

# Estimation accuracy of species abundance based on environmental DNA with relation to its production source and state suggested by meta-analyses

Toshiaki Jo<sup>1</sup> and Hiroki Yamanaka<sup>2</sup>

<sup>1</sup>Ryukoku Daigaku

<sup>2</sup>Ryukoku University

March 07, 2024

## Abstract

Environmental DNA (eDNA) analysis is a promising tool for non-disruptive and cost-efficient estimation of species abundance. However, its practical applicability in natural environments is limited owing to a potential gap between eDNA concentration and species abundance in the field. Although the importance of accounting for eDNA dynamics, such as transport and degradation, has been discussed, the influence of eDNA characteristics, including production source and cellular/molecular state, on the accuracy of eDNA-based abundance estimation was entirely overlooked. We conducted meta-analyses using 44 of previous eDNA studies and investigated the relationships between the accuracy ( $R^2$ ) of eDNA-based abundance estimation and eDNA characteristics. First, we found that estimated  $R^2$  values were significantly lower for crustaceans and mussels than fish. This finding suggests that less frequent eDNA production of these taxa owing to their external morphology and physiology may impede accurate estimation of their abundance *via* eDNA. Moreover, linear mixed modeling showed that, despite high variances,  $R^2$  values were positively correlated with filter pore size, indicating that selective collection of larger-sized eDNA, which is typically fresher, could improve the estimation accuracy of species abundance. Although our collected dataset was somewhat biased to the studies targeting specific taxa, our findings shed a new light on the importance of what characteristics of eDNA should be targeted for more accurate estimation of species abundance. Further empirical studies are required to validate our findings and fully elucidate the relationship between eDNA characteristics and eDNA-based abundance estimation.

## Title:

Estimation accuracy of species abundance based on environmental DNA with relation to its production source and state suggested by meta-analyses

## Authors:

Toshiaki Jo<sup>1,2</sup>, Hiroki Yamanaka<sup>1,2</sup>

## Affiliations:

<sup>1</sup>Faculty of Advanced Science and Technology, Ryukoku University: 1-5, Yokotani, Oe-cho, Seta, Otsu City, Shiga 520-2194, Japan

<sup>2</sup>Ryukoku Center for Biodiversity Science: 1-5, Yokotani, Oe-cho, Seta, Otsu City, Shiga 520-2194, Japan

## Corresponding author:

Toshiaki Jo

Faculty of Advanced Science and Technology, Ryukoku University: 1-5, Yokotani, Oe-cho, Seta, Otsu City, Shiga 520-2194, Japan

Tel: +81-77-544-7113

Email: [toshiakijo@gmail.com](mailto:toshiakijo@gmail.com)

## Running head:

Synthesizing eDNA-based abundance estimation

## Keywords:

abundance; environmental DNA (eDNA); production; meta-analysis;  $R^2$ ; state

## Abstract (< 250 words)

Environmental DNA (eDNA) analysis is a promising tool for non-disruptive and cost-efficient estimation of species abundance. However, its practical applicability in natural environments is limited owing to a potential gap between eDNA concentration and species abundance in the field. Although the importance of accounting for eDNA dynamics, such as transport and degradation, has been discussed, the influence of eDNA characteristics, including production source and cellular/molecular state, on the accuracy of eDNA-based abundance estimation was entirely overlooked. We conducted meta-analyses using 44 of previous eDNA studies and investigated the relationships between the accuracy ( $R^2$ ) of eDNA-based abundance estimation and eDNA characteristics. First, we found that estimated  $R^2$  values were significantly lower for crustaceans and mussels than fish. This finding suggests that less frequent eDNA production of these taxa owing to their external morphology and physiology may impede accurate estimation of their abundance *via* eDNA. Moreover, linear mixed modeling showed that, despite high variances,  $R^2$  values were positively correlated with filter pore size, indicating that selective collection of larger-sized eDNA, which is typically fresher, could improve the estimation accuracy of species abundance. Although our collected dataset was somewhat biased to the studies targeting specific taxa, our findings shed a new light on the importance of what characteristics of eDNA should be targeted for more accurate estimation of species abundance. Further empirical studies are required to validate our findings and fully elucidate the relationship between eDNA characteristics and eDNA-based abundance estimation.

## Introduction

In the past decades, environmental DNA (eDNA) analysis has been developed and remarkably applied in multiple fields of ecology, fisheries, and environmental science (Ficetola et al., 2008; Bálint et al., 2018; Ruppert et al., 2019; Spear et al., 2021). Environmental DNA is defined as a total pool of DNA isolated from environmental samples such as water and sediment (Pawlowski et al., 2020); in a narrower sense, it is generally defined as an extra-organismal DNA released from macro-organisms as a form of feces, skin, mucus, and gamete (Barnes & Turner, 2016; Rodriguez-Ezpeleta et al., in press). Contrary to traditional methods, PCR-based detection of target eDNA does not require capturing nor observing individuals, and thus eDNA analysis is a feasible approach for non-disruptive, highly-sensitive, and cost-effective biomonitoring (Takahara et al., 2013; Yamanaka & Minamoto, 2016; Deiner et al., 2017; Djurhuus et al., 2020). Therefore, eDNA analysis has a potential to improve the monitoring of biodiversity and ecosystem, allowing for more effective conservation and management of biodiversity and resources.

In addition to species presence/absence, eDNA analysis can be used to predict species abundance from target eDNA concentrations. Several studies have reported positive correlations between eDNA concentrations and species abundance for various taxa and environments (Takahara et al., 2012; Pilliod et al., 2013; Klymus et al., 2015; Salter et al., 2019). However, a recent meta-analysis demonstrated that the correlation between eDNA concentration and species abundance was weaker in natural environments than in controlled laboratory conditions (i.e., aquaria, tanks, or mesocosms) (Yates et al., 2019). According to the study, the mean  $R^2$  values were 81 % and 57 % in laboratory conditions and natural environments, respectively. This finding is intuitively unsurprising given that abundance can be precisely set in laboratory experiments, but we

cannot know ‘true’ species abundance in natural environments where some individuals are not analyzable depending on their developmental stage and/or the survey method (Yates et al., 2019). In addition, the effects of diffusion and degradation on eDNA detection/quantification would be more substantial in natural environments due to compounding and complicated environmental conditions, including temperature, water chemistry, flow rate, and substrate (Strickler et al., 2015; Jane et al., 2015; Shogren et al., 2018; Jo et al., 2019a). Such factors could hamper the practical application of eDNA-based abundance estimation in natural environments (Hansen et al., 2018). Therefore, toward an effective conservation management of biodiversity and precise stock assessment *via* eDNA analysis, it is important to value the factors affecting such variabilities with regard to the estimation accuracy, and improve the accuracy of eDNA-based abundance estimation.

The amount of eDNA in a field is determined by a function of its production, transport, and degradation (Strickler et al., 2015; Barnes & Turner, 2016). Thus, in addition to processes of eDNA transport and degradation, the relationships between eDNA concentration and species abundance may also be affected by target eDNA characteristics, including its production source and cellular/molecular state. For example, eDNA production sources and processes may differ among taxa, which could accordingly influence the estimation accuracy of species abundance *via* eDNA analysis, as well as detection sensitivity of target eDNA. Andruszkiewicz et al. (2021) estimated eDNA shedding rates (pg/hour) of multiple taxa under similar experimental conditions and found that crustaceans (*Palaeomenes* spp.) had lower shedding rates than fish (*Fundulus heteroclitus*) and scyphomedusae (*Aurelia aurita* and *Chrysaora* spp.). These findings suggest that external morphology and/or physiology could substantially associate the difference in eDNA production sources and processes among taxa.

Cellular and molecular states of eDNA can also associate with its transport and degradation processes closely, consequently influencing the spatiotemporal range of target eDNA signals and even eDNA-based estimation accuracy of species abundance. Although studies linking eDNA state to its spatiotemporal dynamics are scarce, it has been reported that larger-sized and intra-cellular eDNA contained longer DNA fragments more frequently (Jo et al., 2020a), and eDNA decay rates could be determined by eDNA states, such as target gene (mitochondrial/nuclear) and particle size, as well as abiotic factors, including temperature and water chemistry (Jo & Minamoto, 2021). In the context of abundance estimation, given the rapid degradation of longer eDNA fragments (Jo et al., 2017) and persistence of smaller-sized eDNA particles (i.e., eDNA from smaller size fractions) in water due to the inflow of degraded eDNA from larger to smaller fractions (Jo et al., 2019b), biological signals from longer eDNA fragments and larger eDNA particles (i.e., eDNA from larger size fractions) could be fresher and more spatiotemporally finer in the field, which may consequently improve the accuracy of eDNA-based abundance estimation. Nevertheless, aside from Stewart (2019), who reviewed how biotic factors, such as developmental stage, life history, and species interaction might influence eDNA production and eDNA-based abundance estimation performance, exploration of the effects of eDNA production sources and states on estimation accuracy has been limited.

As far as we know, there is no study to directly value the importance of eDNA production source and state for the accuracy of eDNA-based abundance estimation. However, meta-analyses, synthesizing previous findings and statistically re-analyzing them, may shed a light on the relationship between eDNA-based estimation of species abundance and such eDNA characteristics. In this study, we investigated how different eDNA production sources and states influenced eDNA-based species abundance estimation accuracy by performing meta-analyses of eDNA studies targeting macro-organisms. We conducted a literature search and extracted data on factors influencing eDNA production sources and states. Moreover, since it is unclear how the relationship between species abundance and eDNA concentration differs among various natural environments (e.g., freshwater/marine, lentic/lotic), we also assessed the effect of target environments on eDNA-based abundance estimation accuracy. Integrating and collating previous findings *via* meta-analyses will enable us to elucidate the relationships between species abundance estimation and eDNA characteristics, which would not be recognized in individual eDNA studies.

## Materials and methods

### *Literature search and data extraction*

Using Google Scholar (<https://scholar.google.co.jp/>), we conducted a literature search relating to eDNA production and species abundance published during 2008 to 2021. The literature search was 1 June, 2021). The terms “eDNA” or “environmental DNA”, included in the title and/or text, were targeted for the literature search. We then filtered and selected papers that (i) targeted eDNA from macro-organisms (not from microbes such as bacteria and fungi, or viruses), (ii) were published in international journals, (iii) were peer-reviewed (not preprints), and (iv) described the relationships between eDNA concentration, quantified by real-time PCR or digital PCR, and species abundance (biomass and/or density) by fitting linear or linear-mixed models. For criterion (iv), most studies reported positive relationships with statistical significances, while some of results were not significant.

We then extracted  $R^2$  values from linear fitting, indices of species abundance estimation accuracy based on eDNA concentration, target taxa, filter pore sizes used for water filtration ( $\mu\text{m}$ ), and PCR amplicon sizes (base pair; bp) from the selected studies. Taxa were classified as fish, amphibians, other vertebrates, crustaceans, mussels, and other invertebrates. In studies involving aqueous eDNA collection *via* precipitation or centrifugation, the filter pore size was regarded as 0  $\mu\text{m}$ . We did not include Pearson’s correlation coefficients in our meta-analyses because the index was not available in all collected literatures. If the manuscripts only reported Pearson’s correlation coefficients, we squared the coefficients and substituted them for  $R^2$  values. Different  $R^2$  values based on different experimental conditions within the same study (e.g., species, filter type, and amplicon size) were treated separately. Abundance metrics (biomass/density) were pooled here because of its mere effect on correlation between eDNA concentration and species abundance (Yates et al., 2019). In addition, we extracted information on the study environments and classified them as laboratory, lentic freshwater, lotic freshwater, and marine environments. Moreover, we calculated the sample size (the number of water samples or sampling sites) required for fitting each linear model based on figures and/or text in the corresponding literature.

### *Statistical analyses*

All the statistical analyses were performed by using R version 4.0.4 (R Core Team, 2021). Among the collected literature, we excluded some of them as a matter of convenience in the further analysis (Appendix S1). To test the categorical effects (taxa and environment) on estimation accuracy of species abundance, we estimated the effect sizes and their variances by Fisher’s z-transformation, which allows to avoid ‘Simpson’s paradox’ and compare  $R^2$  values among categories accounting for sample sizes. Using the *metacor* function in the package ‘meta’ (Balduzzi et al., 2019), we produced forest plots to integrate  $R^2$  values with their 95% confidence intervals (CIs) among multiple individual datasets, where weighted-average  $R^2$  values were calculated by using the inverses of variances estimated above, using the *rma* function in the package ‘metafor’ (Viechtbauer, 2010). We adopted the random-effect model, assuming that all datasets share a common effect size but also vary among datasets, as our collected dataset is not functionally identical (i.e., some  $R^2$  values could be derived from the same study). More strictly, the weight (inverse variance) for averaging  $R^2$  values comprised both intra- and inter-study variances of effect size (Borenstein et al., 2010).

In addition, to test the quantitative effects (filter pore size and PCR amplicon size) on estimation accuracy of species abundance, we performed a linear mixed modeling (LMM) by the *lmer* function in the package ‘lmerTest’ (Kuznetsova et al., 2017). Prior to model fitting, we logit-converted  $R^2$  values to meet normality and included the values as the dependent variable. Filter pore size ( $\mu\text{m}$ ) and amplicon size (bp) were explanatory variables. As random effects, we included study groups, abundance metrics (biomass/density), target taxa, and environment, assuming that the effects of filter pore size and amplicon size could be underestimated without consideration of these categorical variables. Because almost all studies included in the model targeted shorter eDNA fragments (<200 bp), we excluded the dataset targeting 719-bp fragments of mitochondrial eDNA (Jo et al., 2017) as an outlier.

## **Results**

We selected and analyzed 44 published papers in total, from which we extracted 100  $R^2$  values ranging from 0.02 to 0.99 (Tables 1 and S1). All the collected studies targeted mitochondrial DNA fragments, and no study

reported the positive correlation between nuclear eDNA concentration and species abundance. Most studies reported the correlation between eDNA concentration and species abundance, targeting fish species (30 of 44 studies). The filter pore size and amplicon size ranged from 0 to 10  $\mu\text{m}$  and 66 to 719 bp, respectively. The majority of studies used filters with less than 3  $\mu\text{m}$  pore size and amplified less than 200 bp target DNA fragments. The number of studies conducted in laboratory, lentic freshwater, lotic freshwater, and marine environments were 15, 14, 14, and 4, respectively.

The effect sizes of  $R^2$  values and their variances were estimated using Fisher's z-transformation and forest plots (Figure S1). Relative to fish ( $R^2 = 0.68$  [95 % CI: 0.61, 0.74]), estimated  $R^2$  values were significantly lower for crustaceans (0.22 [-0.04, 0.46]) and mussels (0.50 [0.44, 0.55]); among the target taxa, the correlation between crustacean eDNA concentration and abundance was the weakest (Figure 1). Laboratory experiments (0.78 [0.69, 0.84]) produced higher estimated  $R^2$  values than natural environments, and there was no substantial difference in the estimation accuracy among natural environment types (Figure 2). In addition, the LMM showed a significant positive effect of filter pore size on  $R^2$  values ( $P < 0.05$ ; Table 2 & Figure 3).  $R^2$  values tended to be higher as filter pore size increased, but some datasets also reported high  $R^2$  values with smaller filter pore sizes. PCR amplicon size did not significantly influence  $R^2$  values ( $P > 0.1$ ; Table 2 & Figure S2).

## Discussion

### *Estimation accuracy and eDNA production source*

Since the first study targeting common carp (*Cyprinus carpio*) (Takahara et al., 2012), most research investigating the correlation between eDNA concentration and species abundance has targeted fish species (e.g., Klymus et al., 2015; Jo et al., 2017; Capo et al., 2021; Table 1). According to these studies, fish eDNA concentration predicts relative species abundance with relatively high accuracy regardless of environmental conditions, which was also confirmed in our meta-analysis. However, the meta-analysis showed that such tendencies might not necessarily be similar among target taxa, and the abundance of some taxa may not be accurately estimated *via* current eDNA applications.

We found that eDNA-based estimation accuracy of species abundance ( $R^2$ ) was significantly lower for crustaceans and mussels than fish. A previous study hypothesized that species morphology and/or behavior might affect eDNA production, reporting lower eDNA shedding rates in grass shrimp than in fish and jellyfish (Andruszkiewicz et al., 2021). Fish and jellyfish are likely to constantly produce eDNA as epidermis and/or muco-substances (Merkes et al., 2014; Sassoubre et al., 2016). In contrast, crustaceans are characterized by their hard exoskeletons and segmented bodies plans (Hadley, 1986) and are thus unlikely to shed large amounts of eDNA from their body surfaces unless they are molting. Similarly, mussel soft tissue is covered with a hard, calcified shell that is less likely to shed eDNA (Sansom & Sassoubre, 2017). Consequently, crustaceans and mussels infrequently and irregularly shed eDNA, which may impede sufficient eDNA detection in the field and prevent accurate abundance estimation *via* eDNA analysis (Dougherty et al., 2016; Machler et al., 2016; Johnsen et al., 2020).

The mean correlation between amphibian eDNA concentration and abundance was similar to that of fish, which is reasonable given amphibians likely shed eDNA constantly from their epidermis and/or mucus. However, the variation (95 % CI) was much larger for amphibians relative to fish. The discrepancy could simply be explained by biases derived from the smaller number of corresponding studies or different experimental conditions. Among the studies collected for our meta-analysis, most of the correlations between amphibian eDNA concentration and abundance were studied in natural environments (e.g., Thomsen et al., 2012; Pilliod et al., 2013). This is likely because of the difficulty in conducting controlled laboratory experiments using amphibians due to its rarity, unless some invasive species such as American bullfrog (*Lithobates catesbeianus*). In addition, possibly for the same reason, many of the collected studies depended on visual counts for abundance estimation, which could ambiguate correlations between amphibian eDNA concentration and abundance. On the other hand, Everts et al. (2021) assessed the correlation between eDNA concentration and abundance of American bullfrog tadpoles and juveniles using mesocosm experiments, and reported rel-

atively high  $R^2$  values (0.64 to 0.99). Thus, amphibians are potentially suitable for accurate estimation of species abundance *via* eDNA analysis. Similarly, given that the study targeting mussels was not conducted under laboratory conditions (Currier et al., 2018), the relatively low  $R^2$  values for mussels may partly include such biases. Accumulating studies targeting various taxa in laboratory conditions and natural environments could help us understand the effects of ecological characteristics (morphology, physiology, and ethology) on the process of eDNA production, and may provide us with keys for improved approach of eDNA-based abundance estimation.

#### *Estimation accuracy and eDNA state*

Our statistical modeling showed that the use of larger pore size filters could improve the accuracy of eDNA-based abundance estimation. According to previous studies, the cellular/molecular structure of larger eDNA particles derived from intra-cellular DNA, such as cells and tissues, is degraded into smaller eDNA particles over time (Jo et al., 2019b). Apparent persistence of such smaller-sized eDNA could thus be longer than that of larger-sized eDNA, and larger eDNA particles collected using larger pore size filters are more likely to be recently released and less degraded than smaller eDNA particles. Such larger-sized ‘fresher’ eDNA is expected to reflect species presence and abundance at a spatiotemporally finer scale, consequently improving the accuracy of eDNA-based abundance estimation. Although individual eDNA studies could not clearly infer the relationship between accuracy and filter pore size (Takahara et al., 2012; Eichmiller et al., 2016), our meta-analysis supports the applicability of larger pore size filters for improved abundance estimation *via* eDNA analysis for the first time.

In contrast, some datasets reported high  $R^2$  values using smaller pore size filters (Figure 3), which can collect both larger-sized fresher eDNA and smaller-sized older eDNA. Thus, these studies may have collected a higher proportion of larger-sized eDNA while using smaller pore size filters; in particular, laboratory experiments with excessively high abundances (e.g., Takahara et al., 2012; Doi et al., 2015) may have collected large quantities of large-sized fresh eDNA. Although studies using larger pore size filters (especially  $>3\ \mu\text{m}$ ) were limited in our dataset, further empirical studies targeting larger-sized eDNA particles would conceivably contribute to the robustness of our results and potentially provide a new approach to improve the accuracy of eDNA-based abundance estimation in the field.

Moreover, the particle size distribution of target eDNA will dictate whether the selective collection of larger-sized eDNA using a larger pore size filter is effective. eDNA particles from fish are generally concentrated at 1 to 10  $\mu\text{m}$  size fractions (Turner et al., 2014; Jo et al., 2019b), whereas eDNA particle size distributions for other taxa remain largely unknown except for Moushomi et al. (2019) targeting *Daphnia magna*. Estimating eDNA particle size distribution provides information on its cellular/molecular state in water (Jo et al., 2019b) and helps determine a suitable filter pore size for its efficient collection (Turner et al., 2014). In particular, research on the particle size distributions of eDNA from crustaceans and mussels, which show weaker correlations between eDNA concentration and abundance, would support the applicability of larger pore size filters for improving the accuracy of their eDNA-based abundance estimation.

In contrast, PCR amplicon size (i.e., DNA fragment length of target eDNA) was not significantly correlated with  $R^2$  values. However, almost all the studies in our meta-analysis targeted shorter eDNA fragments ( $<200$  bp); thus, the effect of PCR amplicon size on the accuracy of eDNA-based abundance estimation may be underestimated. Owing to higher decay rates, detection of longer eDNA fragments may mitigate the effect of degraded eDNA and improve species abundance estimation accuracy (Jo et al., 2017). Nonetheless, this study was conducted in a situation where false-positive inferences of target individuals could be obvious (i.e., the effect of fish markets and dead individuals); thus, the performance of longer eDNA fragments for species abundance estimation in more ‘ordinal’ situations is unknown. Future empirical studies targeting longer eDNA fragments ( $>300$  bp) are needed to elucidate the importance of PCR amplicon size on eDNA-based abundance estimation.

#### *Estimation accuracy and environments*

Higher  $R^2$  values were reported for laboratory conditions than natural environments, which is consistent with

previous findings (Yates et al., 2019) and supports the methodological validity of our statistical analyses. As discussed in Yates et al. (2019), myriad and complex dynamics of eDNA could hamper the prediction of species abundance based on eDNA concentration in natural than controlled environments. However, we observed little difference in estimation accuracy among natural environments. Relative to lentic freshwater environments, such as ponds and lakes, eDNA diffusion and degradation would be substantial in riverine, coastal, and marine environments due to flow and tidal effects. Such factors can transport eDNA very long distances (Deiner & Altermatt, 2014; Andruszkiewicz et al., 2019) while also resulting in rapid dilution (Baker et al., 2018). In contrast, although eDNA may diffuse less in lentic environments, residual eDNA may cause false-positive inferences of species presence and inaccurate estimation of species abundance. It is noted that our results must be cautiously interpreted given the bias in the number of studies among natural environment types, but eDNA-based abundance estimation accuracy might not necessarily be worsened for lotic freshwater and marine environments relative to lentic environments.

Additionally, the spatial location of eDNA sampling within the study environments might also influence the results. Marine eDNA studies generally collected water samples from both near-shore and offshore sites, whereas some lentic freshwater studies relied solely on water sampling from shore sites (e.g., Kamoroff & Goldberg 2018; Kakuda et al., 2019). Even if eDNA detection rates were higher at shore sites (Hänfling et al., 2016; Jo et al., 2021), the lowered diffusion in lentic environments and highly heterogeneous distribution (Hänfling et al., 2016; Li et al., 2019) may not accurately reflect the overall species abundance in studied sites, especially when large lakes and reservoirs were targeted. Further studies are required to clarify how the heterogeneity of eDNA distribution influences the detection sensitivity, yield, and relationship to species abundance.

### *Conclusions and perspectives*

To our knowledge, the present meta-analysis is the first to report variations in eDNA-based estimation accuracy of species abundance among different target taxa and filter pore sizes (reflecting eDNA particle size distribution). Some recent studies have suggested the possibility improving the accuracy of eDNA-based abundance estimation by statistically accounting for the processes of eDNA production, transport, and degradation (Carraro et al., 2018; Cerco et al., 2018; Fukaya et al., in press). In contrast, our meta-analyses shed a new light on the importance of what characteristics of eDNA should be targeted for more accurate estimation of species abundance. In particular, our findings on the effects of eDNA state imply that ‘more recently released’ eDNA, existing as larger eDNA particles and potentially longer eDNA fragments, more precisely reflect species abundance in the field. This knowledge will complement abundance estimation approaches that consider eDNA spatiotemporal dynamics; that is, understanding eDNA characteristics, including production source, particle size, and fragment length, as well as eDNA production, transport, and degradation processes, will enable us to further enhance the potential of eDNA analysis as a non-disruptive and cost-efficient tool for species abundance estimation. Therefore, accumulating knowledge of eDNA states and their interactions with the dynamics is crucial (Jo & Minamoto, 2021), which could facilitate the development of a novel eDNA marker suitable for accurate eDNA-based estimation of species abundance.

There are some potential biases and limitations in our meta-analyses. First, our collected dataset was concentrated toward studies targeting fish species, which might cause biased and over-dispersed estimation for other taxa. Second, our meta-analyses excluded some eDNA studies because these studies were regarded to be inadequate for the methodology of our analyses (see above) or did not directly estimate the indices of abundance estimation accuracy (Pearson’s correlation coefficients or  $R^2$  values; e.g., Jo et al., 2020b). Accumulating additional empirical studies for various taxa and environmental conditions are necessary to validate the findings of our meta-analyses and further elucidate the influence of eDNA characteristics on eDNA-based estimation of species abundances.

Furthermore, although not considered in the present study, the applicability of nuclear eDNA, particularly targeting multiple copies of ribosomal RNA genes, should be noted for more accurate eDNA-based species abundance estimations. Relative to mitochondrial eDNA, targeting multi-copy nuclear eDNA can improve detectability and yield (Minamoto et al., 2017; Jo et al., 2020b) and nuclear eDNA may degrade more rapidly

due to potential differences in membrane and DNA structures (Bylemans et al., 2018; Jo et al., 2020b). In addition, nuclear eDNA production may also be less biased by individual growth and developmental stages, whereas mitochondrial eDNA production is expected to be suppressed with maturity and aging (Jo et al., 2020b). Understanding both the characteristics and dynamics of eDNA will fill a gap between eDNA concentration and species abundance in the field, and update current eDNA analysis as a more refined tool for biodiversity and ecosystem monitoring and stock assessment.

## Acknowledgements

This work was supported by the Environment Research and Technology Development Fund from the Ministry of the Environment, Japan (Grant Number JPMEERF20204004).

## Data accessibility

Detailed information on published studies reporting the relationship between eDNA concentration and species abundance can be found in the Supporting Information.

## Author contributions

T.J. and H.Y. conceived the study. T.J. performed the literature search, analyzed the data, and wrote the first draft of the manuscript. Both authors edited and provided feedback on the manuscript. There is no conflict of interest to declare.

## Supporting Information

Appendix S1. Additional information on prefiltering of the collected literature.

Table S1. A detailed list of all eDNA literatures selected and analyzed in this study.

Figure S1. Forest plots for integrating individual  $R^2$  values for each target taxa, genetic region, and environments.

Figure S2.  $R^2$  values with relation to PCR amplicon size.

## References

- Akamatsu, Y., Kume, G., Gotou, M., Kono, T., Fujii, T., Inui, R., & Kurita, Y. (2020). Using environmental DNA analyses to assess the occurrence and abundance of the endangered amphidromous fish *Plecoglossus altivelis ryukyuensis*. *Biodiversity Data Journal*, 8, e39679.
- Andruszkiewicz, E. A., Koseff, J. R., Fringer, O. B., Ouellette, N. T., Lowe, A. B., Edwards, C. A., & Boehm, A. B. (2019). Modeling environmental DNA transport in the coastal ocean using Lagrangian particle tracking. *Frontiers in Marine Science*, 6, 477.
- Andruszkiewicz, E. A., Zhang, W. G., Lavery, A. C., & Govindarajan, A. F. (2021). Environmental DNA shedding and decay rates from diverse animal forms and thermal regimes. *Environmental DNA*, 3(2), 492-514.
- Baker, C. S., Steel, D., Nieukirk, S., & Klinck, H. (2018). Environmental DNA (eDNA) from the wake of the whales: droplet digital PCR for detection and species identification. *Frontiers in Marine Science*, 5, 133.
- Balduzzi, S., Rücker, G., & Schwarzer, G. (2019). How to perform a meta-analysis with R: a practical tutorial. *Evidence-Based Mental Health*, 22(4), 153-160.
- Bálint, M., Pfenninger, M., Grossart, H. P., Taberlet, P., Vellend, M., Leibold, M. A., ... & Bowler, D. (2018). Environmental DNA time series in ecology. *Trends in Ecology & Evolution*, 33(12), 945-957.
- Barnes, M. A. & Turner, C. R. (2016). The ecology of environmental DNA and implications for conservation genetics. *Conservation Genetics*, 17(1), 1-17.
- Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods*, 1(2), 97-111.

- Brown, G. G., Gadaleta, G., Pepe, G., Saccone, C., & SbisA, E. (1986). Structural conservation and variation in the D-loop-containing region of vertebrate mitochondrial DNA. *Journal of Molecular Biology*, 192(3), 503-511.
- Bylemans, J., Furlan, E. M., Gleeson, D. M., Hardy, C. M., & Duncan, R. P. (2018). Does size matter? An experimental evaluation of the relative abundance and decay rates of aquatic environmental DNA. *Environmental Science & Technology*, 52(11), 6408-6416.
- Capo, E., Spong, G., Koizumi, S., Puts, I., Olajos, F., Konigsson, H., ... & Bystrom, P. (2021). Droplet digital PCR applied to environmental DNA, a promising method to estimate fish population abundance from humic-rich aquatic ecosystems. *Environmental DNA*, 3(2), 343-352.
- 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(46), 11724-11729.
- Cerco, C. F., Schultz, M. T., Noel, M. R., Skahill, B., & Kim, S. C. (2018). A fate and transport model for Asian carp environmental DNA in the Chicago area waterways system. *Journal of Great Lakes Research*, 44(4), 813-823.
- Currier, C. A., Morris, T. J., Wilson, C. C., & Freeland, J. R. (2018). Validation of environmental DNA (eDNA) as a detection tool for at-risk freshwater pearly mussel species (*Bivalvia: Unionidae*). *Aquatic Conservation: Marine and Freshwater Ecosystems*, 28(3), 545-558.
- Deiner, K., & Altermatt, F. (2014). Transport distance of invertebrate environmental DNA in a natural river. *PLoS ONE*, 9(2), e88786.
- Deiner, K., Bik, H. M., Machler, E., Seymour, M., Lacoursiere-Roussel, A., Altermatt, F., ... & Bernatchez, L. (2017). Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. *Molecular Ecology*, 26(21), 5872-5895.
- Djurhuus, A., Closek, C. J., Kelly, R. P., Pitz, K. J., Michisaki, R. P., Starks, H. A., ... & Breitbart, M. (2020). Environmental DNA reveals seasonal shifts and potential interactions in a marine community. *Nature Communications*, 11, 254.
- Doi, H., Uchii, K., Takahara, T., Matsushashi, S., Yamanaka, H., & Minamoto, T. (2015). Use of droplet digital PCR for estimation of fish abundance and biomass in environmental DNA surveys. *PLoS ONE*, 10(3), e0122763.
- Dougherty, M. M., Larson, E. R., Renshaw, M. A., Gantz, C. A., Egan, S. P., Erickson, D. M., & Lodge, D. M. (2016). Environmental DNA (eDNA) detects the invasive rusty crayfish *Orconectes rusticus* at low abundances. *Journal of Applied Ecology*, 53(3), 722-732.
- Eichmiller, J. J., Miller, L. M., & Sorensen, P. W. (2016). Optimizing techniques to capture and extract environmental DNA for detection and quantification of fish. *Molecular Ecology Resources*, 16(1), 56-68.
- Everts, T., Halfmaerten, D., Neyrinck, S., De Regge, N., Jacquemyn, H., & Brys, R. (2021). Accurate detection and quantification of seasonal abundance of American bullfrog (*Lithobates catesbeianus*) using ddPCR eDNA assays. *Scientific Reports*, 11, 11282.
- Ficetola, G. F., Miaud, C., Pompanon, F., & Taberlet, P. (2008). Species detection using environmental DNA from water samples. *Biology Letters*, 4(4), 423-425.
- Fukaya, K., Murakami, H., Yoon, S., Minami, K., Osada, Y., Yamamoto, S., ... & Kondoh, M. (in press). Estimating fish population abundance by integrating quantitative data on environmental DNA and hydrodynamic modelling. *Molecular Ecology*. DOI: 10.1111/mec.15530
- Hadley, N. F. (1986). The arthropod cuticle. *Scientific American*, 255(1), 104-113.

- Hanfing, B., Lawson Handley, L., Read, D. S., Hahn, C., Li, J., Nichols, P., ... & Winfield, I. J. (2016). Environmental DNA metabarcoding of lake fish communities reflects long-term data from established survey methods. *Molecular Ecology*, 25(13), 3101-3119.
- 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(5), 751-768.
- Jane, S. F., Wilcox, T. M., McKelvey, K. S., Young, M. K., Schwartz, M. K., Lowe, W. H., ... & Whiteley, A. R. (2015). Distance, flow and PCR inhibition: eDNA dynamics in two headwater streams. *Molecular Ecology Resources*, 15(1), 216-227.
- Jo, T., Arimoto, M., Murakami, H., Masuda, R., & Minamoto, T. (2019b). Particle size distribution of environmental DNA from the nuclei of marine fish. *Environmental Science & Technology*, 53(16), 9947-9956.
- Jo, T., Arimoto, M., Murakami, H., Masuda, R., & Minamoto, T. (2020b). Estimating shedding and decay rates of environmental nuclear DNA with relation to water temperature and biomass. *Environmental DNA*, 2(2), 140-151.
- Jo, T., Ikeda, S., Fukuoka, A., Inagawa, T., Okitsu, J., Katano, I., ... & Minamoto, T. (2021). Utility of environmental DNA analysis for effective monitoring of invasive fish species in reservoirs. *Ecosphere*, 12(6), e03643.
- Jo, T., & Minamoto, T. (2021). Complex interactions between environmental DNA (eDNA) state and water chemistries on eDNA persistence suggested by meta-analyses. *Molecular Ecology Resources*, 21(5), 1490-1503.
- Jo, T., Murakami, H., Masuda, R., & Minamoto, T. (2020a). Selective collection of long fragments of environmental DNA using larger pore size filter. *Science of the Total Environment*, 735, 139462.
- 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(6), e25-e33.
- Jo, T., Murakami, H., Yamamoto, S., Masuda, R., & Minamoto, T. (2019a). Effect of water temperature and fish biomass on environmental DNA shedding, degradation, and size distribution. *Ecology and Evolution*, 9(3), 1135-1146.
- Johnsen, S. I., Strand, D. A., Rusch, J. C., & Vralstad, T. (2020). Environmental DNA (eDNA) monitoring of noble crayfish *Astacus astacus* in lentic environments offers reliable presence-absence surveillance—but fails to predict population density. *Frontiers in Environmental Science*, 8, 612253.
- Kakuda, A., Doi, H., Souma, R., Nagano, M., Minamoto, T., & Katano, I. (2019). Environmental DNA detection and quantification of invasive red-eared sliders, *Trachemy scripta elegans*, in ponds and the influence of water quality. *PeerJ*, 7, e8155.
- Kamoroff, C., & Goldberg, C. S. (2018). Environmental DNA quantification in a spatial and temporal context: a case study examining the removal of brook trout from a high alpine basin. *Limnology*, 19(3), 335-342.
- Klymus, K. E., Richter, C. A., Chapman, D. C., & Paukert, C. (2015). Quantification of eDNA shedding rates from invasive bighead carp *Hypophthalmichthys nobilis* and silver carp *Hypophthalmichthys molitrix*. *Biological Conservation*, 183, 77-84.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: tests in linear mixed effects models. *Journal of Statistical Software*, 82(13), 1-26.
- Li, J., Lawson Handley, L. J., Harper, L. R., Brys, R., Watson, H. V., Di Muri, C., ... & Hanfing, B. (2019). Limited dispersion and quick degradation of environmental DNA in fish ponds inferred by metabarcoding.

Environmental DNA, 1(3), 238-250.

Machler, E., Deiner, K., Spahn, F., & Altermatt, F. (2016). Fishing in the water: effect of sampled water volume on environmental DNA-based detection of macroinvertebrates. *Environmental Science & Technology*, 50(1), 305-312.

Merkes, C. M., McCalla, S. G., Jensen, N. R., Gaikowski, M. P., & Amberg, J. J. (2014). Persistence of DNA in carcasses, slime and avian feces may affect interpretation of environmental DNA data. *PLoS ONE*, 9(11), e113346.

Minamoto, T., Uchii, K., Takahara, T., Kitayoshi, T., Tsuji, S., Yamanaka, H., & Doi, H. (2017). Nuclear internal transcribed spacer-1 as a sensitive genetic marker for environmental DNA studies in common carp *Cyprinus carpio*. *Molecular Ecology Resources*, 17(2), 324-333.

Pawlowski, J., Apothéloz-Perret-Gentil, L., & Altermatt, F. (2020). Environmental DNA: What's behind the term? Clarifying the terminology and recommendations for its future use in biomonitoring. *Molecular Ecology*, 29(22), 4258-4264.

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(8), 1123-1130.

R Core Team. (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.

Rodriguez-Ezpeleta, N., Morissette, O., Bean, C., Manu, S., Banerjee, P., Lacoursiere-Roussel, A., ... & Deiner, K. (in press). Trade-offs between reducing complex terminology and producing accurate interpretations from environmental DNA: Comment on "Environmental DNA: What's behind the term?" by Pawlowski et al., (2020). *Molecular Ecology*. DOI: 10.1111/mec.15942

Ruppert, K. M., Kline, R. J., & Rahman, M. S. (2019). Past, present, and future perspectives of environmental DNA (eDNA) metabarcoding: A systematic review in methods, monitoring, and applications of global eDNA. *Global Ecology and Conservation*, 17, e00547.

Salter, I., Joensen, M., Kristiansen, R., Steingrund, P., & Vestergaard, P. (2019). Environmental DNA concentrations are correlated with regional biomass of Atlantic cod in oceanic waters. *Communications Biology*, 2, 461.

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(24), 14244-14253.

Sassoubre, L. M., Yamahara, K. M., Gardner, L. D., Block, B. A., & Boehm, A. B. (2016). Quantification of environmental DNA (eDNA) shedding and decay rates for three marine fish. *Environmental Science & Technology*, 50(19), 10456-10464.

Shogren, A. J., Tank, J. L., Egan, S. P., August, O., Rosi, E. J., Hanrahan, B. R., ... & Bolster, D. (2018). Water flow and biofilm cover influence environmental DNA detection in recirculating streams. *Environmental Science & Technology*, 52(15), 8530-8537.

Spear, M. J., Embke, H. S., Krysan, P. J., & Vander Zanden, M. J. (2021). Application of eDNA as a tool for assessing fish population abundance. *Environmental DNA*, 3(1), 83-91.

Stewart, K. A. (2019). Understanding the effects of biotic and abiotic factors on sources of aquatic environmental DNA. *Biodiversity and Conservation*, 28(5), 983-1001.

Strickler, K. M., Fremier, A. K., & Goldberg, C. S. (2015). Quantifying effects of UV-B, temperature, and pH on eDNA degradation in aquatic microcosms. *Biological Conservation*, 183, 85-92.

Takahara, T., Minamoto, T., & Doi, H. (2013). Using environmental DNA to estimate the distribution of an invasive fish species in ponds. *PLoS ONE*, 8(2), e56584.

Takahara, T., Minamoto, T., Yamanaka, H., Doi, H., & Kawabata, Z. (2012). Estimation of fish biomass using environmental DNA. *PLoS ONE*, 7(4), e35868.

Thomsen, P. F., Kielgast, J. O. S., Iversen, L. L., Wiuf, C., Rasmussen, M., Gilbert, M. T. P., ... & Willerslev, E. (2012). Monitoring endangered freshwater biodiversity using environmental DNA. *Molecular Ecology*, 21(11), 2565-2573.

Turner, C. R., Barnes, M. A., Xu, C. C., Jones, S. E., Jerde, C. L., & Lodge, D. M. (2014). Particle size distribution and optimal capture of aqueous microbial eDNA. *Methods in Ecology and Evolution*, 5(7), 676-684.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36(3), 1-48.

Wu, Q., Kawano, K., Uehara, Y., Okuda, N., Hongo, M., Tsuji, S., ... & Minamoto, T. (2018). Environmental DNA reveals nonmigratory individuals of *Palaemon paucidens* overwintering in Lake Biwa shallow waters. *Freshwater Science*, 37(2), 307-314.

Yamanaka, H., & Minamoto, T. (2016). The use of environmental DNA of fishes as an efficient method of determining habitat connectivity. *Ecological Indicators*, 62, 147-153.

Yates, M. C., Fraser, D. J., & Derry, A. M. (2019). Meta-analysis supports further refinement of eDNA for monitoring aquatic species-specific abundance in nature. *Environmental DNA*, 1(1), 5-13.

Yates, M. C., Wilcox, T. M., McKelvey, K. S., Young, M. K., Schwartz, M. K., & Derry, A. M. (2021). Allometric scaling of eDNA production in stream-dwelling brook trout (*Salvelinus fontinalis*) inferred from population size structure. *Environmental DNA*, 3(3), 553-560.

## Tables

Table 1. Published literature on the relationship between eDNA concentration and species biomass/abundance analyzed in this study.

Study	# R <sup>2</sup> values	Sample size	Filter pore size [μm]	Amplicon size [bp]	Environment	Target taxa
Takahara et al. (2012)	3	6 to 12	0 to 3	78	Lab / Fresh_lentic	Fish
Thomsen et al. (2012)	2	9 to 10	0	72 to 81	Fresh_lentic	Amphibian
Pilliod et al. (2013)	4	10 to 13	0.45	78 to 85	Fresh_lentic	Amphibian
Doi et al. (2015)	4	36	0	78	Lab	Fish
Klymus et al. (2015)	2	24 to 25	0	108 to 190	Lab	Fish
Dougherty et al. (2016)	1	12	1.2	128	Fresh_lentic	Crustacean
Eichmiller et al. (2016)	4	9	0.2 to 5	149	Lab	Fish
Erickson et al. (2016)	1	12	0	108	Fresh_lentic	Fish

Study	# R <sup>2</sup> values	Sample size	Filter pore size [µm]	Amplicon size [bp]	Environment	Target taxa
Lacoursière-Roussel et al. (2016a)	10	15	0.2 to 3	139	Lab	Fish
Lacoursière-Roussel et al. (2016b)	2	12	1.2	66	Fresh_lentic	Fish
Schmelzle & Kinziger (2016)	1	19	3	119	Fresh_lentic	Fish
Baldigo et al. (2017)	2	27	1.5	140	Fresh_lotic	Fish
Doi et al. (2017)	6	6 to 7	0.7	131	Fresh_lotic	Fish
Dunn et al. (2017)	1	30	0	88	Lab	Crustacean
Jo et al. (2017)	2	19	0.7	127 to 719	Marine	Fish
Klobucar et al. (2017)	1	5	10	145	Fresh_lentic	Fish
Larson et al. (2017)	2	7 to 9	1.2	184	Fresh_lentic	Crustacean
Mauvisseau et al. (2017)	1	9	0.2	204	Lab	Invertebrate
Minamoto et al. (2017)	1	6	0.7	151	Marine	Invertebrate
Currier et al. (2018)	3	216	0.7	99 to 129	Fresh_lotic	Mussel
Kamoroff & Goldberg (2018)	2	18	1.2	140	Fresh_lentic	Fish
Maruyama et al. (2018)	1	32	0.7	129	Fresh_lotic	Fish
Mizumoto et al. (2018)	5	10 to 66	0.7	124	Lab	Fish
Nevers et al. (2018)	1	5	0.22	147	Fresh_lentic	Fish
Uthicke et al. (2018)	4	6 to 14	0.22 to 1	126	Lab / Marine	Invertebrate
Wu et al. (2018)	1	10	0.7	166	Fresh_lentic	Crustacean
Bracken et al. (2019)	1	10	0.45	72	Fresh_lotic	Vertebrate
Horiuchi et al. (2019)	1	22	0.7	127	Lab	Fish
Itakura et al. (2019)	2	31	0.7	153	Fresh_lotic	Fish
Iwai et al. (2019)	2	53	0.7	96	Fresh_lotic	Amphibian

Study	# R <sup>2</sup> values	Sample size	Filter pore size [µm]	Amplicon size [bp]	Environment	Target taxa
Kakuda et al. (2019)	1	100	0.7	153	Fresh_lentic	Vertebrate
Minegishi et al. (2019)	2	10	0.7	163	Lab	Fish
Salter et al. (2019)	2	8	0.2	NA	Marine	Fish
Takeuchi et al. (2019)	1	18	0.45	107	Lab	Fish
Akamatsu et al. (2020)	2	4 to 9	0.7	124	Fresh_lentic	Fish
Guivas & Brammell (2020)	2	14	0.7	118	Lab	Fish
Itakura et al. (2020)	2	25	0.7	171	Fresh_lentic	Fish
Weldon et al. (2020)	2	18 to 84	3	159	Fresh_lentic	Fish
Brys et al. (2021)	1	6	0.45	119	Lab	Fish
Capo et al. (2021)	4	14 to 15	0.2	134	Fresh_lentic	Fish
Everts et al. (2021)	4	7 to 8	0.8	84	Lab	Amphibian
Spear et al. (2021)	2	22	0.7	175	Fresh_lentic	Fish
Yates et al. (2021)	2	27	1.5	140	Fresh_lentic	Fish

Note. Target amplicon size was not reported in Salter et al. (2019) because they used a commercial kit.

Table 2. Linear mixed model results showing the effects of filter pore size and amplicon size on species abundance estimate accuracy.

Variable	Estimate	SE	P value	P value
Intercept	0.533	0.984	0.596	
Filter pore size	0.280	0.115	0.017	*
Amplicon size	-0.007	0.006	0.247	

Note: Asterisks indicate statistical significance (\*  $P < 0.05$ ).

### Figure legends

Figure 1. Forest plots of estimated R<sup>2</sup> values among target taxa. Circles and error bars represent estimated R<sup>2</sup> values and 95 % CIs, respectively. The gray dotted line shows R<sup>2</sup> = 0. The number of individual R<sup>2</sup> values required for each forest plot is shown above the plot. Note that the ‘Vertebrate’ category does not include fish and amphibians, and the ‘Invertebrate’ category does not include crustaceans and mollusks, respectively.

Figure 2. Forest plots of estimated R<sup>2</sup> values among study environments. Circles and error bars represent

estimated  $R^2$  values and 95 % CIs, respectively. The gray dotted line shows  $R^2 = 0$ . The number of individual  $R^2$  values required for each forest plot is shown above the plot.

Figure 3.  $R^2$  values with relation to filter pore sizes used for water filtration. White and black circles represent the  $R^2$  values based on biomass and density metrics, respectively.

#### Hosted file

image1.emf available at <https://authorea.com/users/343681/articles/710997-estimation-accuracy-of-species-abundance-based-on-environmental-dna-with-relation-to-its-production-source-and-state-suggested-by-meta-analyses>

#### Figures

Figure 1.

#### Hosted file

image2.emf available at <https://authorea.com/users/343681/articles/710997-estimation-accuracy-of-species-abundance-based-on-environmental-dna-with-relation-to-its-production-source-and-state-suggested-by-meta-analyses>

Figure 2.

#### Hosted file

image3.emf available at <https://authorea.com/users/343681/articles/710997-estimation-accuracy-of-species-abundance-based-on-environmental-dna-with-relation-to-its-production-source-and-state-suggested-by-meta-analyses>

Figure 3.