# Estimating fish population abundance by integrating quantitative data on environmental DNA and hydrodynamic modelling

Keiichi Fukaya,[1,2*] Hiroaki Murakami,[3] Seokjin Yoon,[4] Kenji Minami,[5] Yutaka Osada,[6]
Satoshi Yamamoto,[7] Reiji Masuda,[3] Akihide Kasai,[4] Kazushi Miyashita,[8]
Toshifumi Minamoto,[9] and Michio Kondoh[6]

[1] *National Institute for Environmental Studies, Tsukuba, Ibaraki 305-8506, Japan*
[2] *The Institute of Statistical Mathematics, Tachikawa, Tokyo 190-8562, Japan*
[3] *Maizuru Fisheries Research Station, Field Science Education and Research Center, Kyoto University, Maizuru, Kyoto 625-0086, Japan*
[4] *Faculty of Fisheries Sciences, Hokkaido University, Hakodate, Hokkaido 041-8611, Japan*
[5] *Estuary Research Center, Shimane University, Matsue, Shimane 690-8504, Japan*
[6] *Graduate School of Life Sciences, Tohoku University, Sendai, Miyagi 980-8578, Japan*
[7] *Laboratory of Animal Ecology, Department of Zoology, Graduate School of Science, Kyoto University, Kyoto 606-8502, Japan*
[8] *Field Science Center for Northern Biosphere, Hokkaido University, Hakodate, Hokkaido 040-0051, Japan*
[9] *Graduate School of Human Development and Environment, Kobe University, Kobe, Hyogo 657-8501, Japan*
*fukaya.keiichi@nies.go.jp

**Abstract**

We propose a general framework of abundance estimation based on spatially replicated quantitative measurements of environmental DNA in which production, transport, and degradation of DNA are explicitly accounted for. Application to a Japanese jack mackerel (*Trachurus japonicus*) population in Maizuru Bay revealed that the method gives an estimate of population abundance comparable to that of a quantitative echo sounder method. These findings indicate the ability of environmental DNA to reliably reflect population abundance of aquatic macroorganisms and may offer a new avenue for population monitoring based on the fast, cost-effective, and non-invasive sampling of genetic information.

Knowledge on the distribution and abundance of species is crucial for ecology and related applied fields such as wildlife management and fisheries. The detection and quantification of environmental DNA (eDNA) is an emerging methodology for ecological studies and could enhance the ability of investigators to infer occurrence and abundance of species. This approach has been applied, especially but not limited to, to aquatic species such as fish and amphibians and has been identified as a powerful and yet cost-effective tool for species detection (Bohmann *et al.* 2014, Rees *et al.* 2014, Thomsen & Willerslev 2015, Goldberg *et al.* 2016, Deiner *et al.* 2017, Hansen *et al.* 2018). Challenges remain, however, in quantitative applications of eDNA. Since earlier studies revealed positive correlations between species abundance and eDNA concentration (Takahara *et al.* 2012, Thomsen *et al.* 2012, Goldberg *et al.* 2013, Pilliod *et al.* 2013, Eichmiller *et al.* 2014), it has been expected that local population abundance may be inferred by measuring the concentration of eDNA at a given locality. Indeed, an analytical framework proposed recently for eDNA-based

1

abundance estimation assumes a probability distribution that represents the quantitative relation between eDNA concentration and the underlying population size (Chambert *et al.* 2018). Nonetheless, such a definite relation may not always be present, possibly depending on e.g. the shedding rate, transport, and exogenous input of eDNA (Pilliod *et al.* 2013, Eichmiller *et al.* 2014, Lacoursière-Roussel *et al.* 2016, Yamamoto *et al.* 2016, Jo *et al.* 2017).

The fundamental factors that underlie such context dependency are the 'ecology of eDNA': the distribution of eDNA in space and time stems from processes governing the origin, state, transport, and fate of eDNA particles (Barnes & Turner 2016). Thus, in applications of the eDNA methodology, detailed information about such processes may be critical. Without relevant knowledge of these processes, for example, the spatial and temporal scales of information provided by eDNA remain largely uncertain (Thomsen & Willerslev 2015, Goldberg *et al.* 2016, Hansen *et al.* 2018). Therefore, here, our purpose was to develop a general approach to eDNA-based abundance estimation that can fully account for the ecology of eDNA, i.e. the rate of production and degradation of eDNA as well as the transport of eDNA within a flow field in an aquatic area of interest. We use a *tracer model*: a numerical hydrodynamic model that can simulate the distribution of eDNA concentrations within an aquatic area. Under certain assumptions, the behaviour of the model can also be regarded mathematically as a linear function of an input vector representing the distribution of population abundance levels (densities) within the area. The inference of population abundance then boils down to Bayesian estimation of coefficients of a generalised linear model (see Methods for details).

We applied this approach to a population of the Japanese jack mackerel (*Trachurus japonicus*, a commercially important fish species) in Maizuru Bay, Japan (Fig. 1). The bay has a surface area of $\sim$22.87 km$^2$ with a maximal water depth of approximately 30 m, where the jack mackerel is numerically the most dominant fish species. The field work was conducted during a season in which the jack mackerel population in the bay is dominated by new recruits. We cruised the bay on two days in June 2016 to collect water samples at 100 stations and to conduct an acoustic survey. On the basis of the eDNA concentration measurements and a tracer model configured for Maizuru Bay, we obtained an estimate of fish population abundance in the bay. This estimate was then verified via a parallel estimate of abundance obtained by a quantitative echo sounder method.

The abundance estimates yielded by the two methods were comparable; the point estimate of the eDNA method was of the same order of magnitude as that of the quantitative echo sounder estimate, which was covered by the 95% highest posterior density interval (HPDI) of the eDNA-based estimate (Table 1). Moreover, we could identify a coordinate of grids in which density of jack mackerels was estimated to be unrealistically high; fish abundance in this location was estimated at as much as tens of millions of individuals (posterior median and 95% HPDI: $1.35 \times 10^7$ [0.00 to $1.77 \times 10^7$] individuals; Fig. 1b). It is located next to a wholesale fish market (Fig. 1a), which has been suspected as a significant source of exogenous jack mackerel eDNA in Maizuru Bay (Yamamoto *et al.* 2016, Jo *et al.* 2017). We therefore regarded the extreme estimates in these cells as resulting from a massive eDNA input from the market and excluded them from the inference of the bay scale fish abundance. This correction reduced the estimate of fish abundance in the bay, whereas the 95% HPDI still covered the echo sounder estimate (Table 1).

The eDNA methods are rapidly developing technologies that have a great potential to facilitate the understanding and management of aquatic species, although their quantitative

2

| Method | Abundance estimates | 95% Bayesian credible interval |
|--------|--------------------|--------------------------------|
| eDNA + hydrodynamic model | $3.31 \times 10^7$ | $(2.32 \times 10^7, 6.32 \times 10^7)$ |
| (fish market cells omitted) | $2.23 \times 10^7$ | $(0.77 \times 10^7, 5.29 \times 10^7)$ |
| Quantitative echo sounder | $3.91 \times 10^7$ | — |

Table 1. **Estimates of Japanese jack mackerel abundance in Maizuru Bay.** The second row of the eDNA method gives the abundance estimate that excluded the grid cells close to the wholesale fish market (indicated in Fig. 1a), which were identified as extraordinary eDNA sources. The point abundance estimates and credible intervals are presented as posterior medians and highest posterior density intervals, respectively. In both estimation methods, estimates are obtained under the assumption that the size of jack mackerel individuals was 3 cm in body length and 1 g in body weight (see Methods).

applications are still the critical step. A number of quantitative eDNA applications uncovered a positive association between eDNA concentration and abundance of a target species (Takahara *et al.* 2012, Thomsen *et al.* 2012, Goldberg *et al.* 2013, Pilliod *et al.* 2013, Eichmiller *et al.* 2014, Lacoursière-Roussel *et al.* 2016, Yamamoto *et al.* 2016, Jo *et al.* 2017). With the aid of a well-designed sampling scheme and an associated statistical model, such relations can help to quantify abundance at multiple locations, especially in lentic systems where advection of eDNA is limited (Chambert *et al.* 2018). This study presents a novel approach to abundance estimation based on quantitative eDNA measurements into which a numerical hydrodynamic model (i.e. the tracer model) is incorporated to explicitly account for the details of the ecology of eDNA. It may be flexibly applied to a wide array of aquatic systems in which hydrodynamics and rates of eDNA shedding and degradation are modelled, thereby broadening the scope of the general idea implemented recently in a one-dimensional lotic system with a single eDNA source (Sansom & Sassoubre 2017) and in a river network system with multiple eDNA sources (Carraro *et al.* 2018). The application of the proposed approach to the Japanese jack mackerel population in Maizuru Bay indicates that abundance of species can be reliably estimated by means of eDNA in a mesoscale lotic system. Furthermore, the results revealed that the method can distinguish major exogenous sources of eDNA, which have been recognised as a nuisance factor in eDNA applications especially for species subject to fishery (Yamamoto *et al.* 2016, Jo *et al.* 2017).

The proposed framework, however, has several limitations in its current form. It requires several key assumptions, such as the stationarity (i.e. demographic closure) of the population and homogeneity of individuals in terms of their rate of eDNA shedding. In addition, the number of eDNA samples may typically be smaller than the number of grid cells in the tracer model, thus requiring some sort of models explaining the association between population density and measured covariates and/or regularisation (i.e. prior specification) to make a statistical inference (see Methods). Although our results indicated that the method can be applied even with these limitations, further methodological development would be warranted. A promising approach among quantitative eDNA applications is to combine eDNA measurements and classical protocols for abundance estimation (Chambert *et al.* 2018); this strategy is also likely to improve the general approach proposed here.

It has been argued that in an application of the eDNA method, careful consideration of

96 details of the ecology of eDNA is critical (Bohmann *et al.* 2014, Rees *et al.* 2014, Thomsen &
97 Willerslev 2015, Barnes & Turner 2016, Goldberg *et al.* 2016, Deiner *et al.* 2017, Hansen *et al.* 2018).
98 We implemented this idea in a quantitative eDNA method, leading to integration of eDNA
99 concentration measurements and hydrodynamic modelling for abundance estimation. Because the
100 research on aquatic eDNA of macroorganisms is still in its infancy since its discovery (Ficetola *et al.*
101 2008), more work is needed to elucidate the processes that determine a distribution of eDNA in the
102 field; knowledges on the ecology of eDNA will help to improve the accuracy of quantitative eDNA
103 approaches. The relatively less explored field of quantitative eDNA applications lies in the
104 multispecies context, which involves eDNA metabarcoding rather than the targeted quantitative
105 PCR (qPCR) method (Deiner *et al.* 2017). A quantitative metabarcoding technique (Ushio *et al.*
106 2018) may hold great promise for enabling researchers to analyse many aquatic species at a time.
107 Exploring between-species differences in the rate of eDNA shedding and degradation may therefore
108 be worthwhile. In addition to remarkable efficiency in species detection, we expect that eDNA
109 methodologies can enhance the ability of investigators to gain quantitative insights into aquatic
110 ecosystems.

# Methods

## A general framework for abundance estimation

### The tracer model as a linear projection function

114 Here, we define a *tracer model* as a numerical hydrodynamic model that simulates generation,
115 transport, and decay of particles (i.e. eDNA) on the basis of a flow field determined by given
116 physical conditions within an aquatic area of interest. In this study, we assume a tracer model for a
117 three-dimensional discrete space in which the entire aquatic area of interest is discretised into grid
118 cells of known volume. A tracer model can in principle simulate the ecology of eDNA and thus
119 derives a spatial distribution of eDNA within the aquatic area, given that per capita and unit time
120 shedding rates of eDNA, degradation rates of eDNA, and density (or equivalently, abundance) of
121 organisms in each grid cell are specified, in addition to the flow field. The main idea that underlies
122 the framework we propose is that we can regard a tracer model as a function that takes a vector of
123 cell level density of organisms as an input and outputs eDNA concentration in each grid cell at a
124 point in time; thus, the inference of abundance is an inverse problem: finding an input vector of a
125 tracer model (i.e. density of organisms in each grid cell) that best explains measurements of eDNA
126 concentration that are collected at a point in time and are replicated spatially within the aquatic
127 area of interest.

128 Nevertheless, such a problem is difficult to solve under the general conditions where both the
129 environment and abundance vary in a complex manner. We therefore make several key assumptions
130 that simplify the problem. Firstly, we assume that during two time points $t$ and $s$ $(< t)$, key
131 environmental variables for hydrodynamic processes are known from some observations and/or
132 model prediction so that the flow field can be determined and plugged in to the tracer model. Here,
133 $t$ refers to the point in time at which eDNA concentration is observed at multiple locations within
134 the aquatic area, and $s$ denotes some point in time sufficiently far away from $t$ such that eDNA
135 concentration at $t$ is virtually independent from that at $s$. Secondly, we assume that the rates of

4

136 production and degradation of eDNA are known in each grid cell during the period between $s$ and $t$.
137 They may either be regarded as constant across space and time or assumed to vary depending on
138 known environmental variables, such as water temperature, salinity, and pH, so that the rates of
139 generation and disappearance of eDNA can be determined completely in the tracer model. In
140 addition, we assume that these rates are independent of the eDNA concentration, and thus both
141 production and degradation of eDNA are linear processes. Third, we suppose that in each grid cell,
142 all eDNA particles arise exclusively from individuals of the target species that are identical in their
143 eDNA-shedding rate. Finally, we assume that abundance is stationary in each grid cell throughout
144 the period between $s$ and $t$ (i.e. the demographic closure assumption; Williams *et al.* 2002).

145     Under these assumptions, a tracer model can be regarded as a linear function. We denote
146 density of organisms in cell $i$ $(i = 1, \ldots, M)$ by $x_i$ and define $\mathbf{x} = (x_1, x_2, \ldots, x_M)$. Let us denote the
147 water volume of each cell by $\mathbf{v} = (v_1, v_2, \ldots, v_M)$ so that abundance in mesh $i$ and in the whole
148 aquatic area is expressed as $v_i x_i$ and $\mathbf{v}^\top \mathbf{x}$, respectively (here, $\mathbf{a}^\top$ means the transpose of vector $\mathbf{a}$).
149 The tracer model predicts eDNA concentration in each grid cell at time point $t$ that results from the
150 generation, advection, diffusion, and degradation of eDNA occurring between $s$ and $t$ within a given
151 flow field, which we denote (without an explicit index of $t$) by $\mathbf{w} = (w_1, w_2, \ldots, w_M)$. If $a_{ij}$ is defined
152 as the (per unit density) contribution of mesh $j$ to eDNA concentration in mesh $i$ at time $t$, then
153 eDNA concentration can be expressed as $w_i = a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{iM}x_M$. If we designate
154 $\mathbf{A} = (a_{ij})_{M \times M}$, then this equation can be written in a matrix form as $\mathbf{w} = \mathbf{A}\mathbf{x}$. Thus, although a
155 tracer model indeed represents temporal evolution of eDNA concentration within the period between
156 $s$ and $t$ according to some differential equations (presented below), its behaviour can be described
157 simply — under the assumptions noted above — by matrix $\mathbf{A}$, which projects the vector of density $\mathbf{x}$
158 onto the vector of eDNA concentration $\mathbf{w}$. For $i = 1, \ldots, M$, the $i$th column of $\mathbf{A}$ can be obtained
159 numerically as a result of execution of the tracer model between time points $s$ and $t$ with a vector of
160 density in which cell $i$ has a unit density and all other cells have 0 density.

161 **Fitting the tracer model to eDNA concentration data**

162 We assume that eDNA concentration was measured in $N$ samples collected within the aquatic area
163 of interest at a point in time (or, in practice, within a sufficiently short period). Let us denote the
164 observed eDNA concentration in sample $n$ by $y_n$ $(n = 1, \ldots, N)$ and express it with vector
165 $\mathbf{y} = (y_1, \ldots, y_N)$. In the following text, we suppose that all eDNA measurements are positive (i.e.,
166 $y_n > 0$). Note, however, that negative samples could also be included in the analysis given that the
167 detection process of eDNA is modelled jointly (Carraro *et al.* 2018). We define $i(n)$ as an index
168 variable that means the index of the cell in which sample $n$ was obtained. If we let $\mathbf{B} = (a_{i(n)j})_{N \times M}$,
169 the prediction of the tracer model for the data vector, as a function of density vector $\mathbf{x}$ is then
170 expressed as $\mathbf{B}\mathbf{x}$.

171     Because the tracer model yields a linear predictor for $\mathbf{y}$, we can apply the (generalised) linear
172 modelling approach (McCullagh & Nelder 1989) to estimate density vector $\mathbf{x}$; in particular, we can
173 regard $\mathbf{B}$ and $\mathbf{x}$ as a design matrix and a vector of coefficients of a linear regression model,
174 respectively (note that because $\mathbf{x}$ represents density, the searches for estimates should be within the
175 space of parameters such that $x_i \geq 0$ for all $i$). For example, we can consider the following normal
176 linear model:

$$\mathbf{y} \sim \mathcal{N}\left(\mathbf{Bx}, \sigma^2 \mathbf{I}_N\right). \tag{1}$$

where $\mathcal{N}\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right)$ is a multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, $\sigma^2$ is a residual variance of the linear model, and $\mathbf{I}_m$ is a $m \times m$ identity matrix. A maximum likelihood estimation gives estimate $\hat{\mathbf{x}}$ that minimises the residual square error $|\mathbf{y} - \mathbf{B}\hat{\mathbf{x}}|_2^2$. Alternatively, we can fit the model on a logarithmic scale; this approach may be more reasonable than the above model when a lognormal error structure better represents eDNA concentration data as is often the case in quantitative eDNA studies (e.g. Takahara *et al.* 2012, Thomsen *et al.* 2012, Eichmiller *et al.* 2014, Wilcox *et al.* 2016, Jo *et al.* 2017). The alternative model can be written as

$$\log \mathbf{y} \sim \mathcal{N}\left(\log \mathbf{Bx}, \sigma^2 \mathbf{I}_N\right), \tag{2}$$

which is a generalised linear model with an exponential link function: a less popular but still appropriate within the generalised linear modelling framework given that $x_i \geq 0$ for all $i$ (McCullagh & Nelder 1989). The maximum likelihood method for this model yields estimate $\hat{\mathbf{x}}$ that minimises the residual square error $|\log \mathbf{y} - \log \mathbf{B}\hat{\mathbf{x}}|_2^2$.

The standard maximum likelihood approach is, however, not applicable to these models when $M > N$ because the maximum likelihood estimate of $\mathbf{x}$ is not uniquely identified in this setting. This may be a typical situation at a reasonable level of spatial discretisation for the tracer model and sampling effort of eDNA. When some covariates, assumed to covary with density, are available for each cell, a (generalised) linear model for density can be introduced to effectively reduce the number of unknown parameters (Carraro *et al.* 2018). Specifically, density of the target species can be modelled, for example, as $\log \mathbf{x} = \mathbf{Z}\boldsymbol{\beta}$, where $\mathbf{Z}$ is a matrix of covariates, and $\boldsymbol{\beta}$ is a vector of coefficients (including an intercept). Otherwise, additional regularisation is necessary to make an inference based on such singular models. The regularisation method often employed for regression models is to impose a penalty on the size of regression coefficients; a typical example includes ridge regression and lasso, which can be interpreted in general as a Bayesian inference of the model with a specific prior on the regression coefficients (Hastie *et al.* 2009). Thus, inference can be achieved via a Bayesian model-fitting approach such as empirical Bayes and the full-Bayesian inference (Karabatsos 2018).

## An application to a marine fish population

### The Japanese jack mackerel in Maizuru Bay

The study was conducted in Maizuru Bay (Kyoto prefecture, Japan; 35°29′N, 135°23′E) to estimate abundance of the jack mackerel (*T. japonicus*) via concentration of eDNA. The bay has a surface area of $\sim$22.87 km$^2$ with a maximum water depth of approximately 30 m, and connects with Wakasa Bay through a narrow bay mouth in its north (Fig. 1).

According to long-term underwater visual surveys, the jack mackerel is numerically the most dominant fish species in shallow ($< 10$ m in depth) coastal waters in this area (Masuda 2008); their body size ranges from 10 to 45 mm in standard length offshore and 40–120 mm standard length in the shallow rocky reef habitat (Masuda *et al.* 2008). The study was conducted during the peak season of jack mackerel recruitment from the offshore pelagic zone to a coastal shallow reef habitat,

6

213 where the jack mackerel population in the bay is dominated by new recruits. In the following

214 analysis, we therefore assumed that the population is represented by individuals of size $\sim 3$ cm

215 (body length) and $\sim 1$ g (body weight; see Supplementary information).

## Measurement of eDNA concentration

217 We conducted the water sampling on 21 and 22 June 2016 from a research vessel at 100 stations

218 located approximately on $\sim$400 m grids in Maizuru Bay (Fig. 1). Samples were collected at 53

219 stations in the eastern part of the bay on the first day and at 47 stations in the western part on the

220 second day. The average water depth at the 100 stations was $\sim$15m. On both days, the survey began

221 from the mouth of the bay and ended in the inner most part of the bay. The survey was approved by

222 the harbourmaster of Maizuru Bay (Permission number 160 issued on 5th May 2016).

223     At each sampling station, we captured sea water at three depths: the surface, middle, and

224 bottom. The middle and bottom depths were defined as 5 m from the surface, which was just below

225 the pycnocline, and 1 m above the sea floor, respectively. Water samples were collected with a ladle

226 for surface water and vanDorn samplers for middle and bottom water. For each station and depth, a

227 1 L water sample was placed in a plastic bottle, which was rinsed in advance with a subset of

228 captured water. We then immediately added 1 mL of 10% benzalkonium chloride to the samples and

229 mixed them gently to prevent degradation of DNA (Yamanaka *et al.* 2017). The bottles of water

230 samples were stored in opaque containers to avoid sunlight.

231     We filtered water samples on the same day of the field survey through a 47 mm diameter glass

232 microfiber filter (nominal pore size 0.7 $\mu$m, GE Healthcare Life Science [Whatman]) using an

233 aspirator in a laboratory at Maizuru Fisheries Research Station, Kyoto University. The filters were

234 folded so that the filter surface faced inward and were wrapped into aluminium foil to store at

235 $-20$°C until eDNA extraction. It took less than 7 h to complete all operations from the water

236 collection to the filtration. To prevent carryover of eDNA, filtration devices were bleached by means

237 of 0.1% sodium hypochlorite for at least 5 min and then were washed and rinsed with tap water and

238 distilled water, respectively, to clear the remaining sodium hypochlorite. This bleaching process was

239 validated by a series of negative controls of filtration undertaken for every sequence of 15 filtrations

240 in which 1 L of distilled water was filtered with bleached equipment.

241     All samples and negative controls of filtration were subjected to eDNA extraction and

242 subsequent quantitative PCR (qPCR). eDNA extraction was conducted by following the procedure

243 of Yamamoto *et al.* (2016), which eventually yielded 100 $\mu$L of a DNA solution. We determined the

244 concentration of mitochondrial cytochrome b (CytB) of the jack mackerel by qPCR on a LightCycler

245 96 machine (Roche). The primers and probe used in the qPCR were as follows: forward primer,

246 5′–CAG ATA TCG CAA CCG CCT TT–3′; reverse primer, 5′–CCG ATG TGA AGG TAA ATG

247 CAA A–3′; probe, 5′–FAM-TAT GCA CGC CAA CGG CGC CT–TAMRA–3′ (Yamamoto *et al.*

248 2016). This primer set amplifies 127 bp of the CytB gene. The PCR reaction solution was 20 $\mu$L: 2

249 $\mu$L of the extracted DNA solution, a final concentration of 900 nM forward and reverse primers and

250 125 nM TaqMan probe in 1 $\times$ TaqMan master mix (TaqMan gene expression master mix; Life

251 Technologies). The thermal program for the qPCR was as follows: 2 min at 50°C, 10 min at 95°C,

252 and 55 cycles of 15 sec at 95°C and 1 min at 60°C. To draw quantification standard curves, we

253 simultaneously performed PCR on 2 $\mu$L of artificial DNA solutions that contained $3 \times 10^1$ to $3 \times 10^4$

254 copies of our target sequence. qPCR was carried out in triplicate for each sample and standard. In

7

addition, a 2 $\mu$L pure water sample was analysed simultaneously in triplicate as a negative control of PCR. In all the runs, $R^2$ values of calibration curves were more than 0.99, the range of slopes was between $-3.859$ and $-3.512$, and the range of $y$-intercepts was between 38.34 and 40.36. No eDNA of the jack mackerel was detected in any negative control sample of filtration and PCR.

## Development of the tracer model

To obtain the flow field in Maizuru Bay, we configured the Princeton ocean model (POM) with a scaled vertical coordinate (i.e. the sigma coordinate system; Mellor 2002) for the bay. The model represented Maizuru Bay by 20,484 grid cells. Specifically, the bay was discretised by 2,276 horizontal lattice grids at a resolution of 100 m, and the grids had nine non uniform vertical layers, with finer resolution near the surface; the sigma coordinate was set as $\sigma = 0.000, -0.041, -0.088, -0.150, -0.245, -0.374, -0.510, -0.646, -0.796,$ and $-1.000$. The configuration of the model was achieved by means of the bottom topography of the bay, data and model estimates of surface meteorological conditions, estimated river discharges, and the model results of Wakasa Bay as the open boundary conditions (Yoon & Kasai 2017); additional details are described elsewhere (Kasai & Yoon 2018). The model simulated flow fields within the bay from 1 June 2016, under the initial conditions interpolated from the model results of Yoon & Kasai (2017), to the final day for the water sampling (i.e. 22 June 2016). Time steps of the simulation were set to 0.1 s for the external mode and 3 s for the internal mode.

We then incorporated eDNA of jack mackerels into the POM configured for Maizuru Bay as a passive tracer to simulate its concentration within the flow field. The evolution of eDNA concentration in a grid cell, denoted by $c$, is represented as

$$\frac{dc}{dt} = -\lambda c + \beta x + \mathrm{Adv} + \mathrm{Diff}, \tag{3}$$

where $x$ is the density of jack mackerels in the cell, $\lambda$ represents a degradation rate of eDNA, and $\beta$ is a per-capita shedding rate of jack mackerel DNA. Adv and Diff are the advection and diffusion terms, respectively, which were determined implicitly by running the POM for Maizuru Bay. The eDNA degradation rate was assumed to be constant and was adopted from an estimate obtained in tank experiments where the same species-specific primer set was employed ($\lambda = 0.044$ h$^{-1}$; Jo *et al.* 2017). The eDNA shedding rate of the jack mackerel was assumed to be constant; it was derived mathematically and found to be $\beta = 9.88 \times 10^4$ copies per individual per hour, according to the results of tank experiments conducted by Maruyama *et al.* (2014) and Jo *et al.* (2017). Details of this derivation are provided in Supplementary information.

## Estimation of jack mackerel abundance based on eDNA and the tracer model

We fitted the logarithmic model (Eq. 2) to the eDNA concentration data collected in Maizuru Bay. During the model fitting, we omitted negative samples in which the number of remaining observations was $N = 729$. For vector of density $\mathbf{x}$, we specified an independent lognormal prior with unknown prior mean $\mu$ and standard deviation $\tau$:

$$\log \mathbf{x} \sim \mathcal{N} \left( \log \mu \mathbf{1}_M, \tau^2 \mathbf{I}_M \right), \tag{4}$$

8

290 where $\mathbf{1}_M$ represents a vector of all ones with length $M$. Because $N$ was significantly smaller than
291 $M$, we were pessimistic about estimating the spatial variation in cell level density with reasonable
292 precision. Our main goal of the inference was therefore to quantify the bay level abundance $\mathbf{v}^\top \mathbf{x}$
293 along with its uncertainty.

294     With uniform positive priors on $\mu$, $\tau$, and $\sigma$, we fitted the model via a fully Bayesian
295 approach. Posterior samples were obtained by the Markov chain Monte Carlo (MCMC) method
296 implemented in Stan (version 2.18.1; Carpenter *et al.* 2017) in which three independent chains of
297 10,000 iterations were generated after 1,000 warm-up iterations. Each chain was thinned at intervals
298 of 10 to save the posterior sample.

299     Convergence of the posterior was checked for each parameter with the $\hat{R}$ statistic. Posterior
300 convergence was achieved at a recommended degree ($\hat{R} < 1.1$; Gelman *et al.* 2013) in almost all
301 parameters except $\log x$ in four cells. We decided, however, that the results are solid because the
302 posterior of the bay level abundance — the target of the inference — fully converged. The
303 goodness-of-fit assessment of the model, measured by the $\chi^2$-discrepancy statistic (Conn *et al.* 2018),
304 gave no clear indication of a lack of model fit (Bayesian $p$ value: 0.404).

### Estimation of jack mackerel abundance from quantitative echo sounder data

306 An independent estimate of jack mackerel abundance was obtained based on a calibrated
307 quantitative echo sounder by a standard acoustic survey method (Simmonds & MacLennan 2005).
308 The acoustic survey was conducted during the survey cruise for the water sampling (described
309 above). We used the KSE300 echo sounder (Sonic Co. Ltd., Tokyo, Japan) with two transducers
310 (T-182, 120 kHz, and T-178, 38 kHz; beam type, split-beam; beam width, 8.5°; pulse duration, 0.3
311 ms; ping rate, 0.2 s), which were mounted off the side of the research vessel at a depth of 1 m. The
312 acoustic devices were operated during the entire survey cruise to record all acoustic reflections,
313 except when the research vessel stopped at each sampling station where the recording was stopped
314 to avoid reflection from the sampling gear and cables. The research vessel ran at ∼4 knots, on
315 average, between the sampling stations. The echo intensity data were denoised and cleaned in
316 Echoview ver. 9.0 (Echoview Software Pty. Ltd., Tasmania, Australia). We omitted signals between
317 the sea bottom and 0.5 m above it to exclude the acoustic reflection from the sea floor. Additionally,
318 we eliminated signals from sea nettles (*Chrysaora pacifica*) by filtering reflections of $-75$ dB.

319     From the obtained acoustic data, the reflections of jack mackerel were extracted by the
320 volume back scattering strength difference ($\Delta S_V$) method (Miyashita *et al.* 2004, Simmonds &
321 MacLennan 2005). $\Delta S_V$ was defined as the difference in the volume backscattering strength ($S_V$)
322 between the two frequencies as follows:

$$\Delta S_V = S_{V\,120\ \text{kHz}} - S_{V\,38\ \text{kHz}}. \tag{5}$$

323 According to field validation in Maizuru Bay combining acoustic surveys and visual confirmation of
324 jack mackerel schools by snorkelling, we assumed the range of $\Delta S_V$ of jack mackerel between $-6.4$
325 and 5.2 dB. This criterion discriminates the jack mackerel from larval Japanese anchovy (*Engraulis*
326 *japonicus*), the subdominant species in the bay (Masuda 2008), which reflects the high frequency
327 echo strongly as compared to low frequency (Ito *et al.* 2011) and was used to determine $S_V$ of the
328 jack mackerel in 1 m$^3$ water cubes in Echoview ver. 9.0.

9

Density of jack mackerel in a 1 m$^3$ water cube, denoted by $D$, was estimated as

$$D = \frac{10^{\frac{S_{V120\text{ kHz}}}{10}}}{10^{\frac{TS}{10}}} \tag{6}$$

where $TS$ is the target strength of an individual jack mackerel. By assuming that jack mackerel population in the bay was dominated by individuals of the size 3 cm, we chose $TS = -59.6$ dB (Nakamura *et al.* 2013, Yamamoto *et al.* 2016). The fish density on the echo sounder track lines was then matched with the grid specification of the tracer model by a box averaging method. In particular, fish density in each grid cell was estimated by a geometric mean of $D$ taken over a 500 m square block that surrounds the grid cell. For grid cells in which any $D$ was not available in their square block owing to a lack of the acoustic data, fish density was estimated by means of a geometric mean of the fish density taken across the other grid cells. Finally, the bay level abundance was estimated as a sum of the product of grid level density and water volume of each cell.

## Data availability

The datasets generated and analysed during the current study are available from the corresponding author upon reasonable request.

# References

Barnes, M.A. & Turner, C.R. (2016) The ecology of environmental DNA and implications for conservation genetics. *Conservation Genetics*, **17**, 1–17.

Bohmann, K., Evans, A., Gilbert, M.T.P., Carvalho, G.R., Creer, S., Knapp, M., Douglas, W.Y. & de Bruyn, M. (2014) Environmental DNA for wildlife biology and biodiversity monitoring. *Trends in Ecology & Evolution*, **29**, 358–367.

Carpenter, B., Gelman, A., Hoffman, M.D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P. & Riddell, A. (2017) Stan: a probabilistic programming language. *Journal of Statistical Software*, **76**, 1–32.

Carraro, L., Hartikainen, H., Jokela, J., Bertuzzo, E. & Rinaldo, A. (2018) Estimating species distribution and abundance in river networks using environmental DNA. *Proceedings of the National Academy of Sciences*, **115**, 11724–11729.

Chambert, T., Pilliod, D.S., Goldberg, C.S., Doi, H. & Takahara, T. (2018) An analytical framework for estimating aquatic species density from environmental DNA. *Ecology and Evolution*, **8**, 3468–3477.

Conn, P.B., Johnson, D.S., Williams, P.J., Melin, S.R. & Hooten, M.B. (2018) A guide to Bayesian model checking for ecologists. *Ecological Monographs*, **88**, 526–542.

Deiner, K., Bik, H.M., Mächler, E., Seymour, M., Lacoursière-Roussel, A., Altermatt, F., Creer, S., Bista, I., Lodge, D.M., de Vere, N., Pfrender, M.E. & Bernatchez, L. (2017) Environmental DNA metabarcoding: transforming how we survey animal and plant communities. *Molecular Ecology*, **26**, 5872–5895.

10

Eichmiller, J.J., Bajer, P.G. & Sorensen, P.W. (2014) The relationship between the distribution of common carp and their environmental DNA in a small lake. *PloS ONE*, **9**, e112611.

Ficetola, G.F., Miaud, C., Pompanon, F. & Taberlet, P. (2008) Species detection using environmental DNA from water samples. *Biology Letters*, **4**, 423–425.

Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A. & Rubin, D.B. (2013) *Bayesian Data Analysis*. Chapman and Hall/CRC, 3rd edition.

Goldberg, C.S., Sepulveda, A., Ray, A., Baumgardt, J. & Waits, L.P. (2013) Environmental DNA as a new method for early detection of New Zealand mudsnails (*Potamopyrgus antipodarum*). *Freshwater Science*, **32**, 792–800.

Goldberg, C.S., Turner, C.R., Deiner, K., Klymus, K.E., Thomsen, P.F., Murphy, M.A., Spear, S.F., McKee, A., Oyler-McCance, S.J., Cornman, R.S., Laramie, M.B., Mahon, A.R., Lance, R.F., Pilliod, D.S., Strickler, K.M., Waits, L.P., Fremier, A.K., Takahara, T., Herder, J.E. & Taberlet, P. (2016) Critical considerations for the application of environmental DNA methods to detect aquatic species. *Methods in Ecology and Evolution*, **7**, 1299–1307.

Hansen, B.K., Bekkevold, D., Clausen, L.W. & Nielsen, E.E. (2018) The sceptical optimist: challenges and perspectives for the application of environmental DNA in marine fisheries. *Fish and Fisheries*, **19**, 751–768.

Hastie, T., Tibshirani, R. & Friedman, J. (2009) *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. Springer, 2nd edition.

Ito, Y., Yasuma, H., Masuda, R., Minami, K., Matsukura, R., Morioka, S. & Miyashita, K. (2011) Swimming angle and target strength of larval Japanese anchovy (*Engraulis japonicus*). *Fisheries Science*, **77**, 161–167.

Jo, T., Murakami, H., Masuda, R., Sakata, M.K., Yamamoto, S. & Minamoto, T. (2017) Rapid degradation of longer DNA fragments enables the improved estimation of distribution and biomass using environmental DNA. *Molecular Ecology Resources*, **17**, e25–e33.

Karabatsos, G. (2018) Marginal maximum likelihood estimation methods for the tuning parameters of ridge, power ridge, and generalized ridge regression. *Communications in Statistics-Simulation and Computation*, **47**, 1632–1651.

Kasai, A. & Yoon, S. (2018) Reproduction of environmental DNA using a numerical model. *Aquabiology*, **40**, 23–27. (in Japanese with English abstract)

Lacoursière-Roussel, A., Rosabal, M. & Bernatchez, L. (2016) Estimating fish abundance and biomass from eDNA concentrations: variability among capture methods and environmental conditions. *Molecular Ecology Resources*, **16**, 1401–1414.

Maruyama, A., Nakamura, K., Yamanaka, H., Kondoh, M. & Minamoto, T. (2014) The release rate of environmental DNA from juvenile and adult fish. *PLoS ONE*, **9**, e114639.

11

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

Masuda, R., Yamashita, Y. & Matsuyama, M. (2008) Jack mackerel *Trachurus japonicus* juveniles use jellyfish for predator avoidance and as a prey collector. *Fisheries Science*, **74**, 276–284.

McCullagh, P. & Nelder, J.A. (1989) *Generalized Linear Models*. Chapman and Hall/CRC, 2nd edition.

Mellor, G.L. (2002) *Users Guide for A Three-Dimensional, Primitive Equation, Numerical Ocean Model*. (Available at http://www.ccpo.odu.edu/POMWEB).

Miyashita, K., Tetsumura, K., Honda, S., Oshima, T., Kawabe, R. & Sasaki, K. (2004) Diel changes in vertical distribution patterns of zooplankton and walleye pollock (*Theragra chalcogramma*) off the Pacific coast of eastern Hokkaido, Japan, estimated by the volume back scattering strength ($S_V$) difference method. *Fisheries Oceanography*, **13**, 99–110.

Nakamura, T., Hamano, A., Abe, K., Yasuma, H. & Miyashita, K. (2013) Acoustic scattering properties of juvenile jack mackerel *Trachurus japonicus* based on a scattering model and *ex situ* target strength measurements. *Nippon Suisan Gakkaishi*, **79**, 345–354. (in Japanese with English abstract)

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.

Rees, H.C., Maddison, B.C., Middleditch, D.J., Patmore, J.R.M. & Gough, K.C. (2014) The detection of aquatic animal species using environmental DNA – a review of eDNA as a survey tool in ecology. *Journal of Applied Ecology*, **51**, 1450–1459.

Sansom, B.J. & Sassoubre, L.M. (2017) Environmental DNA (eDNA) shedding and decay rates to model freshwater mussel eDNA transport in a river. *Environmental Science & Technology*, **51**, 14244–14253.

Simmonds, J. & MacLennan, D.N. (2005) *Fisheries Acoustics: Theory and Practice*. John Wiley & Sons, 2nd edition.

Takahara, T., Minamoto, T., Yamanaka, H., Doi, H. & Kawabata, Z. (2012) Estimation of fish biomass using environmental DNA. *PloS ONE*, **7**, e35868.

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.

Thomsen, P.F. & Willerslev, E. (2015) Environmental DNA – an emerging tool in conservation for monitoring past and present biodiversity. *Biological Conservation*, **183**, 4–18.

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

Wilcox, T.M., McKelvey, K.S., Young, M.K., Sepulveda, A.J., Shepard, B.B., Jane, S.F., Whiteley, A.R., Lowe, W.H. & Schwartz, M.K. (2016) Understanding environmental DNA detection probabilities: a case study using a stream-dwelling char *Salvelinus fontinalis. Biological Conservation*, **194**, 209–216.

Williams, B.K., Nichols, J.D. & Conroy, M.J. (2002) *Analysis and Management of Animal Populations.* Academic Press.

Yamamoto, S., Minami, K., Fukaya, K., Takahashi, K., Sawada, H., Murakami, H., Tsuji, S., Hashizume, H., Kubonaga, S., Horiuchi, T., Hongo, M., Nishida, J., Okugawa, Y., Fujiwara, A., Fukuda, M., Hidaka, S., Suzuki, K.W., Miya, M., Araki, H., Yamanaka, H., Maruyama, A., Miyashita, K., Masuda, R., Minamoto, T. & 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**, e0149786.

Yamanaka, H., Minamoto, T., Matsuura, J., Sakurai, S., Tsuji, S., Motozawa, H., Hongo, M., Sogo, Y., Kakimi, N., Teramura, I., Sugita, M., Baba, M. & 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.

Yoon, S. & Kasai, A. (2017) Relative contributions of external forcing factors to circulation and hydrographic properties in a micro-tidal bay. *Estuarine, Coastal and Shelf Science*, **198**, 225–235.

# Acknowledgements

# Author contributions

M.K., R.M., A.K., K. Miyashita, and T.M. conceived and designed the eDNA survey. H.M. and S. Yamamoto conducted the molecular experiments. K. Minami and K. Miyashita analysed the echo sounder data. S. Yoon and A.K. developed the tracer model. K.F. and Y.O. designed the methodology and conducted data analyses. K.F., H.M., S. Yoon, and K. Minami led the writing of the manuscript. All the co-authors discussed the results and contributed critically to the manuscript.

13

# Competing interests

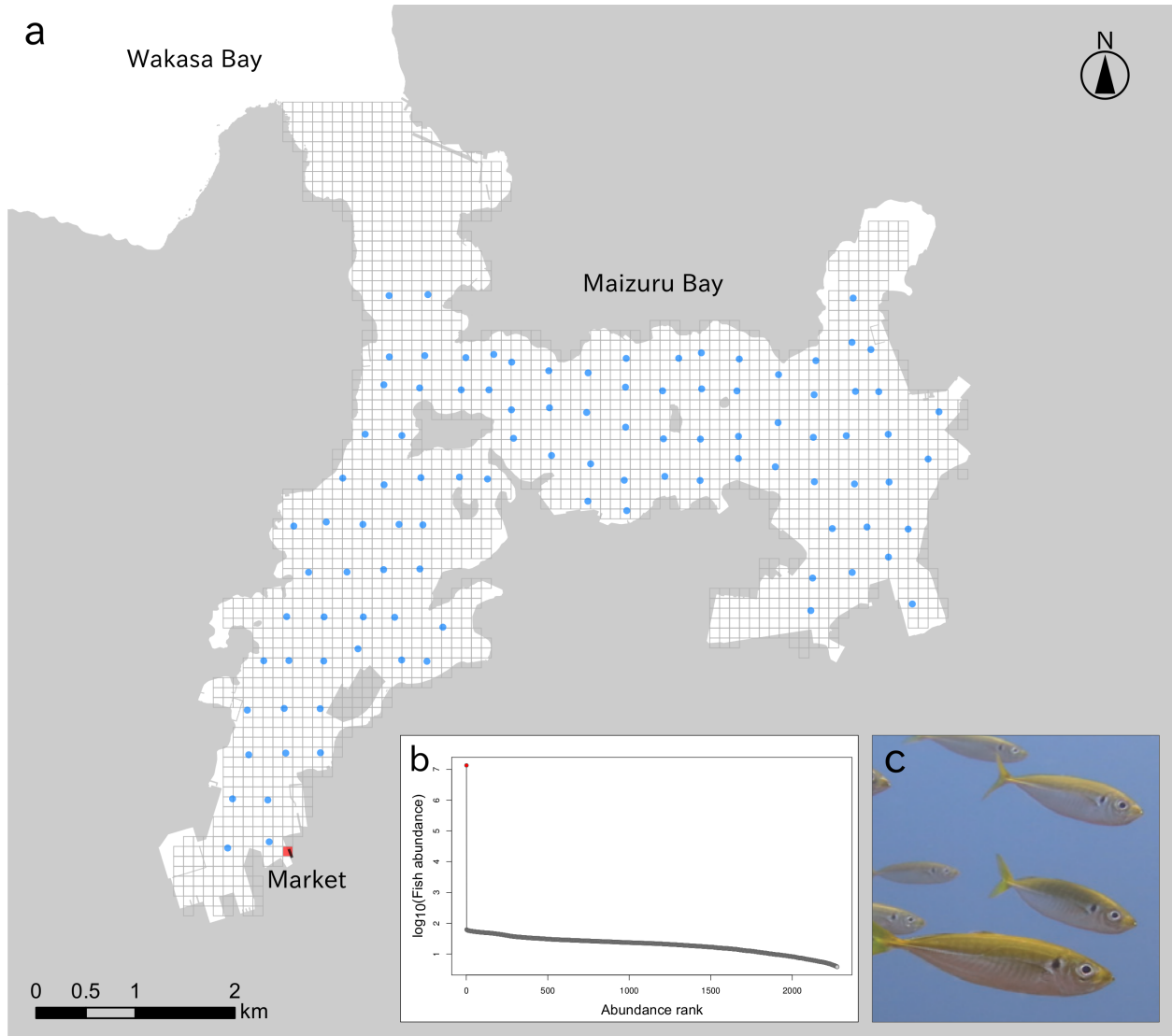The authors declare that they have no competing interests.

Fig. 1. **Maizuru bay, the study site.** a, The 2,276 horizontal lattice grids for the eDNA tracer model (grey boxes) and the 100 water-sampling stations (blue circles). The grid in which estimates of jack mackerel density were extremely high is highlighted in red. The building of the fish market, overlapping with the red lattice grid, is depicted by a filled black box. b, Fish abundance estimates in the 2,276 horizontal lattice grids. Abundance estimates in nine vertical cells were pooled for each grid. The lattice grid next to the market is highlighted in red. c, The Japanese jack mackerel (*T. japonicus*) in Maizuru bay (photo credit: R. Masuda).

# Supplementary information

## Body size of the Japanese jack mackerel

On 23rd June 2016, the day following the survey cruise, a trawl net (79 cm in diameter, 2.4 m long, 5 mm mesh size) was towed horizontally at four locations in Maizuru Bay where strong sonar signals were detected. The size of the collected jack mackerels ($n = 6$) was $35.3 \pm 2.5$ mm (mean $\pm$ SD) in standard length and $0.77 \pm 0.13$ g in body weight. Underwater observation was also conducted on the same day, where approximately 50 individuals of jack mackerel juveniles, of the size 20–30 mm in body length, were found during a 10 min observation period. They were all associating with sea nettles (*Chrysaora pacifica*) either singly or by forming a small group. Given that the study was conducted during the peak season of jack mackerel recruitment from an offshore pelagic zone to a coastal shallow reef habitat where the jack mackerel population in the bay is dominated by new recruits, we supposed that the size of the jack mackerel in the population can effectively be represented by $\sim$30 mm in body length and 1 g in body weight.

## Derivation of the eDNA shedding rate of the Japanese jack mackerel

The eDNA shedding rate of the jack mackerel was derived mathematically from the results of tank experiments conducted by Maruyama *et al.* (2014) and Jo *et al.* (2017).

In the study by Jo *et al.* (2017), three adult Japanese jack mackerels ca. 15 cm in total length and ca. 40 g in body weight, on average, had been kept in three 200 L tanks. Filtered seawater was injected into the tanks as inlet water at a rate of 0.9 L min$^{-1}$. Then, the eDNA concentration in the rearing water ($c$) can be expressed as

$$\frac{dc}{dt} = -(\lambda + \alpha)c + \beta' x \tag{7}$$

where $\lambda$ is a degradation rate of eDNA, which had been identified in the experiment as 0.044 h$^{-1}$ (Jo *et al.* 2017), $\alpha$ is the exponential decay constant due to water injection (0.54 h$^{-1}$), $\beta'$ means the eDNA shedding rate of the adult jack mackerels, and $x$ denotes the fish density in the rearing tank (0.015 individuals per litre).

We assume that the eDNA concentration had reached an equilibrium in experiments by Jo *et al.* (2017), and had been determined as $c_0 = 25365$ copies per litre of seawater (Jo *et al.* 2017). The eDNA shedding rate of juvenile Japanese jack mackerels ($\beta$) is then estimated as

$$\beta = \frac{\beta'}{10} = \frac{c_0(\lambda + \alpha)}{10x} = 9.88 \times 10^4 \text{ copies individual}^{-1} \text{ h}^{-1}, \tag{8}$$

where we assumed that the eDNA shedding rate per fish body weight is four-fold higher in the juvenile fish than in the adult fish (Maruyama *et al.* 2014); this finding indicates that the eDNA shedding rate of adult individuals of size $\sim 15$ cm and weight $\sim 40$ g ($\beta'$) was 10-fold greater than that of jack mackerel juveniles of size $\sim 3$ cm and weight $\sim 1$ g ($\beta$).

16