Investigation of a method for surveying aquatic insects using DNA barcoding of environmental DNA

Sakiko Yaegashi

University of Yamanashi, Department of Civil and Environmental Engineering,

Takeda 4-3-11, kofu, Yamanashi, 400-8511, Japan

Environmental DNA is attracting attention as a new species diversity monitoring method. In the case of eDNA metabarcoding for aquatic insects, which is an indicator of anthropogenic disturbance, non-target taxa (i.e., algae, bacteria) are often detected, which consumes the NGS sequence capacity. In this study, we investigated PCR conditions with high detection efficiency for aquatic insects. In the case of the eDNA metabarcoding to highly detect non-target taxa, it is helpful to detect rare taxa by reducing PCR cycles and by increasing PCR replication.

# Introduction

Biodiversity monitoring of macroinvertebrates, including aquatic insects, has been conducted to assess the effect of anthropogenic disturbance in the river environment. Because the macroinvertebrates have higher species diversity and lower dispersal potentials than freshwater fishes and tend to decline due to anthropogenic effects such as water pollution and the alternation of flow regime, its biodiversity has been employed as a bioindicator of those anthropogenic impacts. Although the general methods of macroinvertebrate diversity consist of field samplings and morphological identification, it is difficult to investigate its species diversity by the methods because of the following problems: 1) the limitation for field samplings (e.g., No. sampling sites, frequency), 2) Misidentification by complex identification keys, 3) the existence of a difficult-to-identify specimen (e.g., young or damaged one), 4) the hard effort to identify a large number of specimens (e.g., hundreds of specimens in a site).

In the last decade, environmental DNA (eDNA) has emerged as a powerful and easy tool to investigate species diversity in aquatic species. The eDNA exists in the environment (e.g., water, soil, air) as a cellar or extra-cellar DNA and has been released by original organisms [1], [2]. Because eDNA in river water is composed of DNA originating from various organisms inhabiting the river, the diversity of organisms can be investigated by eDNA analysis [3]-[6].

In particular, eDNA meta-barcoding is employed to investigate species diversity of particular taxa. In this method, organic matters, including eDNA were concentrated by filtration or other ways at first. Next, PCR is performed using universal primer sets to increase specific taxa [3], [7], and each PCR product reads DNA sequence by a high-throughput/next-generation sequencer (NGS) (e.g., Illumina Miseq). Then, the obtained DNA sequences are searched from DNA databases (e.g., the National Center for Biotechnology Information (NCBI), the Barcode of Life Data System (BOLD)) and estimated origin organisms of each eDNA based on the concordance rate between the obtained DNA sequences and reference sequences in DNA databases (DNA barcoding).

However, eDNA metabarcoding targeting for aquatic insects often detects many non-target taxa (e.g., algae, bacteria, fungi, fish), reducing the number of detectable species [5], [6]. The eDNA contain DNA originated from not only aquatic insects but also other species. Then, the universal PCR primer sets for aquatic insects can amplify DNA from non-target organisms, and DNA barcoding will also detect other organisms, unfortunately.

This study investigated PCR conditions (i.e., No. PCR cycles and No. PCR repetitions) for effective detection of aquatic-insect diversity. PCR method with High number cycles can result in detection of limited species which are more efficiently amplified DNA and is conducted to reduce detected the species diversity of aquatic insect fauna. It is known that the high number of PCR replications increases the detection efficiency of species diversity in a fish eDNA study [9]. However, a few studies confirm the case under conditions other taxa are detected.

# Methods

## Water Sampling and DNA extraction

The river water was collected on November 19, 2019, in the Arakawa river in Fujikawa Basin, Yamanashi, Japan (35°45'13.29"N, 138°35'54.42"E). This sampling site was located downstream of Arakawa Dam, and river water was expected to contain a large amount of organic matter from algae. 500 ml of river water was collected and filtered by a membrane filter (ADVANTEC, material Mixes Cellulose Ester Membranes, pore size = 0.2 µm) in our laboratory. The filter was preserved at -20°C.

DNA was extracted using the DNeasy Blood and Tissue kit (Qiagen). We used two times the volume of buffer ATL, AL, ethanol, and proteinase K solutions as much as the kit protocol. Subsequent operations follow the kit protocol. Then, extracted DNA was purified using OneStep PCR Inhibitor Removal Kits (Zymo Research).

# PCR and NGS Sequencing

The Cytochrome Oxidase I (COI) region of mitochondrial DNA (mtDNA) was amplified using BF2 and BR2 primers targeting aquatic insect DNA [7]. The three types of PCR (i.e., low number of PCR cycles: LC, a high number of PCR cycles: HC, and a high number of replications: HR) were conducted. The 50µl of PCR reaction solution is composed of 15.5μL of PCR Grade Water (Thermo Fisher Scientific), 4μL of 25mM dNTP (TaKaRa), 10μL of 5x Phusion HF buffer (New England BioLabs), 10µM forward and reverse primers with 1st PCR tails for Illumina sequencing system of 5μL, 10% Tween20 of 5μL for PCR enhancing, a total of 0.5μL Phusion high-fidelity DNA polymerase (New England BioLabs) and 5μL of 10-time diluted DNA were mixed. For the HR sample, the reaction solution was divided into five tubes of 10µl each. After PCR, all HR samples were mixed into one sample tube.

PCR steps were conducted by PCR Thermal Cycler Dice (TaKaRa). The PCR cycling conditions were 95°C for 3 minutes, PCR cycling steps (95°C for 30 seconds, 50°C for 30 seconds, and 72°C for 45 seconds), and 72°C for 5 minutes. The PCR products were preserved at 8°C. The number of PCR cycles was 40 cycles for the HC and 30 cycles for the LC and HR. The PCR products were purified using FastGene Gel/PCR extraction kit (Nippon Genetics) and sequenced from both ends using the MiSeq system (Illumina) in the Bioengineering lab (Kanagawa, Japan).

# Data analyses

The sequence data were analyzed following steps by the vsearch plugin in Qiime 2 [10]: 1) The primer sequences were trimmed from the obtained reads. 2) the sequences read from both ends were joined. 3) Short reads and reads with low sequence quality were removed. 4) After de-replication, operational taxonomic units (OTU) were clustered with 97% homology. 5) Chimeric reads that may have been joined during PCR were removed. The default settings were used for these Qiime 2 analyses.

DNA barcoding was performed using the Basic Local Alignment Search Tool (BLAST) [11]. Two DNA databases for this DNA barcoding were employed: 1) COI sequence data of aquatic insects in Fujikawa Basin (sequenced by Authors research group, unpublished) and 2) all Insecta DNA on mtDNA extracted from nt databases in DNA (National Center for Biotechnology Information, NCBI, accessed 2022. March, 14th). After the BLAST analysis, the following OTUs were removed: 1) with short alignment length (i.e., less than 300bp), 2) with a low matching rate (i.e., less than 90%). If more than 10 OTUs were classified in the same DNA database sequence, the BLAST was performed again using all sequences registered in NCBI. If OTUs were classified as taxa other than Insecta, these OUTs were removed. Finally, obtained taxa and the number of reads in each PCR condition were summarized, focusing on EPTDO taxa (i.e., Ephemeroptera, Plecoptera, Trichoptera, Diptera, Odonata) found in river environments commonly.

# Results and Discussions

A total of 111,992 reads (LC: 42,746 reads, HC: 31,444 reads, HR: 38,072 reads) and 12,912 OTUs (LC: 4,690 OTUs, HC: 6,355 OTUs, HR: 4,793 OTUs) were obtained at the OTU clustering analysis. The BLAST analysis classified that 941 reads (LC: 448 reads, HC: 168 reads, HR: 325 reads) and 110 OTUs (LC: 64 OTUs, HC: 30 OTUs, HR: 66 OTUs) were originated from Insecta animals. Unfortunately, the number of reads and OTUs assigned to Insecta were considerably less than the potential of metabarcoding analysis. Insecta taxa were found in 12 orders, 43 families, 59 genera, and 64 species. Of these, aquatic EPTDO taxa were five orders, 26 families, 42 genera, and 50 species. Detailed detection results are shown in Table 1. The highest number of taxa was detected in HR (45 taxa), and the lowest number of taxa was obtained in HC (13 taxa) in Figure 1.

Comparing the LC and the HC, the HC detected three times more species than the HC (Figure 1). In case of high PCR cycling conditions, dominant DNAs from algae or bacteria could be amplified efficiently under the HC PCR condition and could consume the sequencing capacity. As a result, rare DNA from aquatic insects was less likely to be detected. Comparing the LC and the HR, more species were detected by HR (Figure 1). In addition, the number of taxa for which only one read was obtained was greater for the HR than for the LC (LC: 13 taxa, HR: 23 taxa, Table 1). The efficiency of PCR amplification is considered to vary randomly without the effect of primer mismatches. When PCR replications were increased, the randomness of PCR amplification could be evened out, and thus more species could be detected in the HR.



# acknowledgement

This research was supported by the Japan Society for the Promotion of Science (JSPS) (grant numbers: 22K04374), Program to supporting research activities of female researchers, Environment Research and Technology Development Fund (grant numbers: 5-2006) and interdisciplinary research project in University of Yamanashi. Dr. Hidehiro Kaneko kindly supported this study.

REFERENCES

1. Taberlet, P., Coissac, E., Hajibabaei, M. and Rieseberg L.H., "Environmental DNA", *Molecular Ecology*, Vol. 21, (2012), pp 1789–1793.
2. Thomsen, P. F. and Willerslev, E., "Environmental DNA – An emerging tool in conservation for monitoring past and present biodiversity", Biological Conservation, Vol. 183, (2015), pp 4-18.
3. Miya, M., Sato, Y. Fukunaga, T., Sado, T., Poulsen, J.Y., Sato., K., Minamoto, T., Yamamoto, S., Yamanaka, H., Araki, H., Kondoh, M. and Iwasaki, W., "MiFish, a set of universal PCR primers for metabarcoding environmen-tal DNA from fishes: Detection of >230 subtropical marine species". Royal Society Open Science, Vol. 2, (2015), 150088.
4. Fukumoto, S., Ushimaru, A. and Minamoto, T., 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, Vol. 52, (2015), pp 358-365.
5. Deiner, K., Fronhofer, E. A., Mächler, E., Walser, J. and Altermatt, F., "Environmental DNA reveals that rivers are conveyer belts of biodiversity information". *Nature Communications*, Vol. 7, (2016), 12544.
6. Uchida, N., Kubota, K., Aita, S., Kazama, S., "Aquatic insect community structure revealed by eDNA metabarcoding derives indices for environmental assessment". *PeerJ*, Vol. 8, (2020), e9176.
7. Elbrecht, V. and Florian, L., "Validation and Development of COI Metabarcoding Primers for Freshwater Macroinvertebrate Bioassessment", *Frontiers in Environmental Science*, Vol. 5, (2017), 11.
8. Leese, F., Sander, M., Buchner, D., Elbrecht, V., Haase, P., Zizka, V.M.A., "Improved freshwater macroinvertebrate detection from environmental DNA through minimized nontarget amplification", *Environmental DNA* Vol. 3, (2021), pp 261– 276.
9. Doi, H., Fukaya, K., Oka, S., Sato, K., Kondoh, M., Miya, M., "Evaluation of detection probabilities at the water-filtering and initial PCR steps in environmental DNA metabarcoding using a multispecies site occupancy model", *Scientific Reports*, Vol. 9, 2019, pp 1-8.
10. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, et al., "Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2" *Nature Biotechnology*, (2019), Vol. 37, pp 852–857.
11. Edgar, R.C. "Search and clustering orders of magnitude faster than BLAST", *Bioinformatics,* Vol. 26, (2010), pp 2460-2461.