TITLE

Quantification of total fish eDNA concentration to optimize the detection of rare species and sampling effort

Didier Pont[1*], Paul Meulenbroek[2], Maxime Logez[3], Andrea Funk[2], Alice Valentini[4], Tibor Erős [5,6], Thomas Hein[2]

[1] Institute of Hydrobiology and Aquatic Ecosystem Management, BOKU University, Vienna, Austria
[2] Christian Doppler Laboratory for Meta Ecosystem Dynamics in Riverine Landscapes, BOKU University, Vienna, Austria
[3] INRAE, RIVERLY, F-69625, Villeurbanne Cedex, France
[4] SPYGEN, Savoie Technolac, Le Bourget du Lac, France
[5] HUN-REN Balaton Limnological Research Institute, Klebelsberg K. u. 3, H-8237 Tihany, Hungary
[6] National Laboratory for Water Science and Water Security, HUN-REN Balaton Limnological Research Institute, Tihany, Hungary

[*]Corresponding author:
e-mail address: didier.pont@boku.ac.at

Running title: Quantifying total fish eDNA to improve species detection

ABSTRACT

eDNA metabarcoding is now a widely used method for monitoring aquatic biodiversity, but detecting rare species from extra-organism eDNA samples remains a challenge. Yet, rare species, whether threatened or newly colonizing organisms, are the primary targets of biological conservation measures. Replication and sample volume, as well as PCR replication, are among the main parameters studied to optimize eDNA sampling strategies and workflows. To test the hypothesis that the total abundance of fish eDNA copies in a PCR aliquot is a determining factor for the detection of rare fish species, we replicated water samples from three sites and applied a metabarcoding protocol combined with qPCR measurement of the total fish eDNA abundance amplified by our universal primer, allowing to estimate the number of specific eDNA copies in the eluate. Simulations of the species-specific probabilities of detection (multi-species occupancy model) showed that, even for a high effective sequencing depth of approximately 380,000 reads per sample, more than around 2000 eDNA copies per aliquot are required to detect very rare species (< 1 ‰ of the total number of DNA copies in the eluate). This threshold is lowered by a factor of around 4 and 20 for rare (1 ‰≤ < 1 %) and abundant species (> 1 %), respectively. Our results highlight that increasing sampling and technical replication efforts are required to detect rare species. The total volume of water samples required for species detection varies from less than one litre to 100 litres, depending on the eDNA concentration in the environment. Estimating the total concentration of fish eDNA per litre in the studied water body prior to a metabarcoding sampling campaign would enable optimization of both the volume and number of water samples to be collected, as well as the volume of eDNA aliquot per PCR.

KEY WORDS: environmental DNA, metabarcoding, qPCR, species detection, sampling strategy, multispecies occupancy model, fish

1 | INTRODUCTION

The general decline in biodiversity, changes in species distribution, and restoration efforts aimed at promoting biodiversity recovery (Barral et al. 2015) underscore the need for intensive biological monitoring. In most cases, the primary target is small-sized population of threatened or newly colonizing species. Detecting these rare species is complex due to the relative inefficiency and cost of traditional sampling methods (Mackenzie et al. 2006). Yet, they could represent a significant portion of biodiversity. For example, a compilation of historical electrofishing samples shows that fish species with a relative abundance between 1 ‰ and 1 %, and less than 1 ‰, represent on average29 % and 27 % of the total richness, respectively (Figure S1).

Environmental DNA (eDNA) metabarcoding is now recognized as an effective molecular technique for identifying many taxa in a single experiment in aquatic environments without prior knowledge of the target biological community (Hänfling et al. 2016; Stoeckle et al. 2022; Valentini et al. 2016). Unlike other faunal groups (plankton, macroinvertebrates…), aquatic vertebrate species can only be detected from extra-organismal DNA (cellular or extracellular) mainly extracted from water sample (Valentini et al. 2016; Takenaka et al, 2026). Such eDNA is likely to be degraded and present at a low concentration (Altermatt et al. 2023), which makes the effectiveness of eDNA studies highly sensitive at all stages of the process and increases the risk of flaws and errors (Kelly et al. 2019; Muha et al. 2017; Bunholi et al. 2023; Darling et al. 2020; Ficetola et al. 2015). For a given eDNA workflow, increasing the sampling and the replication efforts improves the probability of detecting the target species (Bylemans et al. 2018; Ficetola et al. 2015; Schultz and Lance 2015; Macher et al. 2021). This is particularly true for rare species, as the abundance of eDNA in water is primarily linked to the size of the target population (Rourke et al. 2021). As Stoeckle et al. (2022) pointed out, “The scarcity of environmental DNA remains the main challenge for current environmental DNA analysis methods”.

eDNA sampling protocols for aquatic vertebrates are not standardized (Harper et al. 2019; Takahashi et al. 2023). The volume of water collected per sample (V) generally varies from 1 to 10 litres (Hänfling et al. 2016; Wilcox et al. 2016; Yamamoto et al. 2017; Harper et al. 2019; Takahashi et al. 2023), with extreme values ranging from 0.24 L to 45–100 L (Bessey et al. 2020; Cantera et al. 2019; Pont et al. 2023; Kang et al. 2024). A rough estimate, based on recent fish eDNA studies, indicates a median volume of two litres per biological sample (Table S1). Although the minimum sample volume has increased compared to pioneering studies, no clear consensus has been established (Amberg et al. 2018; Wilcox et al. 2018), but larger sample volumes (after filtration) should be considered when low densities of the target species are expected (Takahashi et al. 2023).

Similarly, the number of biological replicates and their spatial distribution within the studied system still vary considerably (Evans et al. 2017; Amberg et al. 2018). A low number of water samples is sufficient for abundant species but can limit the detection of rare species (Harper et al. 2019, Fukaya et al. 2021; Sellers et al. 2024). Also, collecting multiple eDNA samples spatially distributed or integrative spatial sampling are more efficient than collecting a single large local sample (Altermatt et al. 2023; Peixoto et al. 2023; Pont et al. 2023).

During eDNA amplification, numerous factors (adapter dimers, non-specific amplicons, polymerase errors, over-amplified duplicates...) influence the proportion of non-informative reads (Sims et al. 2014) and thus the optimal effective sequencing depth (ESD), defined in this study as the total number of reads related to species of interest after amplification, sequencing, computational screening, and target species assignment per biological sample. An estimate of ESD from previous studies gives a median value of approximately 90,000 reads per biological sample and 30,000 reads

per PCR replicate (Table S1), a value of the same order of magnitude as a recent estimate of 60,000 reads per sample and 47,000 reads per PCR replicate (Singer et al. 2019; dos Santos and Blabolil, 2025). A lower sequencing depth can lead to failures in detecting rare sequences in the sample (Shelton et al. 2016; Shirazi et al. 2021) but has little effect on the detection of the most abundant species in a biological community (Peixoto et al. 2023). The number of PCR replicates vary from a few to more than ten (Hänfling et al. 2016; Thomsen et al. 2016; Valentini et al. 2016; Yamamoto et al. 2017). A rough estimate from recent literature indicates a median value of three replicates (Table S1), although eight to twelve replicates are recommended when using Miseq platform technology (Ficetola et al. 2015). The technical replication effort has a significant impact on species detection probabilities, with a marked effect on rare species as opposed to more abundant ones (Ficetola et al. 2015; Fonseca et al., 2018).

While both biological and technical replication efforts are paramount, their relative importance on species detection probabilities is debated (Bylemans et al., 2018). Indeed, beyond these two parameters, the abundance of specific eDNA in a water sample and its spatial distribution are key factors in defining field and analysis strategies (Altermatt et al., 2023; Furlan et al., 2016; Song et al., 2020). A high concentration of eDNA in water facilitates species detection, even with a small water volume or a limited number of replicates (Peixoto et al., 2023). Although the importance of the number of eDNA copies in the eluate extracted from biological samples is clearly established (Altermatt et al., 2023; Buxton et al., 2021; Takahashi et al. 2023), studies focused on estimating the limit of detection mainly in single-species qPCR-based studies (Furlan et al., 2016; Schultz and Lance 2025; Klymus et al., 2019; Song et al. 2020), and only few in metabarcoding studies (Stoeckle et al. 2022).

In this article, we hypothesize that the availability of eDNA copies per PCR replicate, combined with sufficient amplification effort, is a key factor in accurately estimating the effectiveness of fish species detection, i.e., the number of species detected (Fukaya et al., 2021), and in optimizing the detection of rare species. We replicated samples from three floodplain sites differing in species composition and environmental conditions. Then we applied a metabarcoding workflow (Pont et al., 2018) combined with an estimation, in each biological sample, of the total number of eDNA copies amplified by teleo, a universal marker (Valentini et al. 2016) using a qPCR analysis (Pont et al., 2023). Our hypothesis was tested using a multi-species occupancy model (Peixoto et al., 2023). Finally, a strategy for optimizing the size/number of biological replicates and the technical effort of replication is proposed, based on the total eDNA concentration per litre of the target zoological group in the studied environment.

We utilized natural variation in eDNA concentration in the field—rather than experimentally induced variation—to account for the variability in fish communities and environmental conditions across sites that differ in their physicochemical composition and spatial structure. As our primary objective was to provide practitioners with recommendations for optimizing their sampling and analysis efforts, we adopted a field protocol that allowed for the inclusion of a wide range of factors likely to influence species detection probabilities.

2 | MATERIALS AND METHODS

2.1 | eDNA sampling and analytical workflow

The three sampling sites were situated in the Danube floodplain (Austrian National Park Donau-Auen) and were sampled from 20th July to 11th August 2021 (Erős et al., 2024). Their water depths were comparable (1.1–1.4 m). The largest site (SW: Schoenauer Wasser, 0.45 km²) was connected to the main channel at the sampling date (flowing water), while the other two were isolated and stagnant (WA: Witzelsdorfer Arm: 0.06 km²; RA: Rosskopf Aussen: 0.10 km²), the last one (RA) being largely covered by macrophyte stands. Each of the ten water samples per site were collected with a disposable sterile tube attached to a stick situated in front of a boat moving slowly along the shoreline and across the centre of the three studied water bodies. Water was pumped intermittently and filtrated each time the boat passed over a different mesohabitat type applying a quasi-stratified sampling strategy. The sampling process was repeated, the operators ensuring that the proportion of the water volume collected per mesohabitat type remains roughly comparable between each of the ten samples. The operator avoided to resample the mesohabitat types at the same spatial location to prevent a previous resuspension of the sediment, and with the distance travelled for each sample depending on the spatial heterogeneity and the size of the site (a minimum of 10 to 20 meters between two pumping points).. A crossflow filtration capsule (0.45 µm, SPYGEN) was used to filter in situ the sample and immediately after refilled it with a conservation buffer CL1 (SPYGEN) to prevent eDNA degradation. The water volume per sample was 30 L, 12 L and 10 L at the RA, WA and SW sites (mean = 17.3 L, Table S2), depending on the clogging speed of the filtration capsule.

We applied a standardized metabarcoding workflow (Section S1) using “teleo” primers (Valentini et al. 2016) to detect fish eDNA, according to the methodology outlined in Pont et al. (2018, 2023). Briefly, 12 PCR replicates were conducted from each extracted eDNA eluate of 200 µL per sample, and with 3 µL per aliquot. After purification, the PCR products were pooled in equal volumes to achieve a theoretical sequencing depth of 500,000 reads per sample. The libraries were sequenced on a NextSeq sequencer, and the resulting sequences underwent analysis using OBITools package (Boyer et al., 2016). Extraction and amplification negative controls were performed for each extraction and amplification batch, and sequenced. Taxonomic assignment was conducted using the ECOTAG program, combining publicly available sequences and a local database (Pont et al., 2023). Only the 44 taxa known to be present in the study area were retained in order to reduce the risk of false positives, which gives a richness comparable to that estimated by traditional sampling methods in the studied stretch of river with about 50 species (Ramler and Keckeis 2019). For convenience, all MOTUs are referred to as species in the text (Section S2).

Metabarcoding was combined with a real-time quantitative PCR-based estimation of the total number of fish eDNA copies in each sample eluate (Pont et al., 2023), using the same “teleo” primers as for metabarcoding (Section S1). The total number of fish eDNA copies per aliquot (TOTDNA) was defined as the product of this estimate and the ratio of the aliquot volume to the eluate volume and was then constant between aliquots of a given sample.”. The number of specific copies in each of the 12 aliquots for a given water sample was estimated by multiplying TOTDNA by the relative abundance of the species-specific sequences obtained after amplification, sequencing and taxonomic assignment in each PCR replicate.

Given that the number of specific eDNA metabarcoding sequences is an acceptable indicator for inferring relative population abundance of species (Kelly et al. 2014; Lawson Handley et al. 2029; Li et

al. 2019; Pont et al. 2018, Rourke et al. 2021), we considered the degree of rarity of a species in an eluate as representative of its rarity in the environment.

2.2 |Statistical analysis

The number of eDNA copies per litre for each sample is given by: TOTDNA_L = (TOTDNA * DNA_E) / (DNA_A * V), where V, DNA_E, DNA_A represent the volumes of the water sample, the DNA eluate and of each aliquot. To compare the spatial heterogeneity of TOTDNA_L between sites, we computed the dispersion parameter from the negative binomial distributions fitted to TOTDNA_L value for each of the three sites (R package MASS, R Core Team 2022). TOTDNA_L was treated as a discrete variable. The dispersion parameter indicates the degree of aggregation, smaller the value, higher is the degree of clumping, whereas higher value indicates a close to random distribution (Furlan et al., 2016).
To analyse the relationship between the total number of sequences per PCR (READS) and TOTDNA, a linear mixed-effects model (R software, package lmerTest, Kuznetsova et al., 2017) was fitted to predict READS as a function of two fixed parameters: TOTDNA and the identity site. Since TOTDNA was a repeated measure between PCRs from the same sample, the sample identity was included as a random effect.
We constructed a multispecies occupancy model (MSOM) within a Bayesian framework (R package SpOccupancy, Doser et al. 2024) to examine methodological and environmental effects on eDNA detection. MSOM have been already used to highlight both methodological and environmental effects on eDNA detection (Peixoto et al., 2023). The model decomposed the probability of detecting the species at a site into two stochastic components: i) the sampling process resulting in biological replicates containing or not the eDNA of the target species, and ii) the PCR process (technical replicates), which resulted in a detection probability of the target species given that the eDNA was present in the sample. For the sampling process, we treated the site identity as a random factor, and two candidate co-variables as fixed effects: TOTDNA and the cumulated sequencing depth across the 12 PCRs (ESD). For detection probability, the two candidate fixed effects were TOTDNA and READS. Rather than using the raw values of ESD and READS, and in order to dissociate these two correlated variables, we used the residuals of the two linear regressions between ESD and TOTDNA (ESD_res), and between READS and TOTDNA (READS_res) as respective explanatory variables of the sampling process and detection components. TOTDNA, READS and ESD variables are log-transformed.
We ran four MCMC chains of each model for 200,000 iterations, after a burn-in of 100,000 iterations, keeping one value on ten per chain, resulting in 40,000 posterior samples for each parameter. We assessed model convergence checking the Rhat statistic (Rhat < 1.1), the number of effective samples (Neff statistic > 1000) and the traceplots from the parameter samples for each MCMC chain (Gelman et al. 2014). There are two candidate variables for each component of the MSOM, which gives a total of 16 possible variable combinations when both components are considered (Table S4). The site effect is always included as a random factor in the sampling process component. We selected the model with the lowest Watanabe-Akaike information criterion value (WAIC, Watanabe, 2010). For the selected model, we computed posterior predictive checks (function ppcOcc) to assess its goodness of fit (GOF) using a Freeman–Tukey discrepancy statistic (R package SpOccupancy, function ppcOcc, Doser et al. 2024). We also examined the correlation between predicted and observed effectiveness of species detection and of the cumulated probabilities of detection per species for the 30 samples. At the species level, we did not considered models for which the Bayesian p-value was less than 0.1 or greater than 0.9, showing an absence of fit of the model with the data (Hooten and Hobbs, 2015).

To better appreciate the influence of each of the explanatory variables (TOTDNA and READS_res), and for 1, 6, and 12 PCR replicates, we simulated the detection probability for each species, firstly across the observed range of TOTDNA values with READS_res fixed at its mean value, and secondly across the observed range of READS_res values with TOTDNA fixed at its mean value. In the second case, The READS value corresponding to the mean of READS_res was back calculated using the regression of READS against TOTDNA (Figure 1).
We considered three categories of species depending on the mean relative abundance of their sequences: abundant species (≥ 1 %), rare species (≤ 1 ‰ and < 1 %) and very rare species (< 1 ‰). The effective limit of detection for each species (LOD, Klymus et al., 2019) for a given number of PCR was calculated as the product of the mean relative abundance of the corresponding specific eDNA sequences and the TOTDNA value required to achieve a 95 % detection probability.
To evaluate the impact of PCR replication on the estimated efficiency of species detection (cumulative species detection probabilities), we simulated for a number of PCRs ranging from 2 to 12 the TOTDNA and READS_res values required to detect 95% of the number of abundant, rare, very rare species, and all species present in the eluate.
Similar graphs were established considering volume gradients (V) and TOTDNA_L, ranging from 1 to 100 L, and from 1,000 to 30,000 copies respectively. For 6 and 12 PCR, and for abundant, rare and very rare species, isolines of 95 % of effectiveness of species detection were plotted at levels corresponding to the mean, 2.5th and 97.5th percentiles of READS_res distribution (Figure 3, A-C).
To highlight the influence of the number of biological samples, we simulated the variability of specific detection effectiveness for a range of one to five 20 L water samples, with 12 PCR replicates per sample (Figure 3 D). Isolines corresponding to a species detection effectiveness of 95% were plotted at levels corresponding to the mean, 2.5th and 97.5th percentiles of the READS_res distribution.
All analyses (see R script, Section S3) were conducted using R Statistical Software (v4.5.2; R Core Team 2022).

3 | RESULTS

3.1 | Sequencing depth and eDNA copies variability

TOTDNA_L ranged from 43 to 19,508 DNA copies. L^{-1} (mean = 4,018) and differed significantly (Kruskal-Wallis test, $p < 0.001$) among the three sites (Table S2). Furthermore TOTDNA_L were more clustered at the RA site than at the WA and SW sites, with values for the negative binomial distribution dispersion parameter of 1.19, 2.90, and 16.30, respectively, and a bell-shaped distribution of TOTDNA_L only at the SW site. For a single PCR replicate, TOTDNA ranged from 8 to 8,779 copies (median and mean of 369 and 1,541 copies).
The species found in the negative controls were only human and domestic animals (e.g. chicken, turkey, pig, dog and/or cat). After the filtering pipeline, the extraction and PCR negative controls were completely clean, and no sequence reads remained in those samples. After bioinformatic filtering and species identification, the cumulated number of sequences was 10,972,794 compared to 14,854,322 initially. The final mean READS value was 31,531 reads (12 to 180,468 reads, Table S2). 12 PCRs failed to detect any sequences, and 25 PCRs did not generate more than 1,000 reads out of a total of 360 PCRs. As all the DNA samples underwent the same laboratory protocol and were analysed in parallel, these failures were mainly due to very low eDNA concentration The number of PCRs per sample with a READS value over 1000 ranged from 4 to 12 (mean = 10.77).

TOTDNA was significantly correlated to ESD and READS with Pearson correlation coefficients of 0.517 and 0.379 respectively ($P < 0.001$). The linear mixed model predicting READS as a function of TOTDNA and the identity site as fixed parameters (Figure S2) had a substantial explanatory power (conditional $R^2 = 0.52$), and with a part related to the fixed effect alone (marginal R^2) of 0.16 (see Table S3 for model parameters). TOTDNA had a significant positive effect ($p < 0.001$) whereas the site identity was unsignificant ($P > 0.05$).

3.2 | Probability of species detection

Of 44 species detected, 44, 23 and 15 were present at the SW, WA and RA site, respectively. The mean specific occurrence per PCR replicate varied between 0.003 and 0.951; and the mean relative abundance of specific reads between 0.001 to 21.39 %.

The MSOM, including only the site identity to the sampling process component and the co-variates TOTDNA and READS_res to the detection probability component, had the lowest WAIC (Table S4). A second model had also a WAIC value close to this model, but including ESD_res to the occupancy component. Following the Ockham's razor principle, the simplest model was retained.

The convergence of the parameters was correctly achieved with $1.00 < Rhat < 1.01$ and $Neff > 1000$. The Bayesian statistics at the community level indicated a correct fit ($P = 0.30$ and 0.59 between samples and replicates, respectively). Furthermore, the predicted and observed number of species for each of the 30 samples were strongly correlated (Pearson correlation coefficient = 0.992, $P > 0.001$) and the regression coefficients of the linear relationship between these two variables did not differ significantly from 0 and 1, respectively (Student's t-test, $p > 0.05$). At the specific level, 31 of the 44 species showed a correct fit with their respective models. Among them, 6, 9 and 16 species were abundant, rare, and very rare species.

The two co-variables had a positive influence on the specific detection probabilities (Figure 1). With one PCR replicate and an average value of READS_res, to achieve a probability of detection of 0.95, at least 157 copies were needed for abundant species, while few rare species and most very rare species did not reach this threshold. With six replicates, all species exceeded 95% detection probabilities (Figure 1 left panels) when TOTDNA was over 25, 119 and 480 copies for abundant, rare and very rare species, respectively. For one replicate, species LOD values ranged from 0.19 to 81.7. The mean LOD values were 13.4, 3.1, 1.8 and 1.2 copies per PCR replicate for one, three, six and 12 replicates, respectively (Figure S3).

With only one replicate and a mean TOTDNA value, the specific detection probability responded positively to an increase in the sequencing depth but remained below 0.95 for rare species (except one) and very rare species. With 6 replicates, most of the rare species reached the detection threshold, but 12 PCR were needed to detect most of the very rare species (Figure1 right panels).

3.3 | Effectiveness of species detection

With only one PCR, the effectiveness of species detection did not reach 95 % of the maximum number of species. For an increasing number of replicates, the required TOTDNA and READS_res values to detect 95 % of the total number of species decreased sharply (Figure 2). For the mean value of READS_res and with 12 PCR, TOTDNA values needed to reach a 95 % effectiveness for all species, very rare, rare and abundant species were 1,476, 2,083, 449 and 95 DNA copies, respectively.

With 12 PCR, a TOTDNA_L of 5,000 copies. L^{-1} and a sequencing depth corresponding to the mean READS_res value, the water volume needed to reach a 95 % effectiveness of species detection was 27.1 L, 6.2 L, and 1.6 L for very rare species, rare species, and abundant species, respectively (Figure

3, A-C). For low TOTDNA_L (500 eDNA copies. L^{-1}), very rare species were not detected, and water volumes of 43.1 L and 9.44 L were needed for rare and abundant species. Abundant species were still detected at TOTDNA_L = 50 copies. L^{-1} for a volume of 25.7 L. For six instead of 12 PCR, the detection of common and rare/very rare species requires volumes approximately 1.2 and 2 times larger, respectively.
Regarding the number of 20 L samples required to achieve a species detection efficiency of 95% (Figure 3D), one and five samples were needed for a TOTDNA_L value of approximately 100 copies L^{-1} for abundant and rare species, respectively. For very rare species, this threshold was reached only at a TOTDNA_L value of around 1,500 copies and with a total of five samples.

4 | DISCUSSION

Our READS and ESD values are in good agreement with published values (Singer et al. 2019; dos Santos and Blabolil 2025) and are positively correlated with TOTDNA. This result seems to contradict the generally accepted idea that metabarcoding PCR tends to generate a number of sequences independent of the initial amount of DNA (Kelly et al. 2019). However, classical sequencers like NextSeq are known to be less capable of sequencing low abundance eDNA (Singer et al. 2019). The higher correlation between TOTDNA and ESD at the sample level confirms the link between numbers of eDNA copies and highlights the high variability of the amplification process across PCR replicates. The detection component of the MSOM demonstrates the positive influence of TOTDNA and READS on species detection whereas the sampling process component depends primarily on the site effect and secondarily on ESD, although the latter variable does not improve the model's predictive performance. Twelve of the thirteen species with poor model fit were present at all three sites, and all were observed in 70 % to 100 % of the water samples (average 91 %). Such unbalanced data, combined with a limited number of samples, are known to affect the estimates of logistic models (Lange, 2024). For correctly modelled species, our specific LOD values are similar to those examined by Klymus et al. (2019), with a mean value for one PCR of 13.4 and 14.7 eDNA copies, respectively, and values below 1.5 copies beyond eight PCRs.

4.1 |Technical replication and eDNA copies per aliquot

TOTDNA and ESD have different impacts on detection by metabarcoding (Anmarkrud et al. 2025): the first is related to the abundance of eDNA in the environment and the sampling effort whereas the second is reflecting the efficiency of the amplification (ESD) and the replication effort.
PCR replication is the classic method for improving the detection of rare species, requiring a larger number of sequences than for abundant ones (Strickland and Roberts 2019, Shirazi et al. 2021). Our results suggest that, for an initial average TOTDNA value of around 1,500 eDNA copies, more than six PCRs are required to detect very rare species, leading to ESD values of 189,000 and 379,000 for six and twelve PCRs, respectively. Several authors have previously recommended such a high number of PCR replicates (Ficetola et al. 2015, Dopheide et al. 2029), far exceeding the median observed in recent metabarcoding literature (Table S1). It is worth noting that a significant proportion of PCRs typically fail (10% in this study) and increasing their number provides also a safety margin. Furthermore, the use of the next-generation NovaSeq sequencer, which generates a very large number of reads, would improve ESD without increasing the number of PCR replicates (Singer et al. 2019).

The combination of the amount of TOTDNA copies, READS and PCR replication determines the effectiveness of species detection. For fewer than five replicates, only the highest TOTDNA/READS pairs are sufficient to detect 95 % of the total number of species. For a mean READS value, and even with 12 PCRs, a TOTDNA value below 1,500 copies will lead to an underestimation of the total species richness. This threshold is of 2000 eDNA copies for very rare species, but 4 and 20 times lower for rare and abundant species, respectively. For a given READS and replication effort (analytical conditions), TOTDNA is the main factor determining the effectiveness of species detection in a metabarcoding experiment and is itself governed by the eDNA concentration in the environment, the sample volume and the aliquot/eluate volume ratio.

4.2 | Water volume and number of biological samples

Previous studies already showed that a combination of large volume and/or multiple eDNA samples are needed to each 95 % effectiveness of species detection, for example ten 2 L samples in lakes (Sellers et al. 2024), 5 to 10 $ L samples in rivers (Van Driessche et al, 2024). In a highly diverse tropical river, only 64 % and 71 % of species were detected in a tropical river with water volumes of 34 L and 68 L (Cantera et al. 2019). Smaller water volumes (1 or 2 L) lead to fewer species detection (Bessey et al. 2020). When only targeting common marine fish species, 8 L sample are recommended (Stoeckle et al. 2022), but 10 water samples of 10 or 20 L were collected to detect the most common species in deep sea water (Yoshida et al. 2023). Detecting rare species when the targeted eDNA is likely to be scarce require sample volume over 2 L (Cantera et al. 2019; Sepulveda et al. 2021), or seven time higher than for an abundant species (Peixoto al. 2023). Our results highlight the importance of TOTDNA_L for optimizing sampling effort based on species rarity. Even for a high effective sequencing depth of approximately 380,000 reads per sample, more than around 2000 eDNA copies per aliquot are required to detect very rare species. This threshold is lowered by a factor of around 4 and 20 for rare and abundant species, respectively. These results have significant implications for sample volume and number, with values ranging from 1 to over 100 L per sample (i.e. from one to more than five 20-liter water samples) depending on TOTDNA_L and the relative abundance of the species-specific sequences. But the increasing number and size of water sample have their own limitations. The volume filtered per water sample is very often limited due to filter clogging or the specificity of the environment studied (Dan et al. 2024), and collecting a larger number of biological samples is cost and time-consuming.

Water volume and the number of biological samples both improve the effectiveness of species detection (Anmarkrud et al. 2025; Macher et al. 2021), but their impacts differ. A larger water volume increases the total number of eDNA copies collected, while, for a similar total water volume, a greater number of biological replicates could improve the representativeness of the fish community, if eDNA is unevenly distributed across the study area (Harper et al. 2019; Peixoto et al. 2023). From this perspective, eDNA sampling presents similar challenges to those of traditional sampling techniques (Shelton et al. 2016, Stein et al. 2014). However, unlike most traditional sampling methods, the spatial distribution of eDNA within a water body is controlled not only by the distribution of target species, but also by a combination of processes regulating its transport, degradation rate, and sedimentation rate (Pont, 2024; Shogren et al. 2018; Wilcox et al. 2016). Regardless of the abundance of the target fauna, eDNA produced in lentic water tends to persist longer in the water column or become trapped around aquatic plants (Harper et al. 2019). The absence of current and a higher habitat diversity also increases the spatial heterogeneity of its distribution and reduces its dispersal (Li et al. 2019). Conversely, in free-flowing environments, only

recently released eDNA is present in the water column and turbulent flow tends to homogenize its spatial distribution (Evans et al. 2017; Wang et al. 2021). Collecting surface water samples from one bank to the other allows for integrated spatial sampling of the stream's cross-section with only one or two large-volume water samples (Pont et al. 2023).

4.3 | False positive

Furthermore, the detection of rare species presents an increased risk of false positives. Broadly speaking, a false positive refers to the detection of eDNA in a water sample when the species is not present at the site, due to: i) contamination / error during sampling or analysis; and ii) the presence of eDNA in a water sample when the target organism has not been observed at the site due to eDNA transport from a distant source or to the rarity of the organism (Barnes et al. 2014, Darling et al. 2020; Pont et al. 2024). It is also worth noting the base rate error, which means that confirming the presence of rare species at a site requires significant sampling and analysis effort (Darling et al., 2020). Considering the detection of a species only in a single PCR replicate and/or with a low number of reads as a false positive is a classic approach, but it also eliminates true positives that are present only at low frequencies in the eDNA sample (Pont et al., 2018; Shirazi et al., 2021). Comparison with a traditional long-term survey is a reliable method for confirming the actual presence of a rare species detected by eDNA. In the absence of such capture data series, the probability of a true local occurrence remains uncertain (Jerde, 2021; Song et al., 2020).

Our recommendations related to the sampling and analytical efforts had several limitations. Among them, the relationship between TOTDNA, sample volume and ESD depends on the proportion of non-informative reads, which is itself linked to the composition of the water sample (PCR inhibition due to co-extracted substances...), eDNA quality, primer selection and the analytical procedure (Jane et al. 2015; Collins et al. 2019; Sims et al. 2014). Also, as previously mentioned, a highly aggregative distribution of species-specific eDNA could lead to a lower detectability than expected for a given sampling effort (Furlan et al. 2016). Nevertheless, PCR is essentially a stochastic process that can be described by probability distributions, even though amplification imperfections affect the results (Weusten & Herbergs 2012; Shelton et al., 2013); the order of magnitude we recommend for the minimum total number of target-species DNA copies per aliquot can therefore be considered applicable to a wide range of situations.

Moreover, we hypothesize that there is a positive relationship between eDNA concentrations and the abundance of focal species (Rourke et al. 2021), even if many other factors can affect fish release DNA rate into the water (Wang et al., 2021). Our MSOM model relied on the assumption that detection probabilities for rare species are comparable to those for non-rare species; an assumption potentially biased by species-specific traits that make certain species difficult to detect. However, among the 44 species of our study, we found no link between rarity and eight ecological and biological species traits (Table S5). Also, no clear link between species rarity and their traits emerges from a large dataset compiling traditional electrofishing samples covering most of French river types (Table S6). Nevertheless, the possibility of such an underestimation of elusive species in certain local situations cannot be ruled out. Comparative studies using varying for their eDNA primer, metabarcoding workflow, and conducted in diverse aquatic environments would be particularly useful for confirming our recommendations.

4.4 | Conclusion

The scarcity of extra-organismal eDNA is one of the main limitations of current metabarcoding methods, making it difficult to detect rare species and accurately assess biodiversity patterns (Stoeckle et al., 2022). Our results demonstrate that abundant species are easy to detect even with reduce sampling and analytical effort. Given sufficient replication effort (6 to 12 PCRs per sample), the detection efficiency for rare species/very rare species depends largely on the number of target eDNA copies in each aliquot subjected to amplification and sequencing; this number being determined by the concentration of target eDNA in the environment, the sample volume and number, and the aliquot volume. Detecting very rare species would require a very high total eDNA concentration per litre, combined with a considerable sampling effort. In practice, detecting rare species appears to be a reasonable objective.

Then, when targeting such rare species, sampling strategy and analytical effort must be tailored to each local situation. Collecting a few water samples from a water body to estimate total fish eDNA concentration prior to a metabarcoding sampling campaign would allow for the optimization of the water volume filtered per sample; the goal would be to obtain an eluate suitable for preparing aliquots containing one to several thousand copies of total fish eDNA. Furthermore, the variability in total eDNA concentration estimated from these "pre-samples" could provide a rough estimate of the spatial heterogeneity of eDNA distribution and help strike the right balance between sample volume and the number of biological samples. It would also be possible to adjust the volume of each aliquot based on the estimated total number of eDNA copies present in the eDNA eluate. Finally, the development of portable thermal cyclers for on-site qPCR analysis would enable rapid estimations of TOTDNA_L at a low cost (Shrestha et al., 2026).

Although comprehensive inventories of aquatic biodiversity using eDNA metabarcoding remain unattainable with metabarcoding method (Shirazi et al., 2021), we anticipate that our recommendations will enhance the detection of rare species and support more effective biodiversity monitoring.

ACKNOWLEDGEMENTS and FUNDING

We thank the Donau-Auen National Park Authority for their support, BOKU University for Open Access funding, Jorgen Steindl, Peter Bader and David Gandolf for help in the field, and the SPYGEN staff for their participation to the eDNA analysis.

This work was supported by the Austrian Science Fund (FWF) project RIMECO (Grant-DOI10.55776/I5006) and a Hungarian ANN-OTKA 141884 grant, and the Széchenyi Plan Plus Program with the support of the RRF 2.3.1 21 2022 00008 project. This study was financially supported by the Austrian Federal Ministry of Economy, Energy and Tourism, the National Foundation for Research, Technology and Development, and the Christian Doppler Research Association.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

All Illumina raw sequence data and the number of reads detected for each assigned taxon in each field are available on DOI: 10.5281/zenodo.2268753

LITERATURE CITED

Altermatt, F., Carraro, L., Antonetti, M., Albouy, C., Zhang, Y., Lyet, A., . . . Pellissier, L. (2023). Quantifying biodiversity using eDNA from water bodies: General principles and recommendations for sampling designs. Environmental DNA, 5(4), 671-682. doi:10.1002/edn3.430

Amberg, J. J., Bonar, S. A., Perez, C., Rees, C. B., Jackson, C., & Ulibarri, R. M. (2018). The Importance of Sound Methodology in Environmental DNA Sampling: Response to Comment. North American Journal of Fisheries Management, 38(3), 597-600. doi:10.1002/nafm.10054

Anmarkrud, J. A., dos Anjos Santa Rosa, F., Thorbek, L., Schrøder-Nielsen, A., Sviggum, S. M., Ready, J. S., . . . Mauvisseau, Q. (2025). eDNA Replicates, Polymerase and Amplicon Size Impact Inference of Richness Across Habitats. Environmental DNA, 7(3), e70095. doi:https://doi.org/10.1002/edn3.70095

Barnes, M. A., Turner, C. R., Jerde, C. L., Renshaw, M. A., Chadderton, W. L., & Lodge, D. M. (2014). Environmental conditions influence eDNA persistence in aquatic systems. Environ Sci Technol, 48(3), 1819-1827. doi:10.1021/es404734p

Barral, M. P., Rey Benayas, J. M., Meli, P., & Maceira, N. O. (2015). Quantifying the impacts of ecological restoration on biodiversity and ecosystem services in agroecosystems: A global meta-analysis. Agriculture, Ecosystems & Environment, 202, 223-231. doi:https://doi.org/10.1016/j.agee.2015.01.009

Bessey, C., Jarman, S. N., Berry, O., Olsen, Y. S., Bunce, M., Simpson, T., . . . Keesing, J. (2020). Maximizing fish detection with eDNA metabarcoding. Environmental DNA, 2(4), 493-504. doi:10.1002/edn3.74

Boyer, F., Mercier, C., Bonin, A., Le Bras, Y., Taberlet, P., & Coissac, E. (2016). obitools: a unix-inspired software package for DNA metabarcoding. Mol Ecol Resour, 16(1), 176-182. doi:10.1111/1755-0998.12428

Bunholi, I. V., Foster, N. R., & Casey, J. M. (2023). Environmental DNA and RNA in aquatic community ecology: Toward methodological standardization. Environmental DNA, 5(6), 1133-1147. doi:https://doi.org/10.1002/edn3.476

Bylemans, J., Gleeson, D. M., Lintermans, M., Hardy, C. M., Beitzel, M., Gilligan, D. M., & Furlan, E. M. (2018). Monitoring riverine fish communities through eDNA metabarcoding: determining optimal sampling strategies along an altitudinal and biodiversity gradient. Metabarcoding and Metagenomics, 2. doi:10.3897/mbmg.2.30457

Cantera, I., Cilleros, K., Valentini, A., Cerdan, A., Dejean, T., Iribar, A., . . . Brosse, S. (2019). Optimizing environmental DNA sampling effort for fish inventories in tropical streams and rivers. Scientific Reports, 9(1), 3085. doi:10.1038/s41598-019-39399-5

Collins, R. A., Bakker, J., Wangensteen, O. S., Soto, A. Z., Corrigan, L., Sims, D. W., . . . Mariani, S. (2019). Non-specific amplification compromises environmental DNA metabarcoding with COI. Methods in Ecology and Evolution, 10(11), 1985-2001. doi:10.1111/2041-210x.13276

Dan, M. E., Portner, E. J., Bowman, J. S., Semmens, B. X., Owens, S. M., Greenwald, S. M., & Choy, C. A. (2024). Using low volume eDNA methods to sample pelagic marine animal assemblages. Plos One, 19(5). doi:10.1371/journal.pone.0303263

Darling, J. A. (2020). How to learn to stop worrying and love environmental DNA monitoring. Aquat Ecosyst Health Manag, 22(4), 440-451. doi:10.1080/14634988.2019.1682912

Dopheide, A., Xie, D., Buckley, T. R., Drummond, A. J., & Newcomb, R. D. (2019). Impacts of DNA extraction and PCR on DNA metabarcoding estimates of soil biodiversity. Methods in Ecology and Evolution, 10(1), 120-133. doi:https://doi.org/10.1111/2041-210X.13086

dos Santos, R. A., & Blabolil, P. (2025). Comparison of Bioinformatic Pipelines for eDNA Metabarcoding Data Analysis of Fish Populations. Fishes, 10(5). doi:10.3390/fishes10050214

Doser, J. W., Finley, A. O., Kéry, M., & Zipkin, E. F. (2022). spOccupancy: An R package for single-species, multi-species, and integrated spatial occupancy models. Methods in Ecology and Evolution, 13(8), 1670-1678. doi:https://doi.org/10.1111/2041-210X.13897

Erős, T., Funk, A., Pont, D., Hein, T., Meulenbroek, P., Preiszner, B., . . . Czegledi, I. (2024). eDNA metabarcoding reveals the role of habitat specialization and spatial and environmental variability in shaping diversity patterns of fish metacommunities. *Plos One, 19*(1). doi:10.1371/journal.pone.0296310

Evans, N. T., Li, Y. Y., Renshaw, M. A., Olds, B. P., Deiner, K., Turner, C. R., . . . Pfrender, M. E. (2017). Fish community assessment with eDNA metabarcoding: effects of sampling design and bioinformatic filtering. Canadian Journal of Fisheries and Aquatic Sciences, 74(9), 1362-1374. doi:10.1139/cjfas-2016-0306

Ficetola, G. F., Pansu, J., Bonin, A., Coissac, E., Giguet-Covex, C., De Barba, M., . . . Taberlet, P. (2015). Replication levels, false presences and the estimation of the presence/absence from eDNA metabarcoding data. Mol Ecol Resour, 15(3), 543-556. doi:10.1111/1755-0998.12338

Fonseca, V. G. (2018). "Pitfalls in relative abundance estimation using eDNA metabarcoding". Molecular Ecology Resources, 18(5), 923-926. doi:10.1111/1755-0998.12902

Fukaya, K., Kondo, N. I., Matsuzaki, S. i. S., & Kadoya, T. (2021). Multispecies site occupancy modelling and study design for spatially replicated environmental DNA metabarcoding. Methods in Ecology and Evolution, 13(1), 183-193. doi:10.1111/2041-210x.13732

Furlan, E. M., Gleeson, D., Hardy, C. M., & Duncan, R. P. (2016). A framework for estimating the sensitivity of eDNA surveys. Mol Ecol Resour, 16(3), 641-654. doi:10.1111/1755-0998.12483

Gelman, A., Hwang, J., & Vehtari, A. (2014). Understanding predictive information criteria for Bayesian models. Statistics and Computing, 24(6), 997-1016. doi:10.1007/s11222-013-9416-2

Hänfling, 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. Mol Ecol, 25(13), 3101-3119. doi:10.1111/mec.13660

Harper, L. R., Handley, L. L., Carpenter, A. I., Ghazali, M., Di Muri, C., Macgregor, C. J., . . . Hanfling, B. (2019). Environmental DNA (eDNA) metabarcoding of pond water as a tool to survey conservation and management priority mammals. Biological Conservation, 238. doi:ARTN 108225 10.1016/j.biocon.2019.108225

Hooten, M. B., & Hobbs, N. T. (2015). A guide to Bayesian model selection for ecologists. Ecological Monographs, 85(1), 3-28. doi:https://doi.org/10.1890/14-0661.1

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. Mol Ecol Resour, 15(1), 216-227. doi:10.1111/1755-0998.12285

Jerde, C. L. (2021). Can we manage fisheries with the inherent uncertainty from eDNA? J Fish Biol, 98(2), 341-353. doi:10.1111/jfb.14218

Kang, Y., & Song, Y. (2026). Environmental insights from non-target detections in urban eDNA metabarcoding. iScience, 29(2). doi:10.1016/j.isci.2026.114632

Kelly, R. P., Port, J. A., Yamahara, K. M., & Crowder, L. B. (2014). Using environmental DNA to census marine fishes in a large mesocosm. *Plos One, 9*(1), e86175. doi:10.1371/journal.pone.0086175

Kelly, M. (2019). Adapting the (fast-moving) world of molecular ecology to the (slow-moving) world of environmental regulation: lessons from the UK diatom metabarcoding exercise. Metabarcoding and Metagenomics, 3. doi:10.3897/mbmg.3.39041

Klymus, K. E., Merkes, C. M., Allison, M. J., Goldberg, C. S., Helbing, C. C., Hunter, M. E., . . . Richter, C. A. (2019). Reporting the limits of detection and quantification for environmental DNA assays. Environmental DNA, 2(3), 271-282. doi:10.1002/edn3.29

Kuznetsova, A., Brockhoff Per B., Rune H. Christensen B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, 82(13), 1--26. doi:10.18637/jss.v082.i13

Lawson Handley, L., Read, D. S., Winfield, I. J., Kimbell, H., Johnson, H., Li, J., . . . Hänfling, B. (2019). Temporal and spatial variation in distribution of fish environmental DNA in England's largest lake. *Environmental DNA, 1*(1), 26-39. doi:10.1002/edn3.5

Lange, C. (2024). *Practical Machine Learning with R: Tutorials and Case Studies* (1st ed. ed.): Chapman and Hall/CRC. London.

Li, J., Lawson Handley, L. J., Harper, L. R., Brys, R., Watson, H. V., Di Muri, C., . . . Hänfling, B. (2019). Limited dispersion and quick degradation of environmental DNA in fish ponds inferred by metabarcoding. Environmental DNA, 1(3), 238-250. doi:10.1002/edn3.24

Macher, T.-H., Schütz, R., Arle, J., Beermann, A. J., Koschorreck, J., & Leese, F. (2021). Beyond fish eDNA metabarcoding: Field replicates disproportionately improve the detection of stream associated vertebrate species. Metabarcoding and Metagenomics, 5, 59-71. doi:10.3897/mbmg.5.66557

MacKenzie, D. I., J. D. Nichols, J. A. Royle, K. H. Pollock, L. L. Bailey, and J. E. Hines. 2006. Occupancy Estimation and Modeling. Amsterdam: Academic Press.

Muha, T. P., Rodriguez-Rey, M., Rolla, M., & Tricarico, E. (2017). Using Environmental DNA to Improve Species Distribution Models for Freshwater Invaders. Frontiers in Ecology and Evolution, 5, 7. doi:ARTN 158, 10.3389/fevo.2017.00158

Peixoto, S., Mota-Ferreira, M., Chaves, C., Velo-Antón, G., Beja, P., & Egeter, B. (2023). Multi-species occupancy modeling reveals methodological and environmental effects on eDNA detection of amphibians in temporary ponds. Environmental DNA, 5(4), 796-811. doi:https://doi.org/10.1002/edn3.443

Pont, D. (2024). Predicting downstream transport distance of fish eDNA in lotic environments. Mol Ecol Resour, 24(4), e13934. doi:10.1111/1755-0998.13934

Pont, D., Meulenbroek, P., Bammer, V., Dejean, T., Erős , T., Jean, P., . . . Valentini, A. (2023). Quantitative monitoring of diverse fish communities on a large scale combining eDNA metabarcoding and qPCR. Mol Ecol Resour, 23(2), 396-409. doi:10.1111/1755-0998.13715

Pont, D., Rocle, M., Valentini, A., Civade, R., Jean, P., Maire, A., . . . Dejean, T. (2018). Environmental DNA reveals quantitative patterns of fish biodiversity in large rivers despite its downstream transportation. Scientific Reports, 8(1), 10361. doi:10.1038/s41598-018-28424-8

Ramler D. & H. Keckeis (2019). Effects of large-river restoration measures on ecological fish guilds and focal species of conservation in a large European river (Danube, Austria), Science of The Total Environment,686: 1076-1089. https://doi.org/10.1016/j.scitotenv.2019.05.373.

R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rourke, M. L., Fowler, A. M., Hughes, J. M., Broadhurst, M. K., DiBattista, J. D., Fielder, S., . . . Furlan, E. M. (2021). Environmental DNA (eDNA) as a tool for assessing fish biomass: A review of approaches and future considerations for resource surveys. Environmental DNA, 4(1), 9-33. doi: 10.1002/edn3.185

Schultz, M. T., & Lance, R. F. (2015). Modeling the Sensitivity of Field Surveys for Detection of Environmental DNA (eDNA). Plos One, 10(10), e0141503. doi:10.1371/journal.pone.0141503

Sellers, G. S., Jerde, C. L., Harper, L. R., Benucci, M., Di Muri, C., Li, J., . . . Hänfling, B. (2024). Optimising species detection probability and sampling effort in lake fish eDNA surveys. Metabarcoding and Metagenomics, 8, e104655.

Sepulveda, A. J., Hoegh, A., Gage, J. A., Eldridge, S. C. L., Birch, J. M., Stratton, C., . . . Barnhart, E. P. (2021). Integrating Environmental DNA Results With Diverse Data Sets to Improve Biosurveillance of River Health. Frontiers in Ecology and Evolution, 9. doi:ARTN 620715 10.3389/fevo.2021.620715

Shelton, A. O., O'Donnell, J. L., Samhouri, J. F., Lowell, N., Williams, G. D., & Kelly, R. P. (2016). A framework for inferring biological communities from environmental DNA. Ecol Appl, 26(6), 1645-1659. doi:10.1890/15-1733.1

Shirazi, S., Meyer, R. S., & Shapiro, B. (2021). Revisiting the effect of PCR replication and sequencing depth on biodiversity metrics in environmental DNA metabarcoding. Ecol Evol, 11(22), 15766-15779. doi:10.1002/ece3.8239

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. Environ Sci Technol, 52(15), 8530-8537. doi:10.1021/acs.est.8b01822

Shrestha, K., Farhana, F. Z., Koo, K. M., Darwish, N., Ross, A. G., & Shiddiky, M. J. A. (2026). Challenges in the Development and Implementation of Point-of-Care Handheld qPCR Devices. *Small, 22*(3), e06233. doi:https://doi.org/10.1002/smll.202506233

Sims, D., Sudbery, I., Ilott, N. E., Heger, A., & Ponting, C. P. (2014). Sequencing depth and coverage: key considerations in genomic analyses. Nat Rev Genet, 15(2), 121-132. doi:10.1038/nrg3642

Singer, G. A. C., Fahner, N. A., Barnes, J. G., McCarthy, A., & Hajibabaei, M. (2019). Comprehensive biodiversity analysis via ultra-deep patterned flow cell technology: a case study of eDNA metabarcoding seawater. Scientific Reports, 9(1), 5991. doi:10.1038/s41598-019-42455-9

doi:10.1371/journal.pone.0090234

Song, J. W., Schultz, M. T., Casman, E. A., Bockrath, K. D., Mize, E., Monroe, E. M., . . . Small, M. J. (2020). A probabilistic model for designing and assessing the performance of eDNA sampling protocols. Mol Ecol Resour, 20(2), 404-414. doi:10.1111/1755-0998.13113

Stein, A., Gerstner, K., & Kreft, H. (2014). Environmental heterogeneity as a universal driver of species richness across taxa, biomes and spatial scales. Ecology Letters, 17(7), 866-880. doi:https://doi.org/10.1111/ele.12277

Stoeckle, M. Y., Adolf, J., Ausubel, J. H., Charlop-Powers, Z., Dunton, K. J., & Hinks, G. (2022). Current laboratory protocols for detecting fish species with environmental DNA optimize sensitivity and reproducibility, especially for more abundant populations. Ices Journal of Marine Science, 79(2), 403-412. doi:10.1093/icesjms/fsab273

Strickland, G. J., & Roberts, J. H. (2019). Utility of eDNA and occupancy models for monitoring an endangered fish across diverse riverine habitats. Hydrobiologia, 826(1), 129-144. doi:10.1007/s10750-018-3723-8

Takahashi, M., Sacco, M., Kestel, J. H., Nester, G., Campbell, M. A., van der Heyde, M., . . . Allentoft, M. E. (2023). Aquatic environmental DNA: A review of the macro-organismal biomonitoring revolution. *Science of the Total Environment, 873*. doi:10.1016/j.scitotenv.2023.162322

Takenaka, M., Uchida, N., Hasebe, Y., Yano, K., & Ueki, G. (2026). Environmental DNA metabarcoding for insect biodiversity surveys: Current status and future prospects for potential applications in highly diverse insect taxa. *Entomological Science, 29*(1). doi:10.1111/ens.12616

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

Valentini, A., Taberlet, P., Miaud, C., Civade, R., Herder, J., Thomsen, P. F., . . . Dejean, T. (2016). Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. Mol Ecol, 25(4), 929-942. doi:10.1111/mec.13428

Van Driessche, C., Everts, T., Neyrinck, S., Deflem, I., Bonte, D., & Brys, R. (2024). Reduced sampling intensity through key sampling site selection for optimal characterization of riverine fish communities by eDNA metabarcoding. *Ecological Indicators, 169*. doi:10.1016/j.ecolind.2024.112807

Wang, S., Yan, Z., Hanfling, B., Zheng, X., Wang, P., Fan, J., & Li, J. (2021). Methodology of fish eDNA and its applications in ecology and environment. Sci Total Environ, 755(Pt 2), 142622. doi:10.1016/j.scitotenv.2020.142622

Watanabe, S. (2010). Asymptotic Equivalence of Bayes Cross Validation and Widely Applicable Information Criterion in Singular Learning Theory. Journal of Machine Learning Research, 11, 3571-3594.

Weusten, J., and J. Herbergs. 2012. “A Stochastic Model of the Processes in PCR Based Amplification of STR DNA in Forensic Applications.” Forensic Science International: Genetics 6, no. 1: 17–25. https://doi.org/10.1016/j.fsigen.2011.01.003

Wilcox, T. M., Carim, K. J., Young, M. K., McKelvey, K. S., Franklin, T. W., & Schwartz, M. K. (2018). Comment: The Importance of Sound Methodology in Environmental DNA Sampling. North American Journal of Fisheries Management, 38(3), 592-596. doi:10.1002/nafm.10055

Wilcox, T. M., McKelvey, K. S., Young, M. K., Sepulveda, A. J., Shepard, B. B., Jane, S. F., . . . Schwartz, M. K. (2016). Understanding environmental DNA detection probabilities: A case study using a stream-dwelling char Salvelinus fontinalis. Biological Conservation, 194, 209-216. doi:10.1016/j.biocon.2015.12.023

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

Yoshida, T., Kawato, M., Fujiwara, Y., Nagano, Y., Tsuchida, S., & Yabuki, A. (2023). Optimization of environmental DNA analysis using pumped deep-sea water for the monitoring of fish biodiversity. Frontiers in Marine Science, 9. doi:10.3389/fmars.2022.965800

FIGURES CAPTIONS

Figure 1. Simulation of the species-specific detection probabilities as a function of the total number of DNA copies per each PCR aliquot (TOTDNA, left panel) and the number of reads per PCR (READS, right panels), per one, six and twelve PCR replicates. For each graph, lines represent the probability of detection of species belonging to one of the three relative abundance class of the total number of sequences (READS): black: ≥ 1 % (abundant species), blue: 1 ‰≤ <1 % (rare species), red: < 1 ‰ (very rare species). Horizontal line: detection probability of 0.95.

Figure 2. Effectiveness of species detection as a function of total DNA copy number per PCR (TOTDNA) and the residuals of the linear regression between the total number of reads per PCR (READS) and TOTDNA (READS_res). Isolines (red lines) along which the ratio of cumulative specific detection probabilities is 95 % of the total number of species for two (upper red curve) to twelve PCR replicates (lower red curve). Mean (solid line) and median (dashed line) values of TOTDNA. A: all species; B: very rare species (relative abundance of species-specific sequences < 1 ‰ of READS); C: rare species with (≥ 1 ‰ and <1 %); D: abundant species (≥ 1 %).

Figure 3. Effectiveness of species detection as a function of total DNA concentration per litre (TOTDNA_L) and volume /number of samples.
Figure 3 A to 3 C: Water volume versus TOTDNA_L for very rare species (relative abundance of species-specific sequences < 1 ‰), rare species (≥ 1 ‰ and <1 %), and abundant species (≥ 1 %) respectively. Isolines along which the ratio of cumulative specific detection probabilities is 95 % of the total number of detected species for six (orange) and twelve (black) PCR replicates.
Figure 3 D: Number of 20 L samples versus TOTDNA_L for very rare species (relative abundance of species-specific sequences < 1 ‰, red lines), rare species (≥ 1 ‰ and <1 %, green lines), and abundant species (≥ 1 %, blue lines) respectively.
Isolines along which the ratio of cumulative specific detection probabilities is 95 % of the total number of detected species for twelve PCR replicates.
Isolines are drawn for an effective sequencing depth corresponding to the mean (solid lines), and to the 2.5 % and 97.5 % percentiles of the READS_res distribution (dashed lines). READS_res is the residuals of the linear regression between READS and TOTDNA.

FIGURES

Figure 1

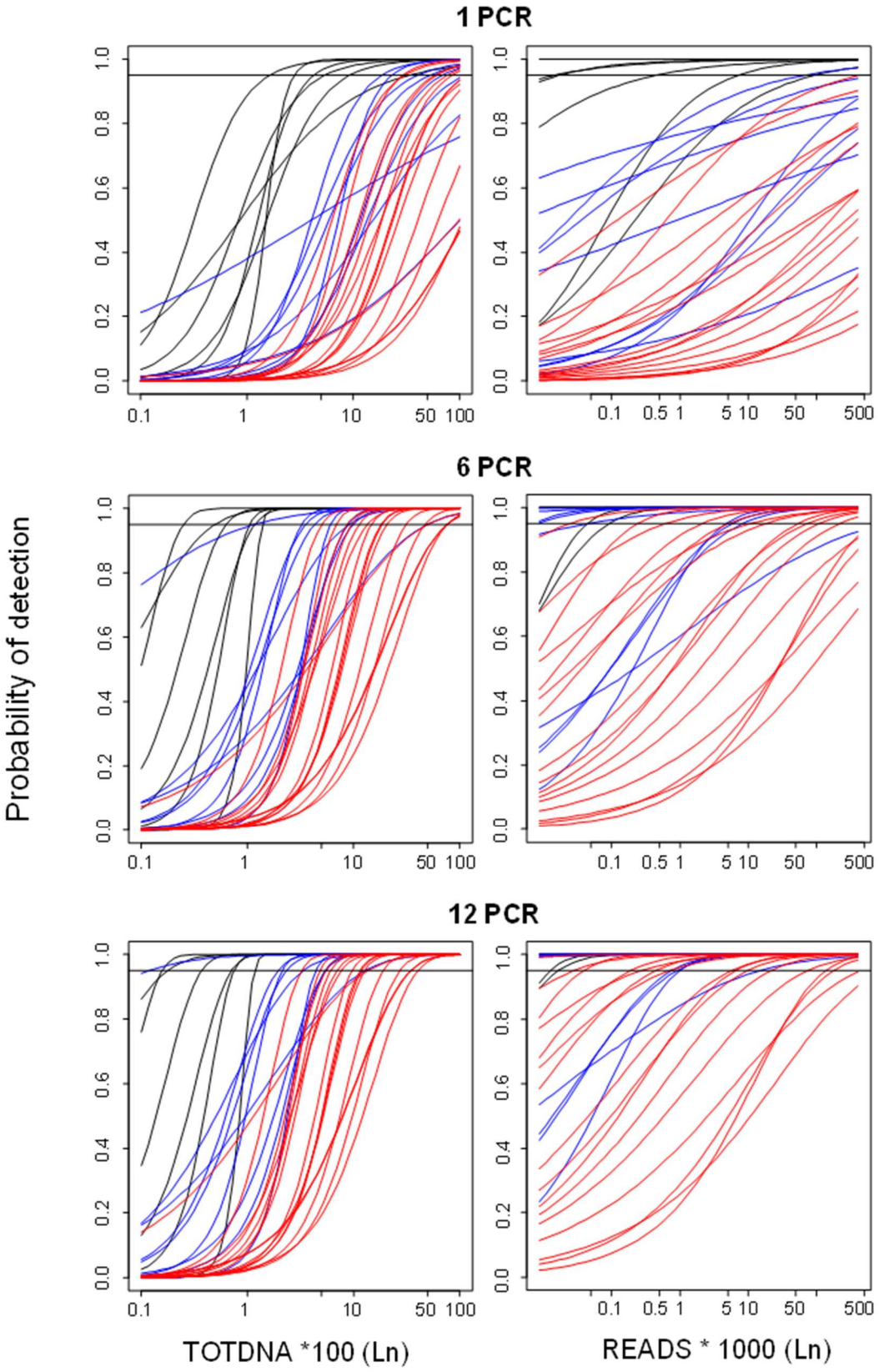


Figure 2

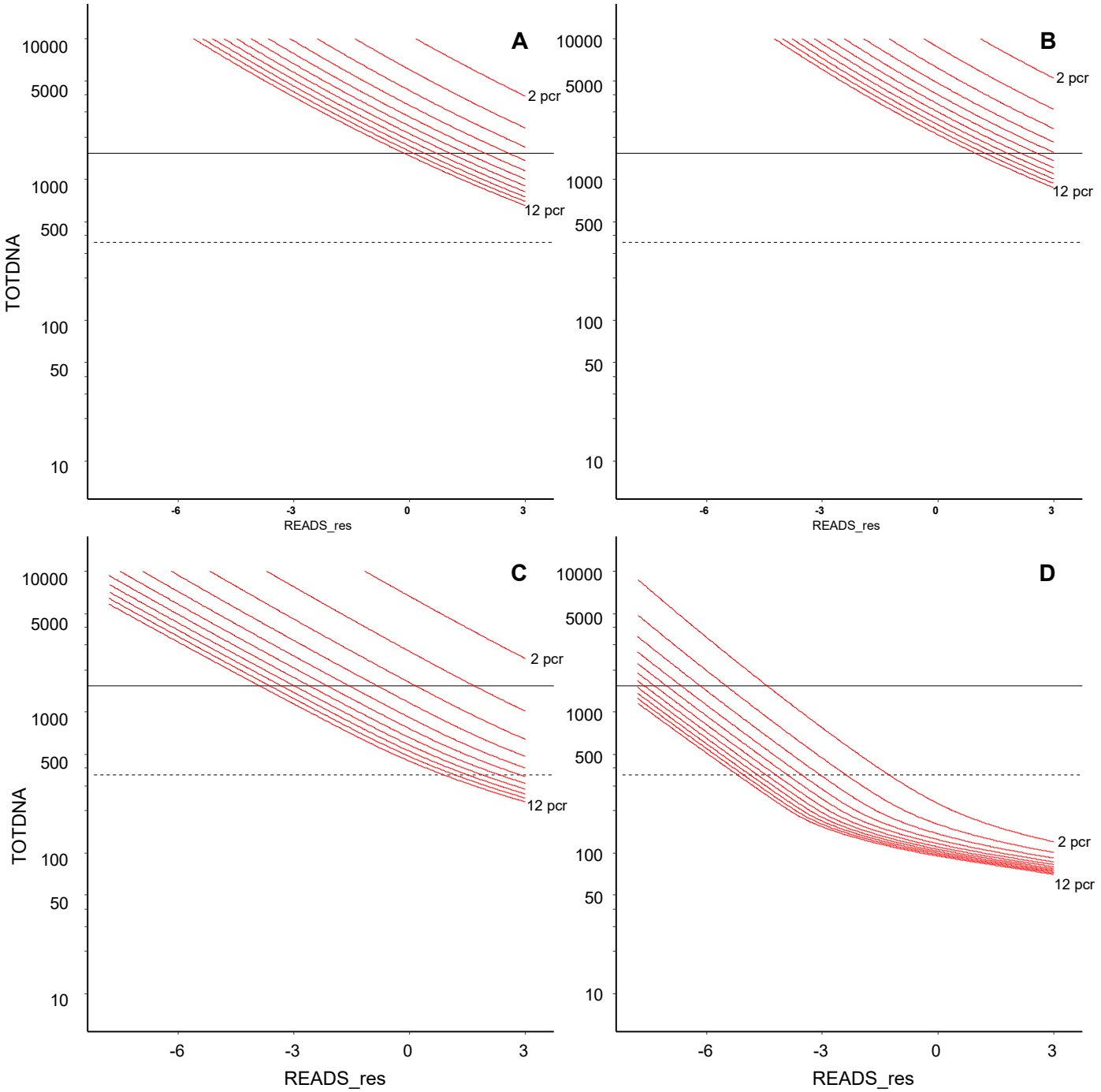
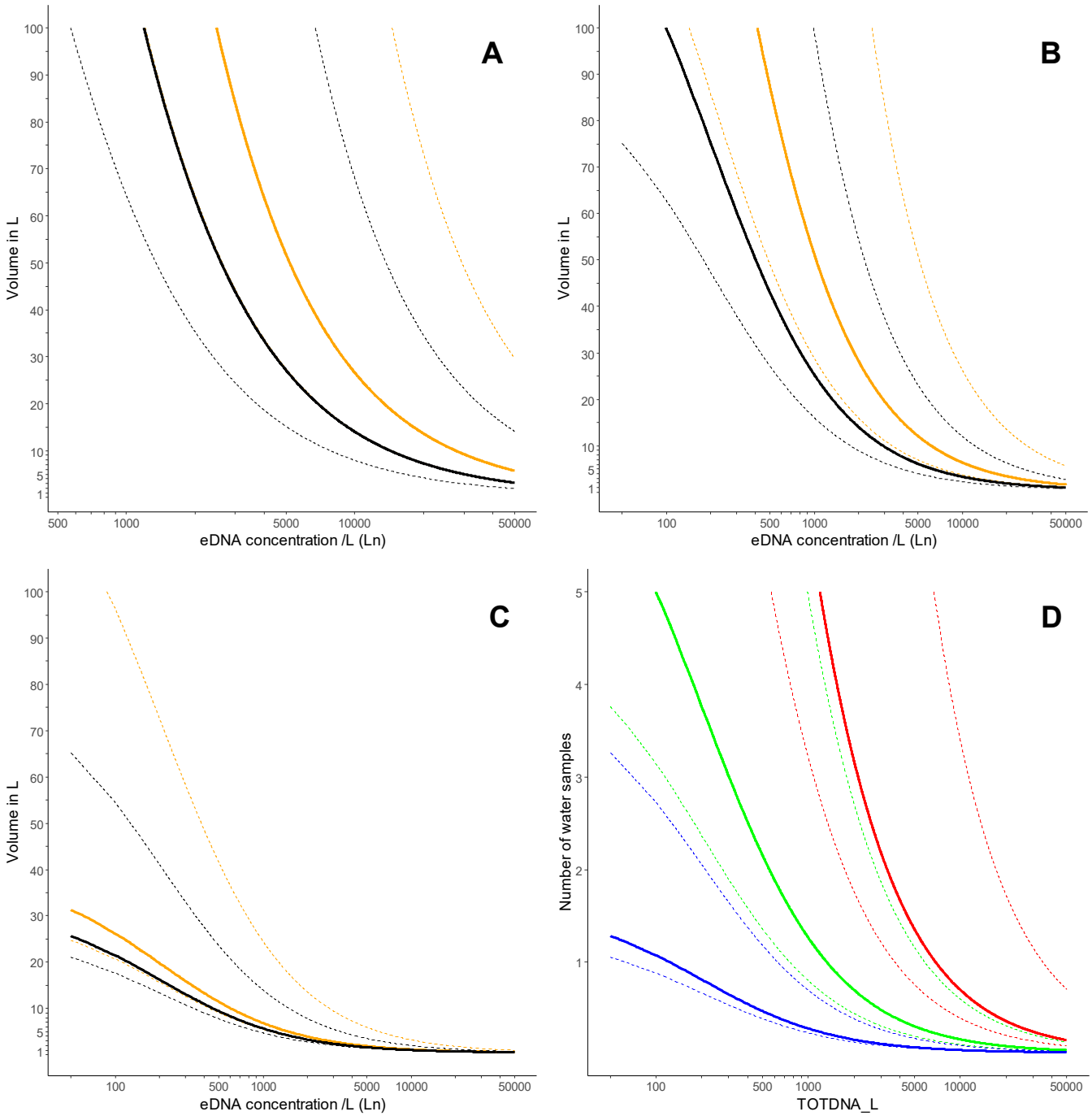

A
B
C
D
TOTDNA
READS_res
2 pcr
12 pcr
10000
5000
1000
500
100
50
10
-6
-3
0
3

Figure 3


A
B
C
D
Volume in L
eDNA concentration /L (Ln)
Number of water samples
TOTDNA_L
100
90
80
70
60
50
40
30
20
10
5
1
4
3
2
500
1000
5000
10000
50000