



OPEN

## Estimation of tuna population by the improved analytical pipeline of unique molecular identifier-assisted HaCeD-Seq (haplotype count from eDNA)

Kazutoshi Yoshitake<sup>1</sup>, Atushi Fujiwara<sup>2</sup>, Aiko Matsuura<sup>3</sup>, Masashi Sekino<sup>3</sup>, Motoshige Yasuie<sup>3</sup>, Yoji Nakamura<sup>3</sup>, Reiichiro Nakamichi<sup>3</sup>, Masaaki Kodama<sup>4</sup>, Yumiko Takahama<sup>4</sup>, Akinori Takasuka<sup>1</sup>, Shuichi Asakawa<sup>1</sup>, Kazuomi Nishikiori<sup>4</sup>, Takanori Kobayashi<sup>5,6</sup> & Shugo Watabe<sup>6</sup>✉

Many studies have investigated the ability to identify species from environmental DNA (eDNA). However, even when individual species are identified, the accurate estimation of their abundances by traditional eDNA analyses has been still difficult. We previously developed a novel analytical method called HaCeD-Seq (Haplotype Count from eDNA), which focuses on the mitochondrial D-loop sequence. The D-loop is a rapidly evolving sequence and has been used to estimate the abundance of eel species in breeding water. In the current study, we have further improved this method by applying unique molecular identifier (UMI) tags, which eliminate the PCR and sequencing errors and extend the detection range by an order of magnitude. Based on this improved HaCeD-Seq pipeline, we computed the abundance of Pacific bluefin tuna (*Thunnus orientalis*) in aquarium tanks at the Tokyo Sea Life Park (Kasai, Tokyo, Japan). This tuna species is commercially important but is at high risk of resource depletion. With the developed UMI tag method, 90 out of 96 haplotypes (94%) were successfully detected from Pacific bluefin tuna eDNA. By contrast, only 29 out of 96 haplotypes (30%) were detected when UMI tags were not used. Our findings indicate the potential for conducting non-invasive fish stock surveys by sampling eDNA.

### Abbreviations

HaCeD Haplotype count from eDNA

The accessibility of new molecular techniques, such as high-throughput sequencing, has enabled the assessment of biodiversity wherein samples are noninvasively acquired from various environments, including aquatic environments. These samples are cells released into the environment<sup>1-4</sup> and there are many studies for estimation of biomass by using DNA extracted from these cells, called environmental DNA (eDNA)<sup>5-7</sup>. However, the method using eDNA reported previously had some limitations. One of the problems when studying eDNA is that different quantities of cells are released into aqueous environments from different individuals even in the same species in the same environment due to variations in size and age-class distributions<sup>8</sup> as well as alteration in environmental factors such as water temperature and flow<sup>9</sup>. We have recently developed a novel method for the reliable and simple estimation of fish populations abundance from eDNA using the haplotype count<sup>10</sup>. This method, called

<sup>1</sup>Department of Aquatic Bioscience, Graduate School of Agricultural and Life Sciences, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-8657, Japan. <sup>2</sup>Fisheries Technology Institute, Japan Fisheries Research and Education Agency, 422-1 Nakatsuhamaura, Minami-ise, Mie 516-0193, Japan. <sup>3</sup>Fisheries Resources Institute, Japan Fisheries Research and Education Agency, 2-12-4 Fuku-ura, Kanazawa, Yokohama, Kanagawa 236-8648, Japan. <sup>4</sup>Tokyo Sea Life Park, 6-2-3 Rinkai-cho, Edogawa-ku, Tokyo 134-8587, Japan. <sup>5</sup>Japan Fisheries Research and Education Agency, 1-1-25 Shinurashima-cho, Kanagawa-ku, Yokohama, Kanagawa 221-8529, Japan. <sup>6</sup>School of Marine Biosciences, Kitasato University, 1-15-1 Kitasato, Minami-ku, Sagami-hara, Kanagawa 252-0373, Japan. ✉email: swatabe@kitasato-u.ac.jp

HaCeD (**H**aplotype **C**ount from **e**DNA)-Seq, was successfully adopted to noninvasively estimate the abundances of various eel species inhabiting the same experimental tank as well as the aquaculture pond at the same time<sup>10</sup>.

In HaCeD-Seq, the accuracy of base identification in the sequencing data is very important. Even a single base difference will be counted as a different haplotype in this method; thus, only reads with a base identification accuracy of 99.9% or higher were used in the previous report on eels<sup>10</sup>. However, even with 99.9% base identification accuracy, the probability of error occurrence at least once in every 300 bases (= the average length of a read) is  $(1 - 0.999^{300}) = 26\%$ . Therefore, even 99.9% accuracy is not high enough for this method. In addition, when samples contain several individuals, there are variations in the yield of eDNA derived from each individual as described above. It is generally difficult to discriminate between true minor haplotypes and false haplotypes derived from sequencing errors of the major haplotypes. In our previous report, due to sequencing errors, eels with less than 3% of those having the major haplotypes were indistinguishable from those with false haplotypes<sup>10</sup>.

Safe-SeqS is a sequencing technique that uses unique molecular identifiers (UMIs) to remove sequencing errors<sup>11</sup>. This technique assigns a random UMI tag to each DNA fragment, amplifies the fragment by PCR, sequences the amplified product and gathers the reads with the same UMI tag during data analysis, thus removing sequencing errors. Numerous studies have used UMI tags to remove PCR amplification bias, mainly in single-cell analysis<sup>12,13</sup>. Using UMI tags to remove PCR errors and sequencing errors has been adopted to analytical tools such as pRESTO<sup>14</sup>, MAGERI<sup>15</sup> and AmpUMI<sup>16</sup>, which facilitate the assembly of high-quality reads from sequenced data. Thus, it is interesting to explore whether UMI tags enable HaCeD-Seq to correctly detect the abundance of individuals, even when the abundance is low.

HaCeD-Seq was established to determine population sizes, and it is applicable to organisms with individual specific D-loop sequences, namely, those with high haplotype diversity. The haplotype diversities of the D-loop of tuna are reportedly as high as 0.9980 for longtail tuna<sup>17</sup>, 0.9987–1.000 for yellowfin tuna<sup>17,18</sup>, and 0.996 for Pacific bluefin tuna<sup>18</sup>. The Tokyo Sea Life Park in Japan planned to add naturally harvested Pacific bluefin tuna juveniles to this aquarium in 2016. This was a good opportunity to validate the improved UMI-assisted HaCeD-Seq method, because fin tissue samples could be taken for the determination of D-loop haplotype before they were transferred into the aquarium.

Tuna are top predators with unique features in terms of swimming ability<sup>19</sup>. These fish continuously swim at a high speed to extract oxygen with their gills and to chase prey. Tuna has long been an international market commodity for canning<sup>20</sup>. The fast muscle of tuna is also consumed raw in fish dishes (sashimi); formerly, this consumption was almost exclusively in Japan but is now common throughout the world. The consumption of tuna has greatly increased owing to the desirable taste of sashimi. This fish also has high contents of unsaturated n-3 fatty acids, such as eicosapentaenoic acid (EPA, 20:5 n-3) and docosahexaenoic acid (DHA, 22:6 n-3), which are thought to be beneficial for human health<sup>21,22</sup>.

Due to the increasing demand from consumption, the tuna stocks are considered to be at risk of overexploitation, although it previously had been thought that tuna harvesting would never cause their populations to become threatened<sup>20,23</sup>. In response to the risk of over fishing, the aquaculture programme of the Pacific bluefin tuna started in 1970 in Japan, resulting in predominantly farm-raised fish by 2002<sup>24,25</sup>. Despite this innovative achievement, most tuna aquaculture business still depends on wild stocks of juveniles because the survival rates of larval and juvenile tuna from artificial reproduction is not high<sup>26</sup>. Therefore, it is urgent to assess natural tuna stocks correctly and easily.

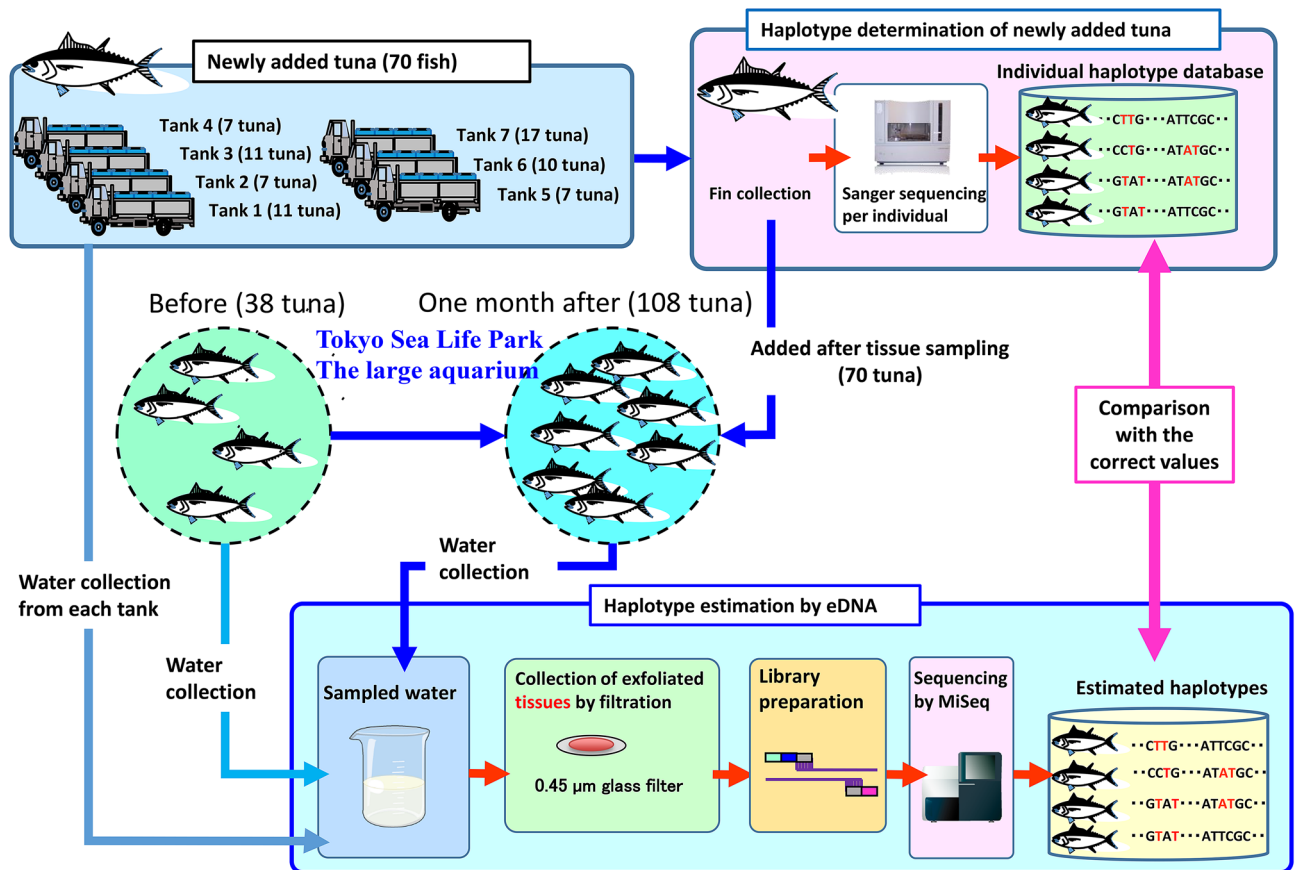
The objective of this study was to improve our HaCeD-Seq method by combining UMI assemblies and to examine the accuracy of this improved method to estimate the abundance of Pacific bluefin tuna that were reared in a large aquarium tank in the Tokyo Sea Life Park.

## Results

**UMI tagging to improve base accuracy.** eDNA was extracted from water samples taken from a total of 70 Pacific bluefin tuna transported to the Tokyo Sea Life Park with 7 live fish transport vehicles (Fig. 1). This was only one opportunity to obtain samples in our experiments.

Using the extracted eDNA as a template, PCR was performed with and without UMI-tagged primers, and the amplified products were sequenced by MiSeq at 300 bp × 2 paired-end sequencing (Fig. 2). The number of reads obtained is summarized in Tables 1 and 2. For sequencing data without the UMI tags, only high precision reads with all bases above Q30 (equal to 99.9% accuracy) were extracted, with an average of 32% of the reads remaining (Table 1). The average quality of bases in the obtained filtered reads was 37.9 (99.98% accuracy). The UMI-tagged reads were error-corrected by MAGERI, resulting in an average of 6% of the reads remaining (Table 2). The average quality of the bases was 39.8 (99.99% accuracy), but the real base accuracy was much higher since the upper limit of base quality output by MAGERI is up to 40.

**Estimation of tuna abundances in water tanks of live fish transport vehicles.** Each of the seven tanks in the transport vehicles contained 7–17 tuna specimens (Table 1). When the D-loop sequences of all 70 individuals were determined by Sanger sequencing, duplicated sequences were found for two sets, whereas one sequence was found for three individuals. Thus, in total, 64 haplotypes were found for the 70 fish in the tanks (Table 3, Supplementary Information Table S1). The sequences of 64 haplotypes determined by Sanger sequencing were identical to those determined by NGS with and without the UMI tags. We also performed HaCeD-Seq with and without the UMI tags on water samples from all the tanks to determine the optimal cluster size cut-off. Since false haplotypes result from sequencing errors, for the most abundant haplotypes, the changes in sensitivity and specificity depend on the cut-off for low-frequency haplotypes. The sequences and frequencies of haplotypes are shown in Supplementary Information Table S2, and we confirmed that the true haplotype sequences with the UMI tags were identical to those without the UMI tags. The data from each tank was normalized by the

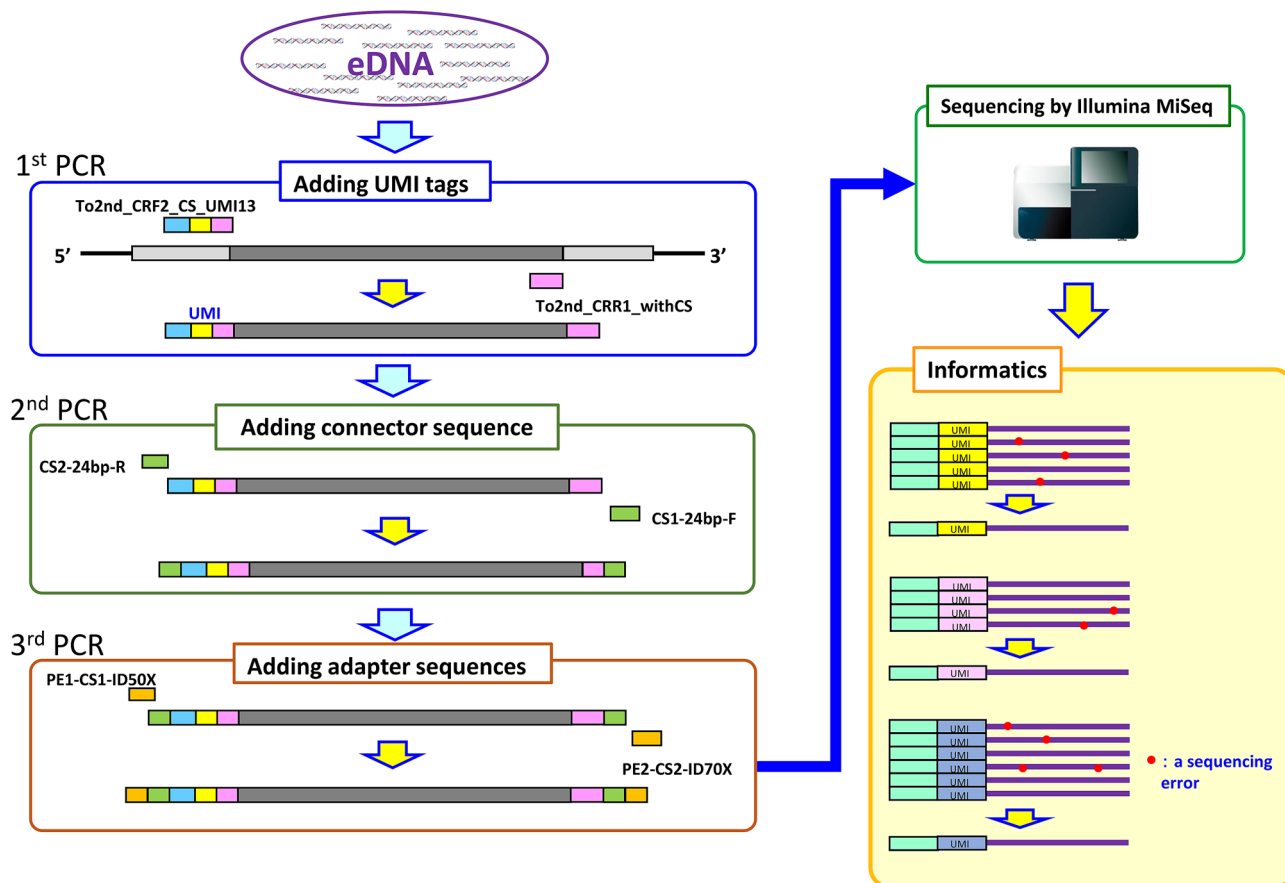


**Figure 1.** Experimental design of the study. Seventy Pacific bluefin tuna caught in the Pacific Ocean off Shikoku Island, Japan, were transported to the Tokyo Sea Life Park at Kasai, Tokyo. eDNA was extracted from water taken from the live fish vehicle tanks and subjected to HaCeD-Seq. The D-loop sequences of the 70 fish were determined by Sanger sequencing using DNA extracted from dorsal fin tissue samples that had been dissected with small forceps before the fish were added to the aquarium. One month after the 70 tuna were added to the large aquarium tank, water was sampled from the tank and the number of haplotypes was determined by HaCeD-Seq. The images of sequencers are from TogoTV (©2016 DBCLS TogoTV/CC-BY-4.0).

maximum haplotype cluster size and the sensitivities of haplotype detection are shown in Fig. 3A. Here, the sensitivity is the percentage of correct haplotypes among the total haplotypes observed. The sensitivity in HaCeD-Seq was not different between those with and without the UMI tags, although it tended to be slightly higher with the UMI tags. This indicates that the UMI tags remove the PCR bias caused by difference in the concentrations of different haplotypes that had been amplified by PCR.

The specificity results are shown in Fig. 3B. Here, the specificity is the percentage of correct haplotypes compared to all the detected haplotypes. The specificity varied markedly depending on whether the UMI tags were used. We determined the minimum percentage that would result in a specificity of at least 85% as the cut-off. Without the UMI tags, the specificity decreased to below the detection threshold of 2%. By contrast, with the UMI tags, the specificity remained almost unchanged until reaching 0.2%. We chose a threshold of 0.2% with the UMI tags and 1.5% without the UMI tags. The HaCeD-Seq results of the transport vehicle water tank samples using these thresholds are shown in Fig. 4. All of the true haplotypes (determined by Sanger sequencing) were detected both with and without the UMI tags, with exception of tank 6 where only 9/10 true haplotypes were detected without the UMI tags. Across all tanks, the absence of the UMI tags generated 8 false haplotype detections, compared to 4 with the UMI tags.

**Estimation of tuna abundance in the aquarium tank.** We next conducted the experiment in the aquarium and compared the results obtained from HaCeD-Seq with and without the UMI tags. A total of 38 tuna were already present in the aquarium tank, and then the 70 tuna were newly added. It was not feasible to determine the haplotypes of the existing 38 tuna by Sanger sequencing because it was difficult to collect tissue samples from live specimens in the large aquarium tank. Therefore, the water in the aquarium containing the 38 tuna was sampled, and 36 haplotypes were successfully obtained from an analysis with the UMI tags (Table 3 and “before” sequences in Supplementary Information Table S1). Among these 36 haplotypes, four were identical to those of fish in the transport vehicle tanks. Thus, together with the additional 70 tuna with 64 haplotypes, the total number of unique haplotypes of the total 108 tuna in the aquarium was 96 (Table 3, Supplementary Information Table S1).



**Figure 2.** Principle of eliminating PCR errors and sequencing errors using unique molecular identifier (UMI) tags. Assign one random UMI tag per molecule in the first PCR; amplify by including the UMI tag in the second PCR; sequence with MiSeq; collect reads with the same UMI tag; perform error correction by removing different bases. The image of a sequencer is from TogoTV (©2016 DBCLS TogoTV/CC-BY-4.0).

Tank	Individuals	Paired-end reads	> Q30 reads	Remaining reads (%)	Bases of > Q30 reads	Average base quality of > Q30 reads
<b>Transport vehicles</b>						
No. 1	11	3,84,457	1,20,193	31.3	4,38,66,841	37.94
No. 2	7	2,23,461	71,883	32.2	2,62,32,822	37.94
No. 3	11	4,47,696	1,50,634	33.6	5,49,41,606	37.95
No. 4	7	3,51,886	1,09,673	31.2	4,00,28,810	37.94
No. 5	7	2,53,536	83,057	32.8	3,03,15,443	37.95
No. 6	10	4,06,575	1,27,593	31.4	4,65,70,918	37.94
No. 7	17	5,48,745	1,76,627	32.2	6,44,67,585	37.94
<b>Aquarium</b>						
Before addition	38	14,47,500	4,79,641	33.1	17,50,66,441	37.95
After addition	108	42,96,541	13,87,557	32.3	50,64,41,781	37.95

**Table 1.** Statistics on sequence data without the UMI tags. The number of reads and the number and percentage of reads with all bases above Q30 are shown.

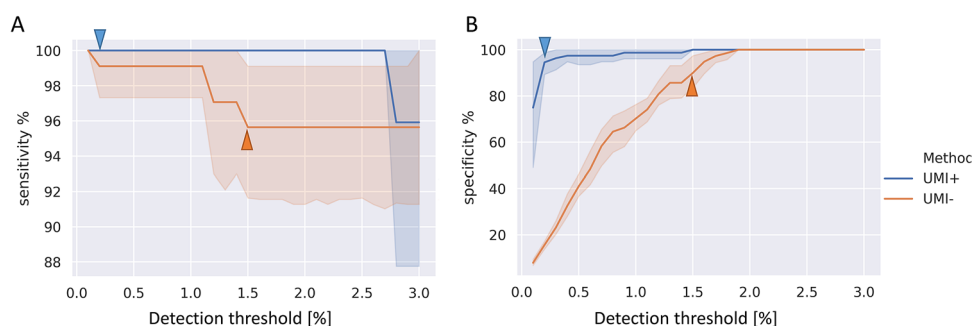
The HaCeD-Seq analysis of water samples from the aquarium tank detected 90 out of the 96 haplotypes (94%) when the UMI tags were used (Fig. 5). However, HaCeD-Seq analysis without the UMI tags detected only 29 haplotypes (30%), with two false haplotypes. These results show that when a large number of haplotypes need to be detected (i.e., greater than 100), the UMI tags are essential to obtain the correct number of haplotypes.

Tank	Paired-end reads	Assembled reads by MAGERI	Remaining reads (%)	Bases of assembled reads by MAGERI	Average base quality of assembled reads by MAGERI
<b>Transport vehicles</b>					
No. 1	3,19,040	17,048	5.3	69,34,013	39.86
No. 2	1,61,459	3132	1.9	12,74,727	39.85
No. 3	3,84,519	37,523	9.8	1,52,23,580	39.87
No. 4	1,85,995	234	0.1	95,584	39.50
No. 5	2,25,437	8074	3.6	32,85,460	39.86
No. 6	1,60,608	5416	3.4	22,04,473	39.86
No. 7	3,29,380	30,637	9.3	1,24,60,852	39.87
<b>Aquarium</b>					
Before addition	10,48,043	87,244	8.3	3,54,64,075	39.85
After addition	26,30,623	2,49,697	9.5	10,15,28,051	39.87

**Table 2.** Statistics on sequence data with the UMI tags. The number of reads and the number and percentage of reads assembled by MAGERI are shown.

Group	Individuals	Unique haplotypes	Haplotypes / Individuals (%)	Duplicated haplotypes (a/a)	Duplicated haplotypes (a/b)	Duplicated haplotypes (b/b)	Triplicated haplotypes (b/b)	Haplotype diversity
Previously present (a)	38	36*	95	2	4	–	–	0.991
Newly added (b)	70	64**	91	–	4	4	1	0.990
Total	108	96***	89	2	4	4	1	0.993

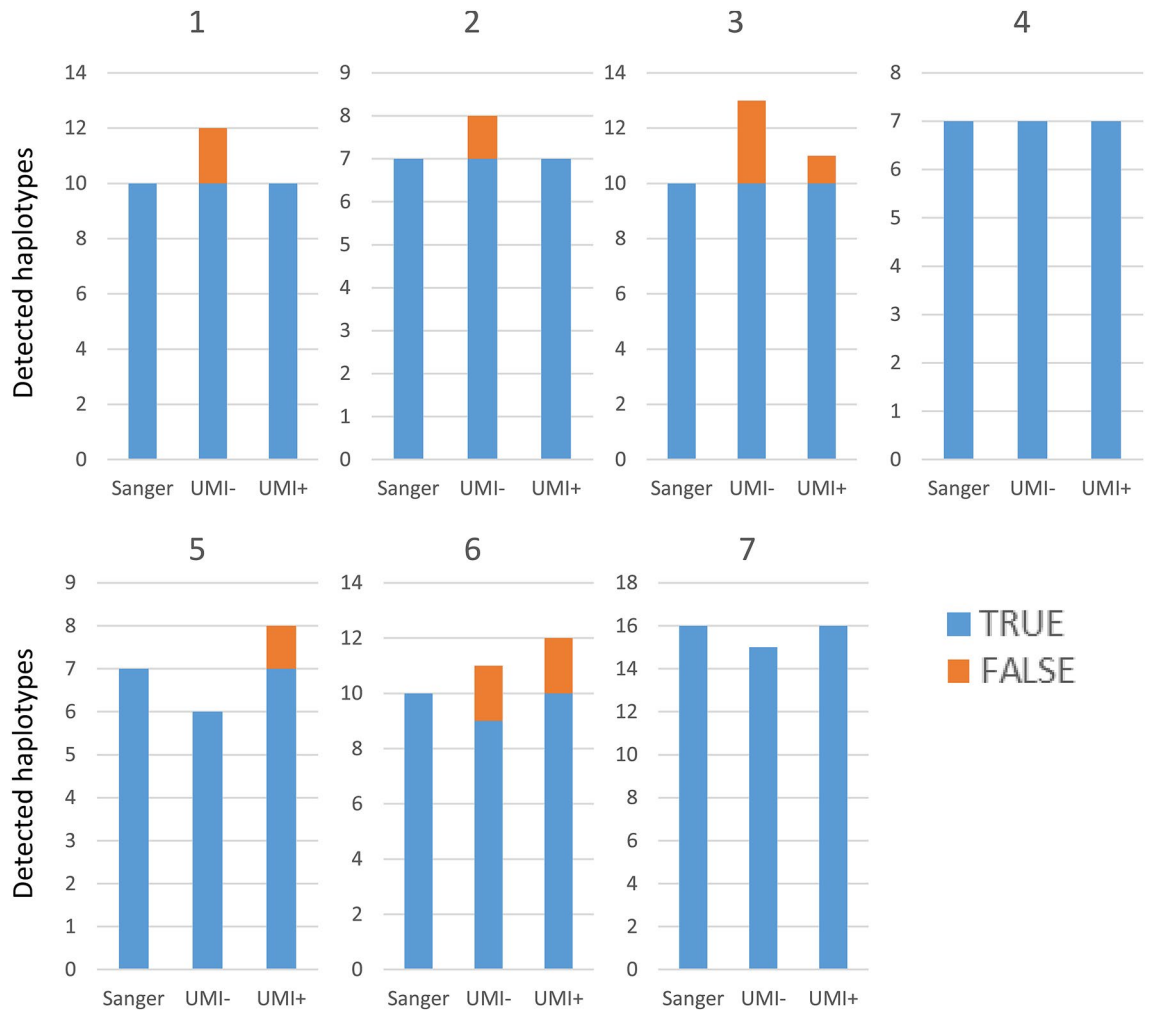
**Table 3.** The number of D-loop haplotypes determined based on UMI-assisted HaCeD-Seq. a/a, a/b and b/b indicate duplicated or triplicated haplotypes in the aquarium tank, the aquarium tank/vehicle tanks, and vehicle tanks, respectively. \*Determined by NGS with the UMI tags from eDNA samples. \*\*Determined by Sanger sequencing from tissue samples. \*\*\*Total number of haplotypes including previously present and newly added.



**Figure 3.** Differences in sensitivity and specificity with and without the UMI tags. (A) The relationship between the detection threshold and sensitivity changed depending on whether the UMI tags were used and depending on the detection threshold for the cluster size. Here, the sensitivity is the percentage of correct haplotypes among the total haplotypes observed. (B) The relationship between the detection threshold and specificity. Here, the specificity is the percentage of correct haplotypes compared to all the detected haplotypes. The horizontal axis shows the detection threshold, which is the size of the cluster when the most abundant haplotype is set to 100%. A threshold that was as low as possible without decreasing specificity was selected. The threshold finally adopted by HaCeD-Seq with the UMI tags (0.2%) is indicated by a blue triangle; the threshold adopted without the UMI tags (1.5%) is indicated by an orange triangle.

## Discussion

In this paper, we report the first application of a random sequence (unique molecular identifier: UMI) to eDNA analysis. In this method, the D-loop is PCR-amplified as the target region by adding the UMI tags to the end of this region with eDNA as a template. The D-loop and UMI sequences are then determined by next-generation sequencing, and identical DNA sequences with the same UMI sequence are counted as one unique sequence. The reads with different UMI tags are considered to be derived from different DNA copies; by counting the number of UMI tags, an accurate number of DNA copies can be obtained. Furthermore, PCR and sequencing errors



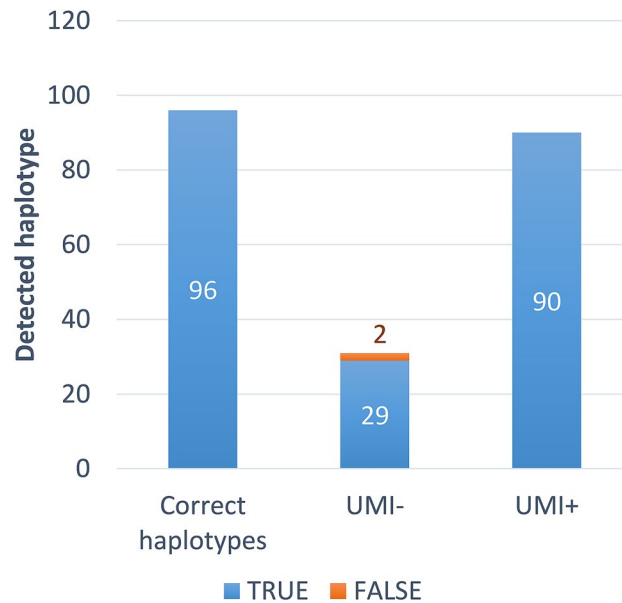
**Figure 4.** Number of detected haplotypes at optimal threshold. Number of haplotypes detected using a threshold of 0.2% with the UMI tags and 1.5% without the UMI tags; TRUE is the number of correct haplotypes detected; FALSE is the number of false haplotypes detected. Tank 1, 10 haplotypes for 11 individuals; Tank 2, 7 haplotypes for 7 individuals; Tank 3, 10 haplotypes for 11 individuals; Tank 4, 7 haplotypes for 7 individuals; Tank 5, 7 haplotypes for 7 individuals; Tank 6, 10 haplotypes for 10 individuals; Tank 7: 16 haplotypes for 17 individuals.

can be avoided by integrating sequences with the same UMI tag, and more accurate D-loop sequences can be obtained referring to the data previously reported<sup>14–16,27,28</sup>.

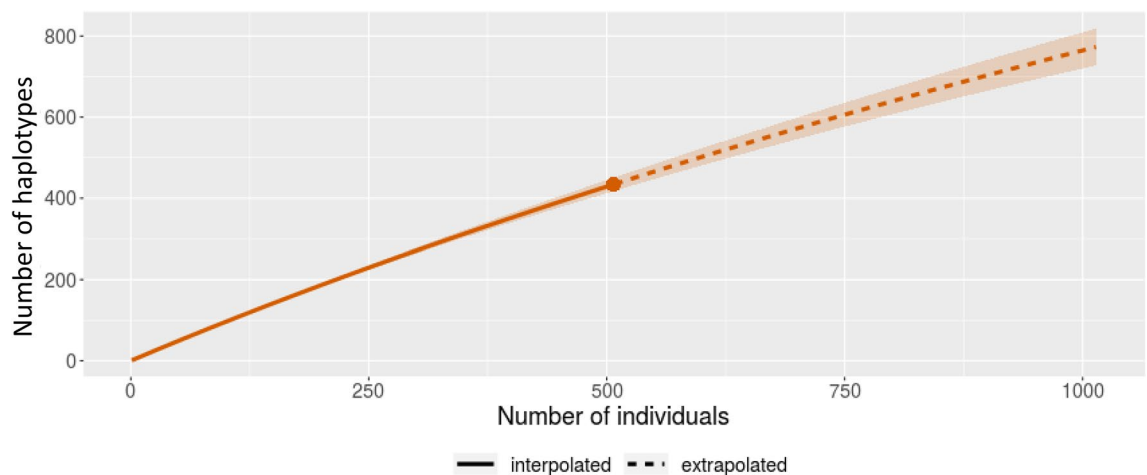
The accuracy of DNA polymerase in DNA synthesis is approximately 99.995%, including that of high-fidelity DNA polymerase<sup>29</sup>. However, as sequencing accuracies have increased over the years, PacBio's HiFi reads exceed 99.99% accuracy, so the effect of PCR errors during library construction has become non-negligible. PCR errors can be also avoided by integrating the results of multiple PCR runs operated in separate tubes. However, because there are cases in which PCR errors might occur at particular sites with similar frequencies even in separate reaction tubes with the same template DNA<sup>27,28</sup>, the removal of PCR errors by software has limitations. For greater accuracy, not only software processing, but also physical processing must be used in combination. We have successfully applied UMI tags to reduce the noise from PCR and next-generation sequencing errors by an order of magnitude when counting haplotypes; these UMI tags were very practical for eliminating PCR errors<sup>14–16</sup>.

By using HaCeD-Seq combined with the UMI tags to estimate the number of overlapping haplotypes in the Pacific bluefin tuna rearing in the aquarium tank at Tokyo Sea Life Park, we were able to detect 90 out of 96 haplotypes (94%) with no false haplotype. These results suggest that almost all the haplotypes present in the aquarium tank could be detected. On the other hand, 4 false haplotypes were detected among 71 haplotypes in 7 vehicle tanks with the UMI tags, suggesting the effect of different DNA concentrations in different sizes of tanks even on the number of false haplotypes for the UMI-assisted method. Meanwhile, only 29 out of 96 haplotypes (30%) could be detected with two false haplotypes without the UMI tags. Nevertheless, we were unable to detect 6 out of the 96 haplotypes (6%). Because 6 individuals died during the first month after addition to the aquarium tank, the six missing haplotypes may have resulted from these deaths. Therefore, it is likely that our haplotype detection rate was greater than 94%. The improved detection sensitivity when using the UMI tags is an important achievement to address the challenge of determining fish abundances from open seawater samples





**Figure 5.** HaCeD-Seq analysis results for the aquarium tank water. The correct number of haplotypes is 96 but only 90 haplotypes were detected when using the UMI tags (UMI+). Only 29 haplotypes and two false haplotypes were detected without the UMI tags (UMI-).



**Figure 6.** Relationship between the number of individuals and the number of unique haplotypes. Estimation using iNEXT from data on 507 Pacific bluefin tuna (DDBJ accession number: AB535754-AB536538).

because the variability in concentration among haplotypes is expected to be greater than tank experiments. HaCeD-Seq with the UMI tags is likely to be adopted in the studies of the species living in open seawater with high haplotype diversity.

Although the haplotype diversity of tuna is high<sup>17,18</sup>, only 96 haplotypes were detected in samples from the aquarium which housed 108 individuals. This discrepancy shows that there are differences between the number of haplotypes and the number of individuals. To estimate the abundance more accurately from the number of detected haplotypes, it is necessary to account for overlap in haplotypes among individuals and thus to make estimates based on a population genetic model as well as the effect of different DNA concentrations in open seawater. The iNEXT tool is a method that can produce such an estimate, and it is possible to estimate the error range of the resulting population numbers from the detected haplotype diversity<sup>30</sup>. In the case of Pacific bluefin tuna, the error range was estimated using the D-loop sequence of 507 bluefin tuna in a public database (Fig. 6). For a haplotype number of 96, the estimated number of individuals ranged from 98 to 101. These results will be an important guideline for setting research policies for field studies.

HaCeD-Seq makes it possible to estimate the number of individual fish in a tank from the eDNA contained in a water sample. Some fish hide behind rocks inside the aquarium, and even fish keepers are often unsure of the number of individuals in a given tank. Most fish in aquariums are wild-caught individuals, and the haplotype

diversity is expected to be high. Therefore, the method tested in this study can also be applied to monitoring the fish abundance in aquarium tanks as well.

The extent of exploitation of commercially important fish has been increasing due to increasing global demand. In particular, the global depletion of top predatory fish species such as the Pacific bluefin tuna is a serious concern in the context of sustainable fisheries management<sup>31,32</sup>. The estimation of the abundance and biomass (i.e., stock assessment) of fishery resources has heavily depended on fishery information, such as catch data from global stock assessment systems. A variety of resource estimation models have been devised to improve the accuracy of conventional resource estimation, which relies on catch data. The close-kin method, which uses kinship information, has been newly established for Pacific bluefin tuna and southern bluefin tuna<sup>33,34</sup>. This method can estimate the absolute abundance of parental fish based on the number of parent–offspring pairs determined through DNA marker-based parentage analysis. However, it is critical to collect specimens without bias in terms of distribution, migration, spawning time and location. Samples for the close-kin method are collected by fishery-independent sampling surveys or from marketplaces. Therefore, the reliability of this method depends on the accuracy of catch information. By contrast, the method proposed in the current study does not require fish sampling because it uses the eDNA contained in environmental water samples to estimate the abundance of a target species. In addition, as the conventional method of stock assessment relies on fisheries, many individuals are to be missed. High haplotype diversity has already been reported for some commercially important small pelagic fish such as jack mackerel (*Trachurus japonicus*)<sup>35</sup>, chub mackerel (*Scomber japonicus*)<sup>36</sup>, and spotted mackerel (*Scomber australasicus*)<sup>37</sup> for which a total allowable catch (TAC) has been set in the fishery management system of Japan<sup>38</sup>. The haplotype diversities of jack mackerel, chub mackerel and spotted mackerel have been reported to be 0.964–1.000, 0.505–0.967 and 0.996, respectively. Hence, the proposed method could be applicable for stock assessments of these major target species with high haplotype diversity in fisheries. It is interesting to investigate how high haplotype diversity is needed to apply this approach. A combination of the proposed method and the conventional methods is expected to improve the accuracy of estimating the abundance and biomass of aquatic organisms. Furthermore, this method would also potentially be applicable to any commercial or non-commercial species if the estimation process is first established based on haplotype diversity. From such a viewpoint, this method is anticipated to provide a possible venue for multispecies approach under the conceptual framework of ecosystem-based fishery management<sup>39</sup>, which is a new direction for the management of fisheries that places priority on the ecosystem rather than on a specific target species.

There are many studies showing the correlation between the abundance of eDNA and the biomass<sup>5–7</sup>. However, abundance estimation in the natural environment has been difficult, because the concentrations of eDNA are variable by variations in size and age-class distributions even in the same species<sup>8</sup> as well as alteration in environmental factors such as water temperature and flow<sup>9,40,41</sup>. This has been an open question linking eDNA to species abundance. Yamamoto et al.<sup>42</sup> used eDNA with quantitative PCR in combination with echo sounder technology to estimate the distribution and biomass of jack mackerel. Our method allows the fish abundance and biomass to be estimated from environmental water samples without any further supplementary method. Such methodological independence is a strong advantage of our approach, which establishes a simple and accurate estimation of fish abundance. Although different numbers of the sequences were distributed across different haplotypes with the UMI tagged data, we can detect almost 100% of the haplotypes. This data would allow us to see how variable eDNA production rates amongst individuals of the same species, which we think would be a very useful addition to our knowledge.

One of the most suitable applications of the present method is to estimate the number of individuals of endangered species. Endangered species are rapidly declining in abundance and biomass, making it more difficult to sample individuals in the field. A direct abundance estimation based on our method using only environmental water samples could be a desirable way to assess the population status of endangered species without disturbing the ecosystem through field sampling. Our proposed method provides a route for the simultaneous pursuit of sustainable management, ecosystem conservation, and necessary assessment.

## Materials and methods

**Specimens.** Pacific bluefin tuna were caught in the Pacific Ocean off Kashiwa Island, Otsuki, Kochi Prefecture; off Kannoura, Toyo, Kochi Prefecture; and off Nachikatsuura, Wakayama Prefecture. All specimens were temporarily kept in cages off Kashiwa Island and then transported by a live fish boat to Misaki, Kanagawa Prefecture. From this location, four live fish transport vehicles with a combined total of seven tanks were used to transport 70 tuna to the Tokyo Sea Life Park on June 12, 2016. Each transport tank (weighing 7–17.1 tons) contained 7–17 tuna (7, 7, 7, 10, 11, 11 and 17 tuna) (Fig. 1). While these transport tanks had no circulation system, aeration with pure oxygen and air were provided. No fish deaths occurred during transportation from Misaki to the park or during transfer from the vehicle tanks to the aquarium. A 10 L sample of environmental water was collected from each tank after transportation. A total of 38 Pacific bluefin tuna were currently housed in a 2,200-ton aquarium tank at the park, and a 10 L sample of the surface water from this tank was collected on June 8, 2016, before adding the 70 new Pacific bluefin tuna. The water temperature of the aquarium tank was 23 °C, and the circulation volume was 2,200,000 L/h. About 60% of the water in the aquarium tank was replaced per month. The water of fresh natural seawater was taken off Hachijojima in Izu Islands and transported to the aquarium. Aeration with pure oxygen and air were provided, where dissolved oxygen concentration was maintained at 6.3–6.7 mg/L. Pieces of dorsal fin tissues of the tuna were dissected with small forceps from the 70 individuals to determine the D-loop sequences by Sanger sequencing, and pit tags were embedded for individual identification. One month after the addition of the 70 tuna, 10 L of surface water was again collected from the aquarium tank on July 6, 2016. The dates of the individuals that died in the aquarium tank between June 8 and July 6, 2016, are as follows. Two individuals previously present in the aquarium tank died on June 13 and June



27, 2016. Four individuals newly added died in the aquarium tank each on June 13, June 14, June 14, and June 16, 2016. All animal handling methods and experimental protocols were carried out in accordance with the Guidelines for Animal Experimentation at Fisheries Research Agency. This study was approved by the subcommittee on institutional animal care and use of Graduate School of Agricultural and Life Sciences, The University of Tokyo (permission # P14-952).

**DNA extraction and sanger sequencing for reference individuals.** For each of the aforementioned 70 individuals (reference individuals), the total DNA to be used for PCR amplification of the whole D-loop was extracted from the excised fin tissue using the DNeasy Blood & Tissue Kit (Qiagen). The DNA concentration was quantified with a Qubit dsDNA BR Assay Kit (Invitrogen, Thermo Fisher Scientific).

The PCR primer design in Primer<sup>3</sup><sup>43</sup> was based on a complete mitogenome sequence of Pacific bluefin tuna obtained from the NCBI nucleotide database (accession: AB185022). In the mitogenome sequence, the forward primer (ToCR3\_F: 5'-CAGGCTGAGCTGAGAACAAA-3') spans from the 17th to 36th upstream site of the 3'-end of the cytochrome *b* gene, whereas the reverse primer (ToCR1\_R: 5'-GGGCCATCTTAACATCTTCA-3') is positioned within the tRNA<sup>Phe</sup> gene. Hence, the primer set is capable of amplifying the whole tRNA<sup>Thr</sup>, tRNA<sup>Pro</sup>, and D-loop (expected amplicon size of 1,092 bases, including 865 in the D-loop in the mitogenome sequence).

PrimeSTAR GXL polymerase (Takara Bio) was used in the PCR reaction. The PCR reaction solution (25 µL total volume) contained 5.0 µL of 5× buffer provided by the supplier, 2.0 µL of dNTP mix (2.5 mM each), 1.0 µL each of the ToCR3\_F and ToCR1\_R primers (10 µM), 0.5 µL of polymerase (1.25 units/µL), 13.5 µL of ddH<sub>2</sub>O, and 2.0 µL of DNA solution (10 ng/µL). Thermal cycling in a ProFlex PCR System (Applied Biosystems, Thermo Fisher Scientific) consisted of 35 cycles of 98 °C for 10 s, 60 °C for 15 s, and 68 °C for 1 min. The amplified products were enzymatically purified with exonuclease I (New England Biolabs) and thermo-sensitive alkaline phosphatase (Promega), as previously described<sup>44</sup>. The PCR primers were used to sequence the amplicons from both directions using a BigDye Terminator v3.1 Cycle Sequencing Kit in combination with a 3730xl DNA Analyzer (Applied Biosystems, Thermo Fisher Scientific). The resulting sequences were aligned and edited in DNASIS Pro version 2.02 (Hitachi Software Engineering). Multiple sequence alignment was conducted using the ClustalW algorithm implemented in MEGA6<sup>45</sup>. After multiple sequence alignment, a stretch of 365 bases of the D-loop sequence determined for each individual served as a reference sequence in downstream eDNA sequencing.

**Extraction of environmental DNA from water.** A 10 L sample of collected environmental water was passed through a 47 mm GF/D (2.7 µm pore size) glass-fibre filter (Whatman). After filtering the samples, the filters were fixed by adding 15 mL of ethanol<sup>46</sup>. The filter was divided into two equal parts with scissors before DNA extraction; one of the two pieces was used for DNA extraction in order to properly fit the extraction kit. DNA extraction from the filter was performed using a DNeasy Blood & Tissue Kit<sup>47</sup>. The extracted eDNA was eluted in 400 µL (twice with 200 µL) of buffer AE supplied with the kit. The extracted eDNA was quantified using a Qubit 2.0 Fluorometer (Thermo Fisher Scientific) and subjected to PCR under the same conditions as used for Sanger sequencing to confirm PCR amplification.

**UMI-tagged PCR and MiSeq sequencing.** To construct the UMI-tagged libraries, another primer set (degenerate primers) was designed within the D-loop region: To2nd\_CRF2 (5'-CTAGTACYAACCATTCA TA-3') and To2nd\_CRR1 (TGACCCYCTAGARAGAACG-3'). The extracted eDNA was adjusted for UMI-tagged libraries according to the following procedures. The construction of the UMI-tagged libraries was performed in a three-step PCR protocol (Supplementary Information Fig. S1).

For the first PCR, 10 ng of DNA was used as a template, and To2nd\_CRF2\_CS\_UMI13: 5'-CTCTTCCGA TCTGTCTNNNNNNNNNNNNNNCTAGTACYAACCATTTCATA-3' and To2nd\_CRR1\_withCS: 5'-CTCTTTC CGATCTCAGTGACCCYCTAGARAGAACG-3' were used as primers. The composition of the PCR reaction medium (25 µL total) was 0.5 µL of PrimeSTAR GXL DNA Polymerase, 5.0 µL of 5× Prime STAR GXL Buffer, 2.0 µL of dNTP mixture (2.5 mM each), 1.0 µL of primer To2nd\_CRF2\_CS\_UMI13 (10 µM), 1.0 µL of primer To2nd\_CRR1\_withCS (10 µM), 10 ng of sample, and adjusted to 15.5 µL with distilled water. Four cycles of 10 s at 98 °C, 30 s at 55 °C and 1 min at 68 °C were carried out in a thermal cycler, and the reaction solution was then stored at 4 °C. Next, 0.1 µL of exonuclease I (Takara Bio), 3 µL of 10× reaction buffer (included with the exonuclease), and 1.9 µL of nuclease-free water were added for a total of 30 µL, and the reaction was carried out at 37 °C for 30 min to remove the primers. The exonuclease was then inactivated at 80 °C for 20 min, and 10 µL of distilled water was added to bring the total volume to 40 µL. The post-reaction mixture was treated with AMPure XP beads (Beckman Coulter), and PCR amplicons were eluted in 17 µL of 10 mM Tris-HCl buffer (pH 8.0), 15.5 µL of which was collected and used as sample for the second PCR.

The second PCR was performed with the primers CS1-24 bp-F: 5'-TACACGACGCTCTTCCGATCTGTC-3', CS2-24 bp-R: 5'-AGACGTGTGCTCTTCCGATCTCAG-3' at a final concentration of 400 nM. The PCR reaction solution (25 µL total volume) was composed of 0.5 µL of PrimeSTAR GXL DNA Polymerase, 5.0 µL of 5× PrimeSTAR GXL Buffer, 2.0 µL of dNTP mixture (2.5 mM each), 1.0 µL of primer CS1-24 bp-F (10 µM), 1.0 µL of primer CS2-24 bp-R (10 µM) and 15.5 µL of sample. After adjusting the reaction solution, five cycles of 10 s at 98 °C, 15 s at 50 °C, and 30 s at 68 °C were performed in a thermal cycler, followed by 10 cycles of 10 s at 98 °C, 15 s at 60 °C, and 30 s at 68 °C. The samples were then stored at 4 °C. Subsequently, the samples were purified using AMPure XP beads as in the first PCR. The third PCR was performed with the primers PE1-CS1-ID50X: 5'-AAT GATACGGCGACCCGAGATCTACACXXXXXXXXXACACTCTTCCCTACACGACGCTCTCCGATCT GTC-3', PE2-CS2-ID70X: 5'-CAAGCAGAAGACGGCATAACGAGATXXXXXXXXXGTGACTGGAGTTTCAG ACGTGTGCTCTTCCGATCTCAG-3' (XXXXXXXXX: Illumina multiplex tags, Illumina) at a final concentration

of 400 nM. The PCR reaction solution (25 µL total volume) was composed of 0.5 µL of PrimeSTAR GXL DNA Polymerase, 5.0 µL of 5× PrimeSTAR GXL Buffer, 2.0 µL of dNTP mixture (2.5 mM each), 1.0 µL of primer PE1-CS1-ID50X (10 µM), 1.0 µL of primer PE2-CS2-ID70X (10 µM) and 15.5 µL of sample. After adjusting the reaction solution, five cycles of 10 s at 98 °C, 15 s at 60 °C, and 30 s at 68 °C were performed in a thermal cycler.

Library construction without the UMI tag was performed with the primers To2nd\_CRF2\_withCS: CTCTTC CGATCTGTCTAGTACYAACCATTTCATA and To2nd\_CRR1\_withCS: CTCTTCCGATCTCAGTGACCC YCTAGARAGAACG as in the first PCR. The subsequent procedures were performed under the same conditions as in the UMI library preparation. The constructed libraries were paired-end sequenced at 600 bp on a MiSeq next-generation sequencer (Illumina).

**HaCeD-Seq analytical pipeline.** *Without UMI tags.* After joining the paired-end reads with FLASH ver 1.2.11<sup>48</sup> using the option “-M 300”, 20 nucleotides from the 5′-end and 22 nucleotides from the 3′-end were deleted as the adapter and primers region, and the reads that contained one or more <Q30 bases were pruned using a self-made script. The variety and abundance of each haplotype was calculated by clustering the remaining reads at 100% agreement.

*With UMI tags.* After joining the paired-end reads with FLASH ver. 1.2.11 using the option “-M 300”, MAGERI ver. 1.1.1 was run with the following command:

```
java -Xmx32G -jar mageri.jar -M3 nnnNNNNNNNNNNNNNN -R1 LibX.extendedFragments.fastq -references.ref.fa -sample.name LibX.
```

Next, 20 nucleotides from the 5′-end and 22 nucleotides from the 3′-end were deleted as the adapter and primers regions of the obtained consensus sequence, and the number of 100% matching consensus sequences was calculated for each haplotype.

## Data availability

All MiSeq sequencing data were registered in the DNA Data Base of Japan (DDBJ) under accession number DRA010682.

Received: 4 November 2020; Accepted: 12 March 2021

Published online: 12 April 2021

## References

- Garlapati, D., Charankumar, B., Ramu, K., Madeswaran, P. & Ramana Murthy, M. V. A review on the applications and recent advances in environmental DNA (eDNA) metagenomics. *Rev. Environ. Sci. Biotechnol.* **18**, 389–411 (2019).
- Rees, H. C., Maddison, B. C., Middleditch, D. J., Patmore, J. R. M. & Gough, K. C. REVIEW: The detection of aquatic animal species using environmental DNA—a review of eDNA as a survey tool in ecology. *J. Appl. Ecol.* **51**, 1450–1459 (2014).
- Ruppert, K. M., Kline, R. J. & Rahman, M. S. Past, present, and future perspectives of environmental DNA (eDNA) metabarcoding: a systematic review in methods, monitoring, and applications of global eDNA. *Glob. Ecol. Conserv.* **17**, e00547 (2019).
- Thomsen, P. F. & Willerslev, E. Environmental DNA—an emerging tool in conservation for monitoring past and present biodiversity. *Biol. Conserv.* **183**, 4–18 (2015).
- Knudsen, S. W. *et al.* Species-specific detection and quantification of environmental DNA from marine fishes in the Baltic Sea. *J. Exp. Mar. Biol. Ecol.* **510**, 31–45 (2019).
- Salter, I., Joensen, M., Kristiansen, R., Steingrund, P. & Vestergaard, P. Environmental DNA concentrations are correlated with regional biomass of Atlantic cod in oceanic waters. *Commun. Biol.* **2**, 1–9 (2019).
- Stoeckle, M. Y. *et al.* Trawl and eDNA assessment of marine fish diversity, seasonality, and relative abundance in coastal New Jersey, USA. *ICES J. Mar. Sci.* <https://doi.org/10.1093/icesjms/fsaa225> (2020).
- Maruyama, A., Nakamura, K., Yamanaka, H., Kondoh, M. & Minamoto, T. The release rate of environmental DNA from juvenile and adult fish. *PLoS ONE* **9**, e114639 (2014).
- Jane, S. F. *et al.* Distance, flow and PCR inhibition: eDNA dynamics in two headwater streams. *Mol. Ecol. Resour.* **15**, 216–227 (2015).
- Yoshitake, K. *et al.* HaCeD-Seq: a novel method for reliable and easy estimation about the fish population using haplotype count from eDNA. *Mar. Biotechnol. N. Y.* **21**, 813–820 (2019).
- Kinde, I., Wu, J., Papadopoulos, N., Kinzler, K. W. & Vogelstein, B. Detection and quantification of rare mutations with massively parallel sequencing. *Proc. Natl. Acad. Sci.* **108**, 9530–9535 (2011).
- Casbon, J. A., Osborne, R. J., Brenner, S. & Lichtenstein, C. P. A method for counting PCR template molecules with application to next-generation sequencing. *Nucleic Acids Res.* **39**, e81–e81 (2011).
- Kivioja, T. *et al.* Counting absolute numbers of molecules using unique molecular identifiers. *Nat. Methods* **9**, 72–74 (2012).
- Vander Heiden, J. A. *et al.* pRESTO: a toolkit for processing high-throughput sequencing raw reads of lymphocyte receptor repertoires. *Bioinformatics* **30**, 1930–1932 (2014).
- Shugay, M. *et al.* MAGERI: Computational pipeline for molecular-barcoded targeted resequencing. *PLOS Comput. Biol.* **13**, e1005480 (2017).
- Clement, K., Farouni, R., Bauer, D. E. & Pinello, L. AmpUMI: design and analysis of unique molecular identifiers for deep amplicon sequencing. *Bioinformatics* **34**, i202–i210 (2018).
- Kumar, G., Kocour, M. & Kunal, S. P. Mitochondrial DNA variation and phylogenetic relationships among five tuna species based on sequencing of D-loop region. *Mitochondrial DNA Part A* **27**, 1976–1980 (2016).
- Nomura, S. *et al.* Genetic population structure of the Pacific bluefin tuna *Thunnus orientalis* and the yellowfin tuna *Thunnus albacares* in the North Pacific Ocean. *Fish. Sci.* **80**, 1193–1204 (2014).
- Gleiss, A. C., Schallert, R. J., Dale, J. J., Wilson, S. G. & Block, B. A. Direct measurement of swimming and diving kinematics of giant Atlantic bluefin tuna (*Thunnus thynnus*). *R. Soc. Open Sci.* **6**, 190203 (2019).
- Juan-Jordá, M. J., Mosqueira, I., Freire, J. & Dulvy, N. K. The conservation and management of tunas and their relatives: setting life history research priorities. *PLoS ONE* **8**, e70405 (2013).
- Shibata, M. *et al.* Transcriptomic features associated with energy production in the muscles of Pacific bluefin tuna and Pacific cod. *Biosci. Biotechnol. Biochem.* **80**, 1114–1124 (2016).

22. Swanson, D., Block, R. & Mousa, S. A. Omega-3 fatty acids EPA and DHA: health benefits throughout life. *Adv. Nutr. Bethesda Md* **3**, 1–7 (2012).
23. Collette, B. B. *et al.* Conservation. High value and long life—double jeopardy for tunas and billfishes. *Science* **333**, 291–292 (2011).
24. Kumai, H. & Miyashita, S. Life cycle of the Pacific bluefin tuna is completed under reared condition. *Nippon Suisan Gakkaishi Jpn.* **69**, 124–127 (2003).
25. Miyashita, S. *et al.* Maturation and spawning of cultured bluefin tuna, *Thunnus thynnus*. *Suisanzoushoku Jpn.* **48**, 475–488 (2000).
26. Cho, J. *et al.* Production performance of Pacific bluefin tuna *Thunnus orientalis* larvae and juveniles fed commercial diets and effects of switching diets. *Aquac. Sci.* **64**, 359–370 (2016).
27. Tsuji, S. *et al.* Environmental DNA analysis shows high potential as a tool for estimating intraspecific genetic diversity in a wild fish population. *Mol. Ecol. Resour.* <https://doi.org/10.1111/1755-0998.13165> (2020).
28. Tsuji, S. *et al.* Evaluating intraspecific genetic diversity using environmental DNA and denoising approach: a case study using tank water. *Environ. DNA* **2**, 42–52 (2020).
29. Ppyun, H. *et al.* Improved PCR performance and fidelity of double mutant Neq A523R/N540R DNA polymerase. *Enzym. Microb. Technol.* **82**, 197–204 (2016).
30. Hsieh, T. C., Ma, K. H. & Chao, A. iNEXT: an R package for rarefaction and extrapolation of species diversity (Hill numbers). *Methods Ecol. Evol.* **7**, 1451–1456 (2016).
31. Myers, R. A. & Worm, B. Rapid worldwide depletion of predatory fish communities. *Nature* **423**, 280–283 (2003).
32. Worm, B. *et al.* Rebuilding global fisheries. *Science* **325**, 578–585 (2009).
33. Bravington, M., Grewe, P. & Davies, C. Fishery-independent estimate of spawning biomass of southern bluefin tuna through identification of close-kin using genetic markers. *FRDC Report2007034CSIRO Aust.* (2014).
34. Bravington, M. V., Grewe, P. M. & Davies, C. R. Absolute abundance of southern bluefin tuna estimated by close-kin mark-recapture. *Nat. Commun.* **7**, 13162 (2016).
35. Song, N., Jia, N., Yanagimoto, T., Lin, L. & Gao, T. Genetic differentiation of *Trachurus japonicus* from the Northwestern Pacific based on the mitochondrial DNA control region. *Mitochondrial DNA* **24**, 705–712 (2013).
36. Zhu, Y., Cheng, Q. & Rogers, S. M. Genetic structure of *Scomber japonicus* (Perciformes: Scombridae) along the coast of China revealed by complete mitochondrial cytochrome b sequences. *Mitochondrial DNA Part DNA Mapp. Seq. Anal.* **27**, 3828–3836 (2016).
37. Tzeng, T.-D. Population structure and historical demography of the spotted mackerel (*Scomber australasicus*) off Taiwan inferred from mitochondrial control region sequencing. *Zool. Stud.* **8**, 656–663 (2007).
38. Ichinokawa, M., Okamura, H. & Kurota, H. The status of Japanese fisheries relative to fisheries around the world. *ICES J. Mar. Sci.* **74**, 1277–1287 (2017).
39. Pikitch, E. K. *et al.* Ecosystem-based fishery management. *Science* **305**, 346–347 (2004).
40. Yates, M. C., Fraser, D. J. & Derry, A. M. Meta-analysis supports further refinement of eDNA for monitoring aquatic species-specific abundance in nature. *Environ. DNA* **1**, 5–13 (2019).
41. Lacoursière Roussel, A., Rosabal, M. & Bernatchez, L. Estimating fish abundance and biomass from eDNA concentrations: variability among capture methods and environmental conditions. *Mol. Ecol. Resour.* **16**, 1401–1414 (2016).
42. Yamamoto, S. *et al.* Environmental DNA as a ‘Snapshot’ of fish distribution: a case study of Japanese jack mackerel in Maizuru bay, Sea of Japan. *PLoS ONE* **11**, e0149786 (2016).
43. Untergasser, A. *et al.* Primer3—new capabilities and interfaces. *Nucleic Acids Res.* **40**, e115 (2012).
44. Sekino, M. & Yamashita, H. Mitochondrial DNA barcoding for Okinawan oysters: a cryptic population of the Portuguese oyster *Crassostrea angulata* in Japanese waters. *Fish. Sci.* **79**, 61–76 (2013).
45. Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* **30**, 2725–2729 (2013).
46. Minamoto, T., Naka, T., Moji, K. & Maruyama, A. Techniques for the practical collection of environmental DNA: filter selection, preservation, and extraction. *Limnology* **17**, 23–32 (2016).
47. Miya, M. *et al.* MiFish, a set of universal PCR primers for metabarcoding environmental DNA from fishes: detection of more than 230 subtropical marine species. *R. Soc. Open Sci.* **2**, 150088 (2015).
48. Magoč, T. & Salzberg, S. L. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27**, 2957–2963 (2011).

## Acknowledgements

This study was financially supported by the Japan Fisheries Research and Education Agency (No. 3BA101). Some computations were performed on the NIG supercomputer at the ROIS National Institute of Genetics, Japan. Some pictures from the DBCLS Togo Picture Gallery were used in this paper.

## Author contributions

Conceptualization, A.F.; Formal Analysis, K.Y.; Funding Acquisition, A.F. and S.W.; Investigation, A.F., A.M., M.S., M.Y., Y.N. and R.N.; Methodology, K.Y. and A.F.; Project Administration, A.F., S.W. and T.K.; Resources, M.K., Y.T. and K.N.; Writing—Original Draft Preparation, K.Y., A.F., T.K. and S.W.; Writing—Review and Editing, K.Y., A.F., A.K., S.A. and S.W.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-86190-6>.

**Correspondence** and requests for materials should be addressed to S.W.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021