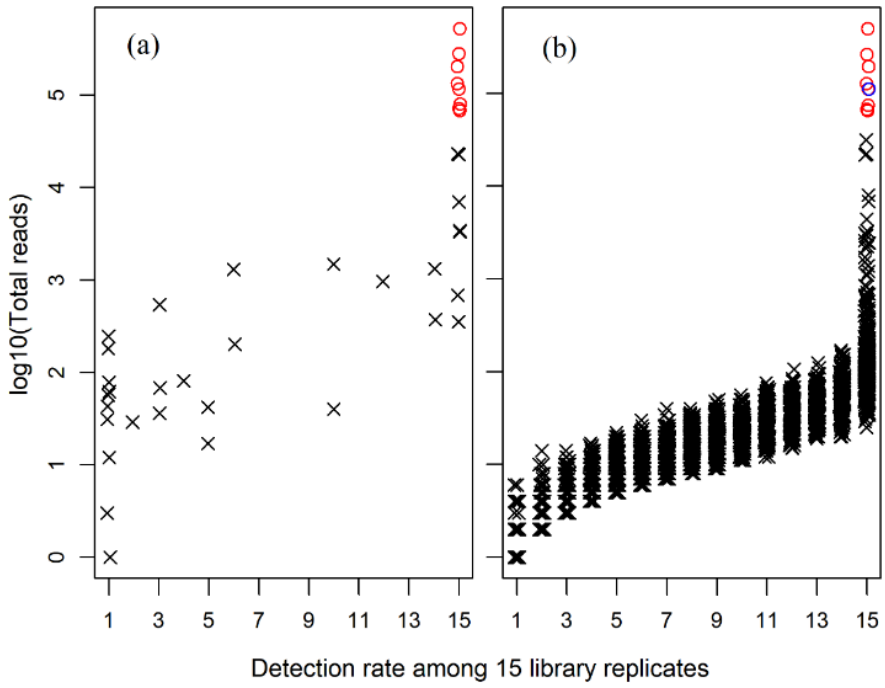

Fig. 3
Relationships between detection rate and total reads for each haplotype, which was detected (a) with and (b) without the use of the ASV method. The red circle and cross indicate the true haplotype and false positive haplotype, respectively. The blue circle indicates the true haplotypes that were detected without using the ASV method but were not detected with the ASV method (false negative haplotype; True haplotype ID I).

being 0.022 (PaaDlp-1) and 0.025 (PaaDlp-2), respectively. Each sequence obtained from eDNA, which was amplified using PaaDlp-1 and was detected from each of the 20 rearing water tanks, was identical to that obtained from tissue-derived DNA of the corresponding individual. In this tank experiment, the target fragments were not detected in any FNC and NTC. Thus, there was no evidence for cross-contamination during sample processing.

Detection of mitochondrial haplotype diversity from an eDNA sample

Based on the bioinformatic analysis using the ASV method, 1,539,351 reads were detected and assigned to 39 haplotypes (Fig. 3a, Table S5). Of these, 1,471,385 (96%) reads were assigned to eight true haplotypes and they were detected from 15/15 of library replicates. The remaining 67,966 (4%) reads consisted of 31 false positive haplotypes. A total of 386 (0.025%) and 235 (0.015%) reads were detected from the three FNC and three NTC (Table S5), respectively. The 31 false positive haplotypes were detected from 1/15 to 15/15 detection rates, but seven false positive haplotypes were detected all 15 library replicates. In addition, the false positive haplotypes with a low detection rate were randomly detected from the 15 library replicates on each filter and were not derived from any particular first PCR replication (Table S5).

Based on the bioinformatic analysis without the ASV method, 1,748,030 reads out of the total reads that passed quality control processes were assigned to Ayu with greater than or equal to 99% identity to the reference haplotypes in the custom-made database. Of these, 1,502,828 (86%) reads were assigned to nine true haplotypes, and they were detected from 15/15 library replicates (Fig. 3b). The remaining 245,202 (14%) reads consisted of 5,683 false positive haplotypes. The 5,683 false positive haplotypes were detected in 1/15 to 15/15 detection rates; however, 335 false positive haplotypes were detected in 15/15 library replicates (Fig. 3b). Despite the efforts to avoid the risk of cross-contamination, 124 (0.007%) and 105 (0.006%) reads were detected from the three FNC and three NTC, respectively.

Regardless of whether the ASV method was used, read abundances of the true haplotypes were significantly larger than those of the false positive haplotypes (Mann-Whitney *U* test; with ASV method, $p < 0.001$, $z = 4.27$; without ASV method, $p < 0.001$, $z = 5.21$; Fig. 3ab). Furthermore, the total reads of false positive haplotypes increased significantly by increasing the detection rate in library replicates (GLM; $p < 0.001$, $p < 0.001$; Fig. 3a and b).

Discussion

We found that correcting erroneous sequences with the ASV method was effective to improve the accuracy of intraspecific diversity evaluates with eDNA analysis. Furthermore, the accuracy of the analysis seems to be further improved by removing of haplotypes with low detection rates. Although some caution is still required for risk of false positives and false negatives, the proposed approach is useful for applying eDNA analysis to evaluation of intraspecific diversity that requires higher accuracy with respect to distinguishing true haplotypes from false ones.

The use of the ASV method in eDNA analysis for evaluating intraspecific diversity considerably decreased the number of false positive haplotypes (Fig. 3a and b). The great performance of the ASV method for eliminating false positive haplotypes is consistent with previous studies that identified microorganisms from mock community samples at fine taxonomical resolutions (Callahan et al., 2016; Kopylova et al., 2016; Hughes et al., 2017). The ASV method can correct a large proportion of erroneous sequences in HTS data, and thus, the combination of the ASV method and eDNA analysis has great potential to advance studies of intraspecific diversity of aquatic macro-organisms. In this study, we used the DADA2 package as an ASV method, but some other Illumina denoisers, which are based on different algorithms also have been published, including UNOISE (Edgar & Flyvbjerg, 2014), MED (Eren et al., 2015) and UNOISE2 (Edgar et al., 2016). The number and variety of false positive and false negative are likely to change depending on the method used, and the detection accuracy is expected to be improved by future development of ASV methods. Thus, novel ASV methods must be evaluated critically in future studies to increase the accuracy intraspecific diversity evaluates based on eDNA analysis.

There was one true haplotype (ID I), which was a false negative when the ASV method was used for bioinformatic analysis (Fig. 3a and b; Table S5). If there was variation within the priming site of the PaaDlp-2 primers, it could cause failure in PCR amplification (cf. Miya et al., 2015). However, haplotype ID I was a perfect match with the primer sequence of PaaDlp-2. In addition, when the HTS data was analyzed without the ASV method, it was detected from all 15 library replicates with a higher number of reads than false positive haplotypes. These results suggest that the DNA fragment of haplotype ID I was successfully amplified in PCR for library preparation. In other words, the sequence of true haplotype ID I was identified incorrectly as an error sequence and was judged as a false haplotype by ASV method even though the haplotype reads were present in all 15 library replicates. This false negative result is known to be caused by a failure to infer unique biological variants in the DADA algorithm (Rosen et al., 2012). In the DADA2 algorithm, haplotypes at low abundance (sequence reads) that have sequences similar to highly abundant (and true) haplotypes are more likely to be corrected and merged with the similar (true) haplotypes (cf. Callahan et al., 2016). The true haplotype ID I had one or two base pair differences with the other true haplotypes (ID B, E and F) and was mapped in close proximity to the abundant true haplotypes (Fig 4).

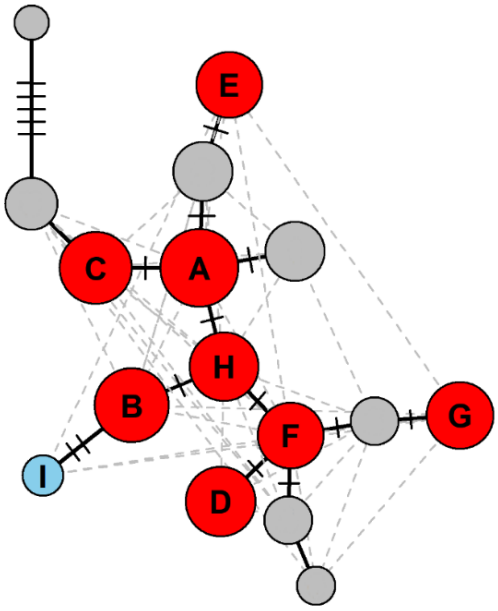

Fig. 4
A haplotype network based on the p-distance of the false negative haplotype and all haplo-types detected from 15/15 library replicates. Red, grey and blue circles indicate true haplotypes, false positive haplotypes and false negative haplotypes, respectively. The alphabet of red and blue circles indicates each haplotype ID. The circle size is proportional to the read abundance of each haplotype. Each of the tick marks corresponds to one base pair difference between the two connect-ed haplotypes. Dashed lines among haplotypes represent one or two base pair difference between the connected haplotypes.

In addition, the eight true haplotypes were detected with read abundances higher than the false positive haplotypes. Considering these results, the sequences of true haplotype ID I might have been identified as an error haplotype and corrected by DADA2.

The present results suggest that detection rates of each haplotype in library replicates provide an important clue to discriminate true haplotypes from false positive haplotypes. True haplotypes, especially predominant haplotypes, would be amplified at an early stage of PCR in all library replicates. However, unlike true haplotype, false positive haplotypes would not be contained in the initial eDNA template, but they are stochastically generated in some library replicates at a low rate (cf. Nakamura et al., 2011; Fukui et al., 2013). The false positive haplotypes with low detection rates in this study were incidentally generated in the first and second PCR step with no detection pattern. Furthermore, detection rates of true haplotypes in library replicates are expected to be much higher than those of false positive haplotypes. This expectation was strongly supported in this study (Fig. 3a and b). We found that the false positive haplotypes were successfully eliminated by combination of the use of the ASV method and the remove sequences that had detection rates less than 15/15 in library replicates. Therefore, the accuracy of evaluation of intraspecific diversity using eDNA analysis would be increased further by selecting haplotypes with high detection rates among multiple library replicates.

The present study also has implications for understanding the relationship between the total reads of each haplotype and the number of individuals owning that haplotype. Previous studies have suggested that the eDNA concentration increases with an increase in abundance and/or biomass of organisms (Takahara et al., 2012; Pilliod et al., 2013; Doi et al., 2016; Yamamoto et al., 2016). In addition, eDNA sequence reads were likely correlated with biomass and the number of individuals belonging to each fish family in the community (Thomsen et al., 2016). Thus, there is a possibility that the total number of reads of each true haplotype reflects the number of individuals that represent corresponding haplotypes (hereafter called 'owner individuals'). In this study, using the GLM analysis with the Poisson distribution, the number of owner individuals had a significant positive effect on the total reads of the haplotypes

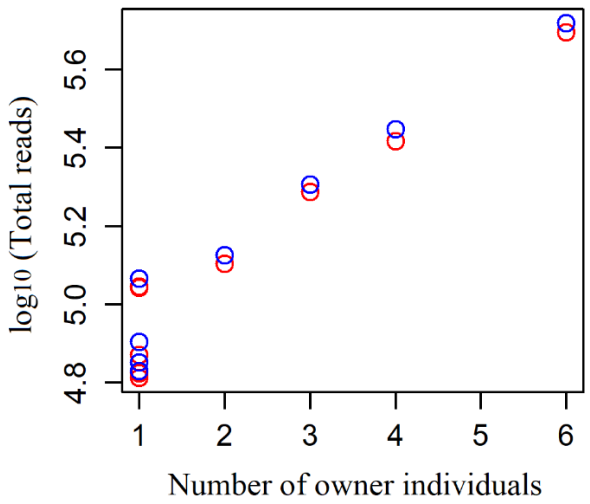

Fig. 5
Relationship between the number of owner individuals and total reads for each detected true haplotype. Blue and red circles indicate true haplotypes detected with and without the ASV meth-od, respectively. The number of owner individuals of each haplotype had a significant positive effect on the total reads with and without using the ASV method ($p < 0.01$, $p < 0.01$, respectively).

regardless of whether the ASV method was used ($p < 0.01$ in both cases; Fig. 5). This result suggests that the use of eDNA analysis has the potential to evaluate not only the diversity of haplotypes but also the relative dominance of each haplotype in a population. Furthermore, the total read abundance potentially can be used to eliminate false positive haplotypes based on an appropriate threshold value. However, we consciously avoided this approach. In a field setting, *a priori* information on the haplotype composition in the target population and the relative concentration of eDNA corresponding to each haplotype are usually unknown, and thus, the use of a higher threshold value for the total read abundance may lead to eliminate true haplotype and increase false negative rates.

In future studies, it will be necessary to address whether we can accurately detect true haplotypes derived from wild populations, because eDNA concentrations of field samples will be lower than those in tank experiments (e.g. Takahara et al., 2012; Minamoto et al., 2017). In addition, heterogeneous distributions of eDNA in field water has been also reported (Jerde et al., 2011; Dejean et al., 2012; Thomsen et al., 2012; Pilliod et al., 2013; Yamamoto et al., 2016). Therefore, it is necessary to

determine optimal sampling strategies, including appropriate volumes of sample water and distances between sampling points to accurately evaluate intraspecific diversity using eDNA analysis. Further development of eDNA analysis to evaluate intraspecific diversity would contribute to more effective genetic resource management and ecosystem monitoring.

Acknowledgements

We thank Mr. Sakurai, S. (Ryukoku University) for supporting the experiments. We thank Dr. Yamamoto, S. (Kyoto University), Dr. Tanabe, A.S. (Ryukoku University), Dr. Fukaya, K. (National Institute for Environmental Studies), and Doi, H. (Hyogo University) for constructive comments and suggestions to this study. This work was supported by JSPS KAKENHI Grant Number JP16K18610 and Grant-in-Aid for JSPS Research Fellow Grant Number JP18J10088.

Authors' contributions

S.T., T.M., and H.Y conceived the ideas and designed the study; S.T. and M.M. conducted the laboratory work and developed primer set (PaaDlp-1: S.T., PaaDlp-2: M.M.); S.T., M.U., and H.S. analyzed the data; and S.T. and H.Y. led the writing of the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

Data accessibility

The minimal raw dataset is uploaded to the DDBJ Sequence Read Archive (accession number: DRA006638).

References

Allendorf, F.W., G.H. Luikart, and S.N. Aitken. (2012) Conservation and the genetics of populations (2nd ed.). Hoboken, NJ: Wiley-Blackwell.

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, 581–583. doi: 10.1038/nmeth.3869

Callahan, B. J., McMurdie, P. J., Holmes, S. P. (2017) Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *The ISME Journal*, **11**, 2639–2643. doi: 10.1038/ismej.2017.119

Coissac, E., Riaz, T., Puillandre, N. (2012) Bioinformatic challenges for DNA metabarcoding of plants and animals. *Molecular Ecology*, 21, 1834–1847. doi: 10.1111/j.1365-294X.2012.05550.x

Dejean, T., Valentini, A., Miquel, C., Taberlet, P., Bellemain, E., Miaud, C. (2012) Improved detection of an alien invasive species through environmental DNA barcoding: the example of the American bullfrog *Lithobates catesbeianus*. Journal of Applied Ecology, 49, 953–959. doi: 10.1111/j.1365-2664.2012.02171.x

Doi, H., Inui, R., Akamatsu, Y., Kanno, K., Yamanaka, H., Takahara, T., Minamoto, T. (2016) Environmental DNA analysis for estimating the abundance and biomass of stream fish. *Freshwater Biology*, 62, 30–39. doi: 10.1111/fwb.12846

Edgar, R. C. (2016) UNOISE2: improved error-correction for Illumina 16S and ITS amplicon sequencing. *bioRxiv*. doi: 10.1101/081257

Edgar, R. C. and Flyvbjerg, H. (2014) Error filtering, pair assembly and error correction for next-generation sequencing reads. *Bioinformatics*, 31, 3476–3482. doi: 10.1093/bioinformatics/btv401

Eren, A. M., Maignien, L., Sul, W. J., Murphy, L. G., Grim, S. L., Morrison, H. G., …, Freckleton, R. (2013) Oligotyping: differentiating between closely related microbial taxa using 16S rRNA gene data. *Methods in Ecology and Evolution*, 4, 1111–1119. doi: 10.1111/2041-210X.12114

Eren, A. M., Morrison, H. G., Lescault, P. J., Reveillaud, J., Vineis, J. H., Sogin, M. L. (2015) Minimum entropy decomposition: unsupervised oligotyping for sensitive partitioning of high-throughput marker gene

sequences. *The ISME Journal*, **9**, 968–979. doi :10.1038/ismej.2014.195

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

Fukui, K., Bessho, Y., Shimada, A., Yokoyama, S., Kuramitsu, S. (2013) Thermostable Mismatch-Recognizing Protein MutS Suppresses Nonspecific Amplification during Polymerase Chain Reaction (PCR). *International Journal of Molecular Sciences*, 14, 6436–6453. doi: 10.3390/ijms14036436

Fukumoto, S., Ushimaru, A., Minamoto, T. (2015) A basin-scale application of environmental DNA assessment for rare endemic species and closely related exotic species in rivers: a case study of giant salamanders in Japan. *Journal of Applied Ecology*, 52, 358–365. doi: 10.1111/1365-2664.12392

Hughes, R., Alkan, Z., Keim, N. L., Kable, M. E. (2017) Impact of sequence variant detection and bacterial DNA extraction methods on the measurement of microbial community composition in human stool. *bioRxiv*. doi: 10.1101/212134

Iguchi, K., Ohkawa, T., Nishida, M. (2002) Genetic structure of land-locked ayu within the Biwa Lake system. Fisheries science, 68, 138–143. doi: 10.1046/j.1444-2906.2002.00399.x

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. doi: 10.1016/j.biocon.2017.04.023

Iwasaki, W., Fukunaga, T., Isagozawa, R., Yamada, K., Maeda, Y., Satoh, T. P., …, Nishida, M. (2013) MitoFish and MitoAnnotator: A Mitochondrial Genome Database of Fish with an Accurate and Automatic Annotation Pipeline. *Molecular Biology and Evolution*, 30, 2531–2540. doi: 10.1093/molbev/mst141

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, 150–157. doi: 10.1111/j.1755-263X.2010.00158.x

Katano, I., Harada, K., Doi, H., Souma, R., Minamoto, T. (2017) Environmental DNA method for estimating salamander distribution in headwater streams, and a comparison of water sampling methods. *Plos One*, 12, e0176541. doi: 10.1371/journal.pone.0176541

Kelly, R. P., Port, J. A., Yamahara, K. M., Crowder, L. B. (2014) Using Environmental DNA to Census Marine Fishes in a Large Mesocosm. *Plos One*, 9, e86175. doi: 10.1371/journal.pone.0086175

Kopylova, E., Navas-Molina, J. A., Mercier, C., Xu, Z. Z., Mahé, F., He, Y., …, Knight, R. (2016) Open-Source Sequence Clustering Methods Improve the State Of the Art. *mSystams*, 1, e00003-15. doi:10.1128/mSystems.00003-15

Laikre, L., Lundmark, C., Jansson, E., Wennerström, L., Edman, M., Sandström, A. (2016) Lack of recognition of genetic biodiversity: International policy and its implementation in Baltic Sea marine protected areas. *Ambio*, 45, 661–680. doi: 10.1007/s13280-016-0776-7

Lodge, D. M., Turner, C. R., Jerde, C. L., Barnes, M. A., Chadderton, L., Egan, S. P., Feder, J. L., Mahon, A. R. (2012) Conservation in a cup of water: estimating biodiversity and population abundance from environmental DNA. *Molecular Ecology*, 21, 2555–2558. doi: 10.1111/j.1365-294X.2012.05600.x

Maddison, W. P., Maddison, D. R. (2010) Mesquite: a modular system for evolutionary analysis, v. 2.75. See http://mesquiteproject.org.

Minamoto, T., Fukuda, M., Katsuhara, K. R., Fujiwara, A., Hidaka, S., Yamamoto, S., Takahashi, K., Masuda, R. (2017) Environmental DNA reflects spatial and temporal jellyfish distribution. *Plos One*, 12, e0173073. doi: 10.1371/journal.pone.0173073

Miya, M., Sato, Y., Fukunaga, T., Sado, T., Poulsen, J. Y., Sato, K., …, 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, 150088. doi: 10.1098/rsos.150088

Moritz, C., Dowling, T.E., Brown, W.M. (1987) Evolution of animal mitochondrial DNA: Relevance for population biology and systematics. *Annual Review of Ecology and Systematics*, 18, 269–292. doi:10.1146/annurev.es.18.110187.001413

Nakamura, K., Oshima, T., Morimoto, T., Ikeda, S., Yoshikawa, H., Shiwa, Y., Ishikawa, S., Linak, M. C., Hirai, A., Takahashi, H., Altaf-Ul-Amin, Md., Ogasawara, N., Kanaya, S. (2011) Sequence-specific error profile of Illumina sequencers. Nucleic Acids Research, 1–13. doi: 10.1093/nar/gkr344

Needham, D. M., Sachdeva, R., Fuhrman, J. A. (2017) Ecological dynamics and co-occurrence among marine phytoplankton, bacteria and myoviruses shows microdiversity matters. *The ISME Journal*, 11, 1614–1629. doi: 10.1038/ismej.2017.29

Palumbi S. (1996) Nucleic acids II: the polymerase chain reaction. In Molecular systematics (eds D Hillis, C Moritz, B Mable), pp. 205–247. Sunderland, MA: Sinauer Associates.

Paradis, E, Claude, J., Strimmer, K (2004). APE: analyses of phylogenetics and evolution in R language. Bioinformatics.

Pilliod, D. S., Goldberg, C. S., Arkle, R. S., Waits, L. P. (2013) Estimating occupancy and abundance of stream amphibians using environmental DNA from filtered water samples. *Canadian Journal of Fisheries and Aquatic Sciences*, 70, 1123–1130. doi: 10.1139/cjfas-2013-0047

R Core Team. R (2016) A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available from: https://www.R-project.org/.

Ramachandran, S., Zhai, X., Thai, H., Campo, D. S., Xia, G., GanovaRaeva, L. M., Drobeniuc, J., Khudyakov, Y. E. (2011) Evaluation of Intra-Host Variants of the Entire Hepatitis B Virus Genome. *Plos One*, 6, e25232. doi: 10.1371/journal.pone.0025232

Rees, H. C., Bishop, K., Middleditch, D. J., Patmore, J. R. M., Maddison, B. C., Gough, K. C. (2014) The application of eDNA for monitoring of the Great Crested Newt in the UK. *Ecology and Evolution*, 4, 4023–4032. doi: 10.1002/ece3.1272

Rees, H. C., Gough, K. C., Middleditch, D. J., Patmore, J. R. M., Maddison, B. C. (2015) Applications and limitations of measuring environmental DNA as indicators of the presence of aquatic animals. *Journal of Applied Ecology*, **52**, 827–831. doi: 10.1111/1365-2664.12467

Rosen, M., Callahan, B. J., Fisher, D. S., Holmes, S. P. (2012) Denoising PCR-amplified metagenome data. BMC Bioinformatics, 13, 283. doi: 10.1186/1471-2105-13-283t

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. doi: 10.1093/molbev/msy074

Schloss, P. D., Gevers, D., Westcott, S. L. (2011) Reducing the Effects of PCR Amplification and Sequencing Artifacts on 16S rRNA-Based Studies. *Plos One*, 6, e27310. doi: 10.1371/journal.pone.0027310

Sigsgaard, E. E., Nielsen, I. B., Bach, S. S., Lorenzen, E.D., Robinson, D. P., Knudsen, S. W., …, Thomsen, P. F. (2016) Population characteristics of a large whale shark aggregation inferred from seawater environmental DNA. *Nature Ecology & Evolution*, 1, 0004. doi: 10.1038/s41559-016-0004

Takahara, T., Minamoto, T., Yamanaka, H., Doi, H., Kawabata, Z. (2012) Estimation of fish biomass using environmental DNA. *PloS One*, 7, e35868. doi: 10.1371/journal.pone.0035868

Takeshima, H., Iguchi, K., Hashiguchi, Y., Nishida, M. (2016) Using dense locality sampling resolves the subtle genetic population structure of the dispersive fish species *Plecoglossus altivelis*. Molecular Ecology, 25, 3048–3064. doi: 10.1111/mec.13650

Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S. (2013) MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Molecular Biology and Evolution*, **30**, 2725–2729. doi: 10.1093/molbev/mst197

Thomsen, P. F., Kielgast, J., Iversen, L. L., Wiuf, C., Rasmussen, M., Gilbert, M. T. P., Orlando, L., Willerslev, E. (2012) Monitoring endangered freshwater biodiversity using environmental DNA. *Molecular Ecology*, 21, 2565–2573. doi: 10.1111/j.1365-294X.2011.05418.x

Thomsen, P. F., Møller, P. R., Sigsgaard, E. E., Knudsen, W., Jrgensen, O. A., Willerslev, E. (2016) Environmental DNA from Seawater Samples Correlate with Trawl Catches of Subarctic, Deepwater Fishes. *Plos One*, 11, e0165252. doi:10.1371/journal.pone.0165252

Thomsen, P. F., Willerslev, E. (2015) Environmental DNA – An emerging tool in conservation for monitoring past and present biodiversity. *Biological Conservation*, 183, 4–18. doi:10.1016/j.biocon.2014.11.019

Uchii, K., Doi, H., Minamoto, T. (2015) A novel environmental DNA approach to quantify the cryptic invasion of non‐native genotypes. *Molecular Ecology Resources*, **16**, 415–422. doi: 10.1111/1755-0998.12460

Uchii, K., Doi, H., Yamanaka, H., Minamoto, T. (2017) Distinct seasonal migration patterns of Japanese native and non-native genotypes of common carp estimated by environmental DNA. *Ecology and Evolution*, **7**, 8515–8522. doi: 10.1002/ece3.3346

Xing, X., Ma, X., Hart, M. M., Wang, A., Guo, S. (2013) Genetic Diversity and Evolution of Chinese Traditional Medicinal Fungus Polyporus umbellatus (Polyporales, Basidiomycota). *Plos One*, 8, e58807. doi:10.1371/journal.pone.0058807

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

Yamamoto, S., Minami, K., Fukaya, K., Takahashi, K., Sawada, H., Murakami, H., …, Kondoh, M. (2016) Environmental DNA as a 'Snapshot' of Fish Distribution: A Case Study of Japanese Jack Mackerel in Maizuru Bay, Sea of Japan. *Plos One*, 11, e0153291. doi: 10.1371/journal.pone.0153291

Yamanaka, H., Minamoto, T., Matsuda, J., Sakurai, S., Tsuji, S., Motozawa, H., …, Kondo, A. (2017) A simple method for preserving environmental DNA in water samples at ambient temperature by addition of cationic surfactant. *Limnology*, 18, 233–241. doi: 10.1007/s10201-016-0508-5