# Optimizing an eDNA protocol for monitoring endangered Chinook Salmon in the San Francisco Estuary: balancing sensitivity, cost and time.

Thiago M. Sanches[1*], Andrea M. Schreier[1]

Animal Sciences, University of California-Davis, Davis, California, US

*Corresponding Author

E-mail: tmsanches@ucdavis.edu

## Abstract

Environmental DNA (eDNA) analysis has gained traction as a precise and cost effective method for species and waterways management. To date, publications on eDNA protocol optimization have focused primarily on DNA yield. Therefore, it has not been possible to evaluate the cost and speed of specific components of the eDNA protocol, such as water filtration and DNA extraction method when designing or choosing an eDNA pipeline. At the same time, these two parameters are essential for the experimental design of a project. Here we evaluate and rank different eDNA protocols in the context of Chinook salmon (*Oncorhynchus tshawytscha*) eDNA detection in an aquatic environment, the San Francisco Estuary. We present a comprehensive evaluation of multiple eDNA protocol parameters, balancing time, cost and DNA yield. For estuarine waters, which are challenging for eDNA studies due to

high turbidity, variable salinity, and the presence of PCR inhibitors, we find that a protocol combining glass filters and magnetic beads, along with an extra step for PCR inhibitor removal, is the method that best balances time, cost, and yield. In addition, we provide a generalized decision tree for determining the optimal eDNA protocol for other studies on aquatic systems. Our findings should be applicable to most aquatic environments and provide a clear guide for determining which eDNA pipeline should be used for a given environmental condition.

# Author Summary

The use of environmental DNA (eDNA) analysis for monitoring wildlife has steadily grown in recent years. Though, due to differences in the ecology of the environment studied and the novelty of the technique, eDNA currently shows a lack of standards compared to other fields. Here we take a deep look into each step of an eDNA assay, looking at common protocols and comparing their efficiencies in terms of time to process the samples, cost and how much DNA is recovered. We then analyze the data to provide a concise interpretation of best practices given different project constraints. For the conditions of the San Francisco Estuary we suggest the use of glass fiber filtration, the use of paramagnetic beads for DNA extraction and the use of a secondary inhibitor removal. We expect our findings to provide better support for managers to decide their standards ahead of project submission not only for estuarine conditions but for other waterine conditions alike.

# Introduction

Environmental management policies rely heavily on measurements of the spatial distribution of habitat occupancy of species. In the past decade, environmental DNA (eDNA) has gained traction as one of the most sensitive and cost effective monitoring methods [1], allowing researchers to better estimate species occupancy rates in a given habitat. Due to high variability in the studied environments, currently,

there are no clear guidelines to assist investigators in choosing an optimal protocol for their particular eDNA monitoring studies.

In this study, we separate and optimize four important steps for eDNA biomonitoring of delta estuarine waters, which are characterized by elevated concentrations of solid suspended particles and fluctuating levels of salinity [2]. We comment on the specifics of each step for eDNA biomonitoring and develop a concise guideline to help determine which approach is most suitable for a set of scenarios. To lessen the burden of comparing DNA isolation methods on future investigators, we provide a framework that should help make more informed decisions, taking into account the specifics of their study requirements.

## The four main steps of an eDNA pipeline

The protocol to isolate environmental DNA from water samples can be described in four steps: filtration, DNA extraction, inhibitor removal and DNA amplification in order to estimate the initial concentration of eDNA [3]. In the filtration step, the water samples, with preferred volumes ranging from 50 mL to 1 liter in previous studies, are pressure-pumped through a membrane filter which captures the free DNA as well as tissue and cells suspended in the water. The next step is to extract DNA from the filter using conventional extraction methods, which were developed to isolate large nuclear DNA fragments from tissue. However, in the case of eDNA, it is preferable to target small fragments of mitochondrial genes as they have a higher copy number per cell compared to nuclear DNA. Then, as the filter may also capture high concentrations of PCR inhibitors, it is often necessary to use a secondary inhibitor removal step to further isolate the DNA from contaminants [4]. Last, the isolated DNA is amplified using quantitative PCR (qPCR) with primers specific to the target species, and the initial amount of the target eDNA is determined based on the Cq method. Although it is possible to use conventional PCR, this method significantly underperforms compared to qPCR in terms of sensitivity, and even when DNA is successfully amplified it provides less informative data [5].

## PCR inhibitors

PCR inhibitors are a group of substances that can inhibit PCR amplification. Their inhibiting mechanism varies between affecting the template DNA, the polymerase or other reagents necessary for the reaction. PCR inhibitors can be catalytic (e.g. proteases degrading proteins and phenol degrading DNA) or work through competitive binding (e.g. melanin forming a complex with the polymerase and humic acid interacting with the DNA template) [6]. Humic matter and proteases are typical PCR inhibitors present in high concentrations in turbid waters and other environmental samples [6,7].

## Estuarine waters and fish detection

In this study we optimize eDNA biomonitoring for estuarine waters, as this habitat is essential for the early developmental stages of several anadromous species, including our target species, Chinook salmon (*Oncorhynchus tshawytscha*). The estuarine environment provides a challenge for eDNA biomonitoring as the elevated density of solid particles, measured by turbidity levels, can bind to the DNA and clog the pores of the filters, limiting the volume of water that might be filtered. Also, estuaries have been shown to have elevated levels of PCR inhibitors [8,9]. Therefore, we assume that if our DNA amplification-based experiments work in these complex conditions, the same approach could also be applied to less turbid freshwater and marine conditions.

## Chinook Salmon as a target

We targeted Chinook Salmon in our experiments for a variety of reasons. First, as a widespread species in the North American Pacific Northwest, it has invaluable importance for the stability of the marine ecosystem of the region [10] and at the same time, provides a critical source of income for historic fishing communities [11,12]. In California, the Central Valley Spring-run, Fall-run and Winter-run Chinook salmon are listed as vulnerable while the Sacramento River winter-run and Central Valley late Fall-run are listed

as endangered [13]. Little is yet known about the spatio-temporal distribution and estuarine habitat usage of pre-smolt juvenile Chinooks during their annual out migration to the ocean. In this life-stage, juvenile Chinook salmons runs blend together and use this period to grow before leaving the estuarine environment. Larval survival in this stage has major implications to the population size of the species [14]. Conventional survey methods have encountered numerous difficulties to sample the rare juvenile Chinook in the marshland conditions of the SFE. Developing a high precision, high throughput eDNA protocol optimized for estuarine waters will allow managers to have a better understanding of the habitats used by Chinook in their early life-stages.

## Accounting for cost and time in experimental design

To decide on the most practical estuarine eDNA protocol, we first need to determine what it means for a protocol to be efficient. In our case we listed our priorities in the following order: 1) The eDNA yield must be adequately sensitive in realistic scenarios; 2) The protocol must be fast and scalable, and 3) The protocol must be cost effective, considering that reagent cost is the main driver of cost per sample. This order of priorities is influenced by several factors that include species abundance and costs. If the target species is ubiquitous and present at high densities, the DNA yield constraint can be loosened, allowing the use of faster and more cost-conscious protocols. If labor cost is inexpensive, choosing a more time intensive yet cheaper protocol will maximize the number of sampling points. On the other hand, in situations where labor accounts for much of the costs, choosing less time intensive protocols will allow more sampling points for the project.

# Results and Discussion

# Interference between filter type and extraction method is negligible

We first examined different methods for the initial two steps of an eDNA pipeline, filtration and DNA extraction, and tested for interference between these steps. In general, when optimizing a protocol consisting of several steps, it is important to identify if previous steps interfere with the effectiveness of subsequent steps. In our case the main possible interference is between the filter used and the extraction method. The yield percentage of a certain extraction method could change depending on which filter was used. Possible reasons for interference between filter and extraction method include different particles binding differentially to filters and extraction methods not isolating DNA from all types of particles at the same yield percentage. The models that we tested are the following:

Model with interference:

$$Y \sim Y_f(Y_e - interference(E|F)) - I_f I_e * \begin{cases} 0, \text{ if secondary inhibitor removal is used} \\ 1, \text{ otherwise} \end{cases} \quad (1)$$

Model without interference:

$$Y \sim Y_f Y_e - I_f I_e * \begin{cases} 0, \text{ if secondary inhibitor removal is used} \\ 1, \text{ otherwise} \end{cases} \quad (2)$$

For both widely-applicable information criterion (WAIC) and Leave-one-out cross validation (LOO) the model without interference was selected with a weight of 1 in both cases [17,20]. One form of visualizing the absence of interactions is that the ranking order of filter yield (1st *Cellulose nitrate* - *2nd Glass fiber* - *3rd Filter paper N1)* does not change independently of which extraction method is chosen (Fig 1A). Similarly, the yield ranking for extraction methods is not affected by filter choice. (Fig 1B). Only the NaOH

method breaks the independence rule for the nitrocellulose filter. In this case, target DNA could not be amplified from NaOH extractions without secondary inhibitor removal, resulting in an upwards skewed average of the DNA yield as the samples without secondary inhibitor removal were not taken into account.

**Figure 1: Relationship between total DNA yield and filtration and extraction protocols.** (A) No crossing between lines indicates that on average, DNA yield ranking for the filters is independent of the extraction protocol. Error bars represent the 95% confidence intervals. The combination of Whatman filter and NaOH extraction wasn't able to amplify the target Chinook DNA. (B) No crossing between lines indicates that on average, DNA yield ranking for the extraction method is independent of the filter type. The NaOH extraction protocol is the only case where the ranking order is not maintained and can be explained by the added effects of carry-on inhibitors.

The lack of interference between steps shows that for future optimization tests, it is not necessary to test all the possible combinations of filters and extraction methods at the same time. Instead one might test each section of the protocol independently and converge on the optimal method. Therefore, it is possible to test more methods for each step and increase the number of replicates for each test in future optimization experiments.

## The nitrocellulose filter can retain the most DNA per volume while the glass filter is the most resilient to high levels of turbidity

Next, we compared DNA yields from three different filters. The nitrocellulose filter outperformed the glass fiber filter in terms of DNA yield by 1.6 times and the Whatman n°1 filter by 3.75 times on average (Fig 2). In other words, 1.6L and 3.75L of water would need to be filtered through a glass fiber filter or Whatman n°1 filter respectively to isolate the same amount of DNA as filtering 1L of water through a nitrocellulose filter. However, the glass filter outperforms the nitrocellulose and Whatman filters in terms

of filtration time, with the glass filter not only being drastically faster but also more consistent and resilient to variations in turbidity (Fig 3). Therefore, we can conclude that for the context of estuarine waters, the glass filter is optimal in terms of DNA yield, speed, and cost.

**Figure 2: Distribution of DNA capture ratio for each filter type from the Automatic Differentiation Variational Inference model.** The broadness of the curve shows the variability of the ratio of the input DNA that binds to the filter. The peak of each distribution is the mean yield ratio of DNA recovery for that filter type. The nitrocellulose filter yielded the highest recovery ratio with little efficiency overlap compared to glass and Whatman filters.

**Figure 3: Percentiles and medium filtration time in order to filter 1L of estuarine water for each filtration method.** The glass filter outperforms nitrocellulose and Whatman by a significant margin in terms of average filtering time and consistency in the filtering time. Dots are filtration events while the black line represents the median value filtering time. Boxes indicate 10% quantiles.

## Qiagen DNA extraction is the most sensitive and reliable, paper extraction is the fastest and most cost effective, and magnetic beads is the most balanced method.

All extraction methods could yield enough eDNA to be detectable by qPCR amplification. The Qiagen DNEasy kit had the highest DNA yield, outperforming NaOH by 1.7 times, magnetic beads by 2.26 times, direct to qPCR dipsticks by 9.71 times and regular dipsticks by 358 times (Fig 4). At the same time, the Qiagen kit is by a considerable margin the most time-consuming method, requiring 77 minutes to process 18 samples. In contrast, the direct to qPCR dipstick approach was the fastest and most cost-efficient method by a wide margin. Currently, the major bottleneck of our experiments is the time required to process the samples. Yet, subsequent tests have shown that the use direct dipstick extraction drastically

lower the probability of amplification in cases where the species of interest is rare. Therefore, we consider the magnetic beads to be the optimal method for estuarine waters.

We estimated our costs for the most-used DNA extraction kits. Alternative kits might be used in order to reduce costs. As an example, Ampure XP is 100 times more expensive than making a magnetic beads solution in-house [21], though this cost reduction is at the expense of lower reproducibility and therefore not optimal for certain projects. Buying in bulk is also other alternative to reduce costs, though that might be limited to initial funding of the project.

**Figure 4: Modelled distribution of percentage yield for each extraction protocol.** The width of each of the curves shows the variability in modelled yield. The peak of the distribution is the mean yield per extraction type. QIagen DNeasy yields the best yield with little overlap with other methods. Meanwhile magnetic beads and NaOH have shown similar distributions with significant overlap, while both dipstick methods underperform the other methods. It is important to note that this plot does not take into account PCR inhibitor carryover, which might vary significantly between methods.

## Yield is mostly dependent on extraction method

The extraction method was shown to be the most influential factor for the eDNA yield from the random forest aggressor analysis (Fig 5). Therefore, further optimization experiments should focus on this step, experimenting with different protocols to extract the eDNA in order to maximize protocol eDNA yield. Meanwhile, in the context of our experiments, the removal of inhibitors was shown to have little impact to the total DNA yield estimated by qPCR, although published data [4] have shown that inhibitor removal highly influences the amplification probability of the qPCR reaction.

**Figure 5: Influence of each eDNA protocol segment to total eDNA yield ratio estimate.** The extraction method was the factor that had the highest influence on the total eDNA yield of the total protocol, while inhibitors didn't have a significant impact compared to the other sections of the total protocol.

Our experiments have shown comparatively little influence of the use of secondary inhibitor removal to the total eDNA yield. Though this result contradicts at some degree our pilot study for estuarine juvenile chinook as well as other published data. Causes for this discrepancy might be resultant of several nonexclusive factors. First inhibitors in general work by binding to DNA strands and not as catalysts, therefore if the ratio between eDNA:inhibitors is significantly elevated, as we would expect in tank experiments, we would expect minimal effects of the inhibitors. Another possibility is that at the sampling location, on the sampling date and time, there were fewer inhibitors than usually observed in estuaries. Another possible explanation is that the yield variance between the extraction methods and filters surpass the yield variance due to the inhibitor removal step, which doesn't mean this step won't significantly influence the DNA yield of the experiment. Last, PCR inhibitors might not affect the eDNA retrieval but only the probability of amplification. This last observation might also explain why the probability of amplification and DNA yield aren't always fully correlated. Therefore, considering this experiment results and previous findings we consider that secondary inhibitor removal is advised if possible as it improves the DNA yield and amplification probability in the context of estuarine samples.

## Effects of secondary inhibitor removal varies between filters and extraction method

We observed that the secondary inhibitor removal step always outperformed skipping this step. Regressions from Figs 6A and 6C were always positive and the distributions from Figs 6B and 6D were always greater than zero. The nitrocellulose filter and the NaOH extraction were the methods that carried

240 the most PCR inhibitors, while the other methods for each step showed a high overlap of their carryover
241 inhibitor distributions. Secondary inhibitor removal was essential to observe any amplification using the
242 NaOH extraction method, which also suggested that this method is inefficient at removing PCR inhibitors.
243
244 **Figure 6: Effects of adding a secondary inhibitor removal step to the eDNA estimation protocol.**
245 (A-B) DNA yield variation of using a OneStep PCR Inhibitor Removal©. The nitrocellulose filter produced
246 the highest inhibitor carryover levels at the same time it captured the highest percentage of free eDNA.
247 This suggested that the nitrocellulose filter captured particulates with indiscriminately with high efficiency
248 (C-D) Estimated distributions for inhibitor carryover for filter and extraction method. Aside from NaOH
249 extraction, other methods had similar distributions of carryover PCR inhibitors with high overlap.
250 Therefore, NaOH extraction, even if it has an elevated eDNA yield, doesn't properly address the high
251 levels of PCR inhibitors commonly encountered in environmental samples.
252
253 In most cases, using a glass fiber filter and magnetic beads would be the most practical method
254 to generate the maximum amount of information obtained about fish distribution given the constraints of
255 our study. Our experiment suggests that DNA extraction from the filters is the most time-consuming step
256 and most variable in terms of efficiency; therefore, this is the step which should be decided with utmost
257 care in order to maintain the high-throughput and useful detection limit of the desired methodology. For
258 this reason, magnetic beads DNA extraction in a promising alternative to silica column extraction, as this
259 method strikes the balance between yield, amplification probability, carryover PCR inhibitors and time to
260 process samples. Meanwhile the cost of using magnetic beads can be mitigated by developing the
261 necessary reagents in-house. Though in specific cases different pipelines might yield better results. For
262 those scenarios, we constructed a simple decision tree for choosing the best methodology various
263 possible study for each scenario (Fig 7). We also ranked the pipelines, sorting them by DNA yield, which
264 should be the main parameter for the pipeline selection. Then, once established which pipelines have a
265 DNA yield that fits the project, balancing time and cost of the pipelines (Fig 8).

**Figure 7: Decision tree for choosing the protocol which will yield the most information given research constraints.** A glass filter is recommended in most cases, as long as the focus isn't maximizing DNA yield with no time or cost constraints. Magnetic beads also are advised in general for its balance between DNA yield and time to process the samples, while cost can be mitigated by producing magnetic beads solution in-house (~$0.55/mL) instead of buying Ampure XP ($15–$70/mL) [21].

**Figure 8: Comparison between eDNA protocols for DNA yield, cost and time to process 96 samples.** Methods were sorted by yield and shown in $\log_{10}$ scale.

# Methods

## Ethics statement

We sampled water in accordance with the University of California Davis Institutional Animal Care and Use Committee (USDA registration: 93-R-0433, PHS Animal Assurance A3433-01) under the protocol number #20608.

## Experimental Design

We tested three biological replicates for every combination of filter, extraction method and inhibitor removal and measured the amount of recovered eDNA using qPCR Cq values and a DNA standard curve, obtained from a fin clip serial dilution on the same plate (Fig 1). Input DNA estimation is described in the methods section. We defined an equation that describes how the efficiency of each step influences the total amount of recovered eDNA:

$$Y \sim Y_f Y_e(E|F) - I_f I_e * \begin{cases} 0, \text{ if secondary inhibitor removal is used} \\ 1, \text{ otherwise} \end{cases} \quad (3)$$

where:

$Y$:ratio of input eDNA that was amplified by the qPCR

$Y_f$:ratio of input eDNA that binds to filter

$Y_e(E|F)$:ratio of eDNA bound to the filter that is isolated by the extraction method

$I_f$:filter inhibitor carryover

$I_e$:extraction method inhibitor carryover

$I_f I_e$:ratio of input eDNA not available to amplification due to inhibitors

**Figure 9: Scheme of steps for an eDNA protocol with tested methods for each step**. Cost and processing times (in minutes) of each method are shown next to the method name. Number of samples for the measured times varies as the number of samples that can be run in parallel varies between steps. Costs are estimated per sample.

Then, based on this equation we used Automatic Differentiation Variational Inference (ADVI) [15] to estimate the distribution of the parameters that maximize the likelihood of the observed yields.

# Sampling

To replicate realistic water conditions in terms of salinity, temperature and turbidity while also controlling the presence and amount of Chinook DNA, we combined water samples from a tank containing a high density of juvenile Chinook with an estuarine water sample from a representative location of pre-smolt Chinook habitat in the San Francisco Estuary. The estuarine water biological replicates consisted of 500mL of surface water taken with a 1L measuring cup (sterilized by rinsing in 20% bleach solution and then rinsing in DI water) from Suisun Bay, California (38°11'16.7"N 121°58'34.5"W ) and collected in a 1L

313 Nalgene bottle. Next, using another sterile measuring cup, we added 500ml of tank water known to hold
314 Chinook salmon DNA to each estuarine water biological replicate. The 680L tank contained 906 Chinook
315 salmon of approximately 11 cm in length. This mixture allowed us to both control Chinook density and
316 observe similar PCR inhibitor levels as those observed in the estuary. In total, we produced 85 samples,
317 which including a deionized water sample control, a tank water only control and a Suisun Bay water
318 control.
319

320 # Estimation of average input eDNA

321 To estimate the average input DNA from the tank water we spiked 10 samples of 1L surface water
322 from the Suisun Bay, California (38°11'16.7"N 121°58'34.5"W) with varying concentrations of isolated
323 Chinook and green sturgeon (*Acipenser medirostris*) DNA totaling 3 samples with 1ng/L, 3 samples with
324 0.1ng/L and 3 samples with 0.01ng/L for the Chinook salmon samples and 3 samples with 10ng/L, 3
325 samples with 1ng/L and 3 samples with 0.1ng/L for green sturgeon. We tested for green sturgeon
326 concomitantly to validate the protocol in a multispecies manner and verify that probe specificity and
327 detection limit doesn't affect the DNA yield of the protocol. The last sample was not spiked and used as a
328 negative control. Then we filtered the samples using a glass filter, extracted the DNA using the Qiagen
329 DNeasy Blood & Tissue Kit (Cat No./ID: 69504) and removed PCR inhibitors using Zymo OneStep™ (Cat
330 No./ID: D6030). Our serial dilution consisted of the same extracted DNA solution used to spike the
331 samples. We estimated the average yield in percentage for this protocol by qPCR amplification. From the
332 Qiagen protocol average yield we could estimate the average input DNA from the tank water. We also
333 estimated that pipelines DNA yield percentage is mildly inverse correlated (p-value = 0.0023) to the initial
334 DNA concentration (Fig S1), while the probability of amplification is logistically correlated to the $Log_{10}$(initial
335 DNA concentration) (Fig S2).
336

337 # Filtration

We filtered the samples one day after sampling to simulate real conditions, where it is not always possible to complete filtration on the same day as sampling. In each filtration run, 4 samples of 1L were filtered in parallel at the speed of 310rpm on a peristaltic pump and we timed each filtration event. Filters were folded in half 3 times and stored in a 2mL microcentrifuge tube and stored at -20°C. Between runs, tubing and casing were sterilized using a bath of 20% bleach [16], rinsed twice using DI water to remove any remaining bleach and dried.

# Extractions

For all DNA extraction protocols except the dipstick-based ones, we added 180µL of ATL buffer and 20µL of 5U Proteinase K to the microcentrifuge tube and incubated at 56°C overnight using a rotisserie attachment for 2mL microcentrifuge tubes. As the incubation time step doesn't require labor, we didn't add it to the total time of the protocol. Next, the filter was compressed inside of the microcentrifuge tube using a pipette tip and the supernatant was transferred to a clean 0.5mL (NaOH extraction) or 1.5 mL microcentrifuge tube (magnetic beads and QIagen).

## NaOH-based extraction

For each 100uL of supernatant we added 5.26µL of 1M NaOH. In a benchtop thermocycler, we incubated the samples at 95°C for 20min and ramped down the temperature at a pace of 0.7°C/min until reaching 4°C. Next, we added 10% of the total volume of 1M Tris-HCL. Samples were vortexed and centrifuged for 15min at 4680rpm. Without disturbing the pellet, 100µL of the supernatant was extracted and transferred to a new 1.5mL microcentrifuge tube and stored at -20°C.

## Magnetic Beads

For each sample, 180µL of Agencout AMPure XP (Beckman Coulter™; Cat No./ID: A63881) was added to the solution and incubated at room temperature for five minutes. Then the microcentrifuge tubes

were placed onto the magnetic plate (DynaMag™-2; Cat No./ID: 123.21D) for 2 minutes. We removed the supernatant and washed the magnetic beads twice using 200µL of a freshly made 70% ethanol solution with an incubation time of 30s in the magnetic plate. We then air dried the beads for 3minutes. A total of 100µL of TE solution was used to resuspend the particles and elute the DNA. The solution was incubated for 1minute at room temperature before pulling down the magnetic beads with the plate for 2minutes. Lastly, the supernatant was transferred to clean 1.5mL microcentrifuge tubes.

## Qlagen DNeasy cell and tissue

The Qlagen DNeasy extraction was performed following the manufacturer's recommendations. A total of 200µL of AL buffer was added and the samples were incubated at 56°C for 10minutes. We added 200µL of ethanol to each sample and the solution was transferred to the column and centrifuged at 8000rpm for 3minutes. Then the column was washed using 500µL of Wash Solution N°1 and centrifuged at 6000rpm for 1minute. Then the column was washed again with Wash Solution N°2 and centrifuged for 3minutes at 1400rpm. Next 100µL of AE solution was added and incubated for 20minutes before centrifuging at 8000rpm for one minute. The flowthrough was then stored at -20°C

## Whatman paper dipstick

Dipsticks were made following the protocol described in [17]. We used the qualitative Whatman filter n°1 to make our dipsticks and used an effective surface area of $8mm^2$ (2mm width and 4mm height). We added 200µL of lysis buffer and ground the filter using a pipette tip until the filter was dissolved. Then we dipped the dipstick in the lysis buffer solution (20mM Tris [pH 8.0], 25mM NaCl, 2.5mM EDTA, 0.05% SDS) 3 times, then dipped 3 times in 100µL of wash solution (10mM Tris [pH 8.0], 0.1% Tween-20), and 3 times in a final solution of nuclease free water which then was stored at -20°C. In the case of “straight to qPCR” dipstick extraction, we directly dipped the dipstick after the wash step into the qPCR reaction.

## Secondary Inhibitor removal

Zymo OneStep™ PCR Inhibitor Removal Kit (Zymo Research; Cat No./ID: D6030) was used following the manufacturer's protocol in order to remove any carryover PCR inhibitors from previous steps. We added 600µL of Prep-solution to the column and centrifuged at 8000g for 3 minutes, the flow through was discarded, then 50µL of DNA elute from previous steps were added to the column and centrifuged at 16000g. The flowthrough was then stored at -20°C.

## qPCR amplification oligos

Quantitative PCR detection for Chinook was developed by adapting the protocol from [18]Reaction solution totaling 20µL was composed of 1× TaqMan™ Environmental Master Mix 2.0 (ThermoFisher Scientific; Cat No./ID: 4396838), 0.9µM concentration of each primer, and 0.7µM of the Taqman probe, and 6µl isolated DNA extract from previous steps. Thermocycling was performed on a Bio-Rad CFX96 real-time detector using the following profile: 10 min at 95°C, 40 cycles of 15s denaturation at 95°C and 1 min annealing–extension at 60°C.

**Table 1: List of used DNA oligonucleotides.**

| Oligonucleotide | Sequence |
|---|---|
| Probe sequence | FAM-5'-AGCACCCTCTAACATTTCAG-3'-ZEN / Iowa Black FQ |
| Forward primer | 5'-CCTAAAAATCGCTAATGACGCACTA-3' |
| Reverse primer | 5'-GGAGTGAGCCAAAGTTTCATCAG-3' |
| Gblock sequence | 5'-ACCATCGTTGTTATTCAACTACAAGAACCT |

| | |
|---|---|
| | AATGGCCAACCTCCGAAAAACCCATCCTCT<br>CCTAAAAATCGCTAATGACGCACTAGTCGA<br>CCTCCCAGCACCCTCTAACATTTCAGTCTG<br>ATGAAACTTTGGCTCACTCCTAGGCCTATG<br>TTTAGCCACCCAAATTCTTACCGGGCTCTT<br>CTTAGCCATACACTATACCT-3’ |

Primers were used for DNA amplification. Probe was used for the qPCR step for DNA quantification. Gblock was used for creating a standard ladder for the qPCR reaction and made possible the conversion from initial DNA concentration to copy number.

## Data analysis

Data analysis was performed in Python 3.7 and is available on https://github.com/sanchestm/eDNA-Protocol-Optimization. We measured interference between filter type and extraction method using two competing models, one that includes the interference effect and one that does not. Using ADVI inference we fitted the data to the models [15]. From the ADVI fitting for the best model we estimated the distribution of filter eDNA yield percentage (Fig 2), extraction eDNA yield percentage (Fig 4) and PCR inhibitor carryover for filtration and extraction (Fig 6). To estimate which step of an eDNA experiment has the most variance between methods, and therefore can lead to the most significant gains when optimized, we trained a random forest regressor [19] with the collected data and estimated importance of each step of the experiment.

PHS/NIH Assurance (A3433-01)

# References

1. Kelly RP, Port JA, Yamahara KM, Martone RG, Lowell N, Thomsen PF, et al. Environmental monitoring. Harnessing DNA to improve environmental management. Science. 2014;344: 1455–1456.
2. Schoellhamer DH. Influence of salinity, bottom topography, and tides on locations of estuarine turbidity maxima in northern San Francisco Bay. Proceedings in Marine Science. 2000. pp. 343–357.
3. Hunter ME, Oyler-McCance SJ, Dorazio RM, Fike JA, Smith BJ, Hunter CT, et al. Environmental DNA (eDNA) sampling improves occurrence and detection estimates of invasive burmese pythons. PLoS One. 2015;10: e0121655.
4. Williams KE, Huyvaert KP, Piaggio AJ. Clearing muddied waters: Capture of environmental DNA from turbid waters. PLoS One. 2017;12: e0179282.
5. Piggott MP. Evaluating the effects of laboratory protocols on eDNA detection probability for an endangered freshwater fish. Ecol Evol. 2016;6: 2739–2750.
6. Schrader C, Schielke A, Ellerbroek L, Johne R. PCR inhibitors - occurrence, properties and removal. J Appl Microbiol. 2012;113: 1014–1026.
7. Gregory JB, Litaker RW, Noble RT. Rapid one-step quantitative reverse transcriptase PCR assay with competitive internal positive control for detection of enteroviruses in environmental samples. Appl Environ Microbiol. 2006;72: 3960–3967.
8. Nielsen KM, Calamai L, Pietramellara G. Stabilization of Extracellular DNA and Proteins by Transient Binding to Various Soil Components. Soil Biology. pp. 141–157.
9. Matheson CD, Gurney C, Esau N, Lehto R. Assessing PCR Inhibition from Humic Substances. Open Enzym Inhib J. 2014;3: 46–52.
10. Vélez-Espino LA, Antonio Vélez-Espino L, Ford JKB, Andres Araujo H, Ellis G, Parken CK, et

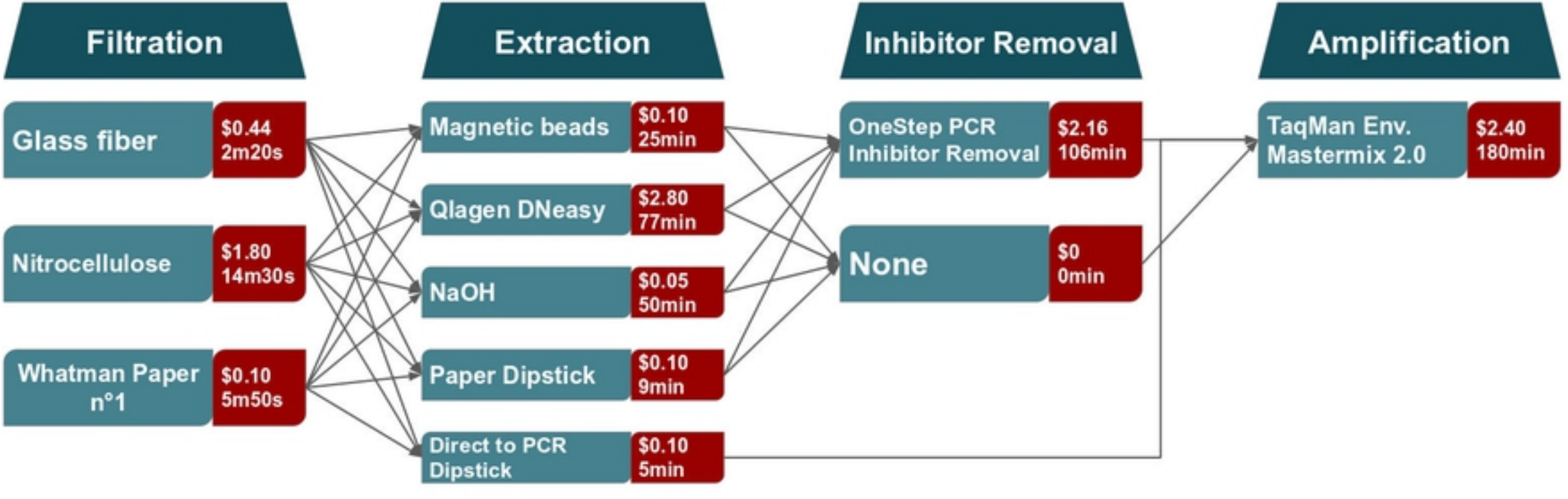
al. Relative importance of chinook salmon abundance on resident killer whale population growth and viability. Aquat Conserv. 2014;25: 756–780.

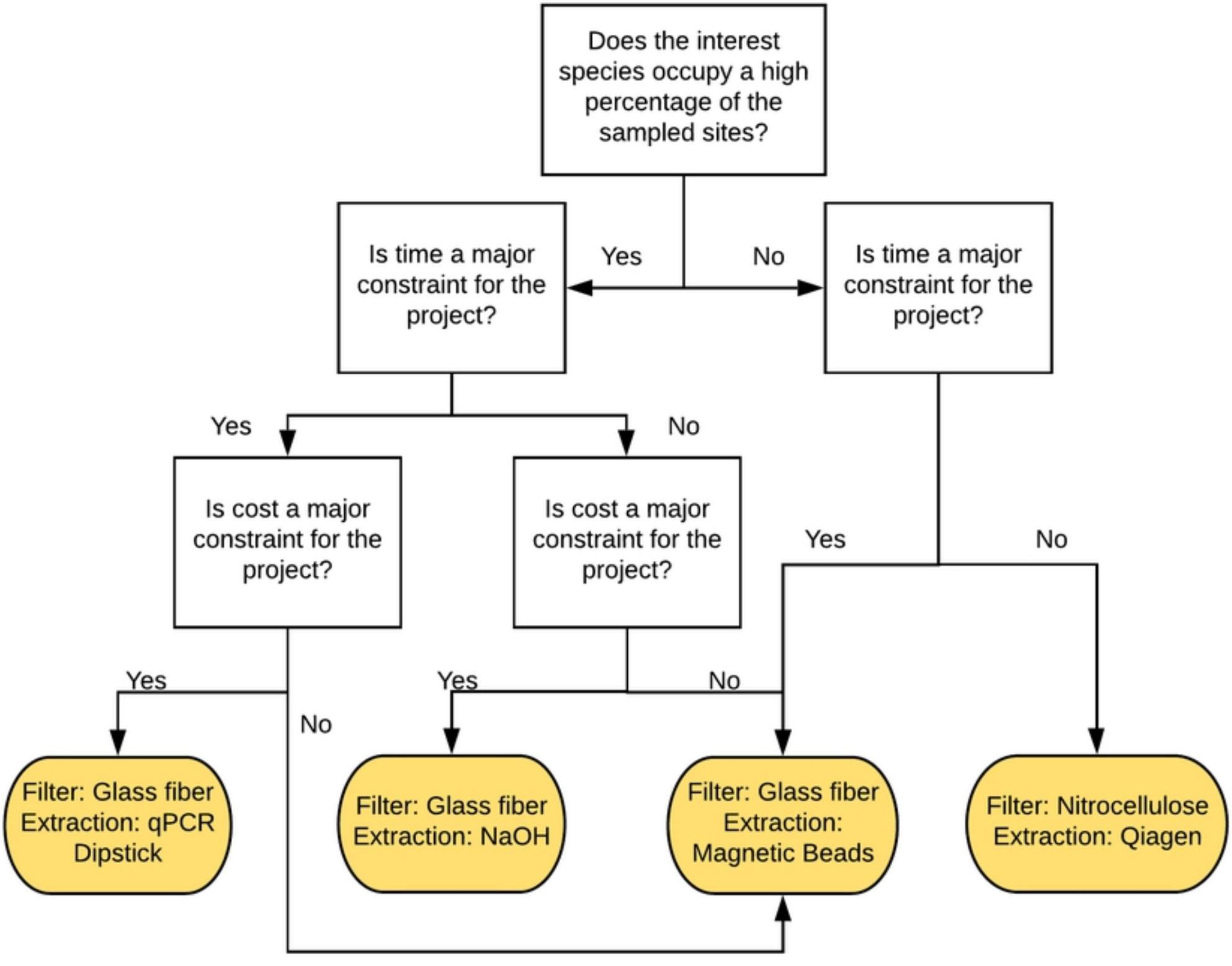
11. Yoshiyama RM, Fisher FW, Moyle PB. Historical Abundance and Decline of Chinook Salmon in the Central Valley Region of California. N Am J Fish Manage. 1998;18: 487–521.
12. Yoshiyama RM. A History of Salmon and People in the Central Valley Region of California. Rev Fish Sci. 1999;7: 197–239.
13. Katz J, Moyle PB, Quiñones RM, Israel J, Purdy S. Impending extinction of salmon, steelhead, and trout (Salmonidae) in California. Environ Biol Fishes. 2012;96: 1169–1186.
14. Helfman G, Collette B, Facey D. The Diversity of Fishes. Wiley-Blackwell; 1997.
15. Salvatier J, Wiecki TV, Fonnesbeck C. Probabilistic programming in Python using PyMC3 [Internet]. 2016. doi:10.7287/peerj.preprints.1686
16. Wilcox TM, McKelvey KS, Young MK, Sepulveda AJ, Shepard BB, Jane SF, et al. Understanding environmental DNA detection probabilities: A case study using a stream-dwelling char Salvelinus fontinalis. Biol Conserv. 2016;194: 209–216.
17. Zou Y, Mason MG, Wang Y, Wee E, Turni C, Blackall PJ, et al. Nucleic acid purification from plants, animals and microbes in under 30 seconds. PLoS Biol. 2017;15: e2003916.
18. Brandl S, Schumer G, Schreier BM, Conrad JL, May B, Baerwald MR. Ten real-time PCR assays for detection of fish predation at the community level in the San Francisco Estuary-Delta. Mol Ecol Resour. 2015;15: 278–284.
19. Geurts P, Ernst D, Wehenkel L. Extremely randomized trees. Mach Learn. 2006;63: 3–42.
20. Vehtari A, Gelman A, Gabry J. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. Stat Comput. 2016;27: 1413–1432.
21. DeAngelis MM, Wang DG, Hawkins TL. Solid-phase reversible immobilization for the isolation of PCR products. Nucleic Acids Res. 1995;23: 4742–4743.
22. *Anaconda Software Distribution*. Computer software. Vers. 2-2.4.0. Anaconda, Nov. 2016. Web. <https://anaconda.com>.

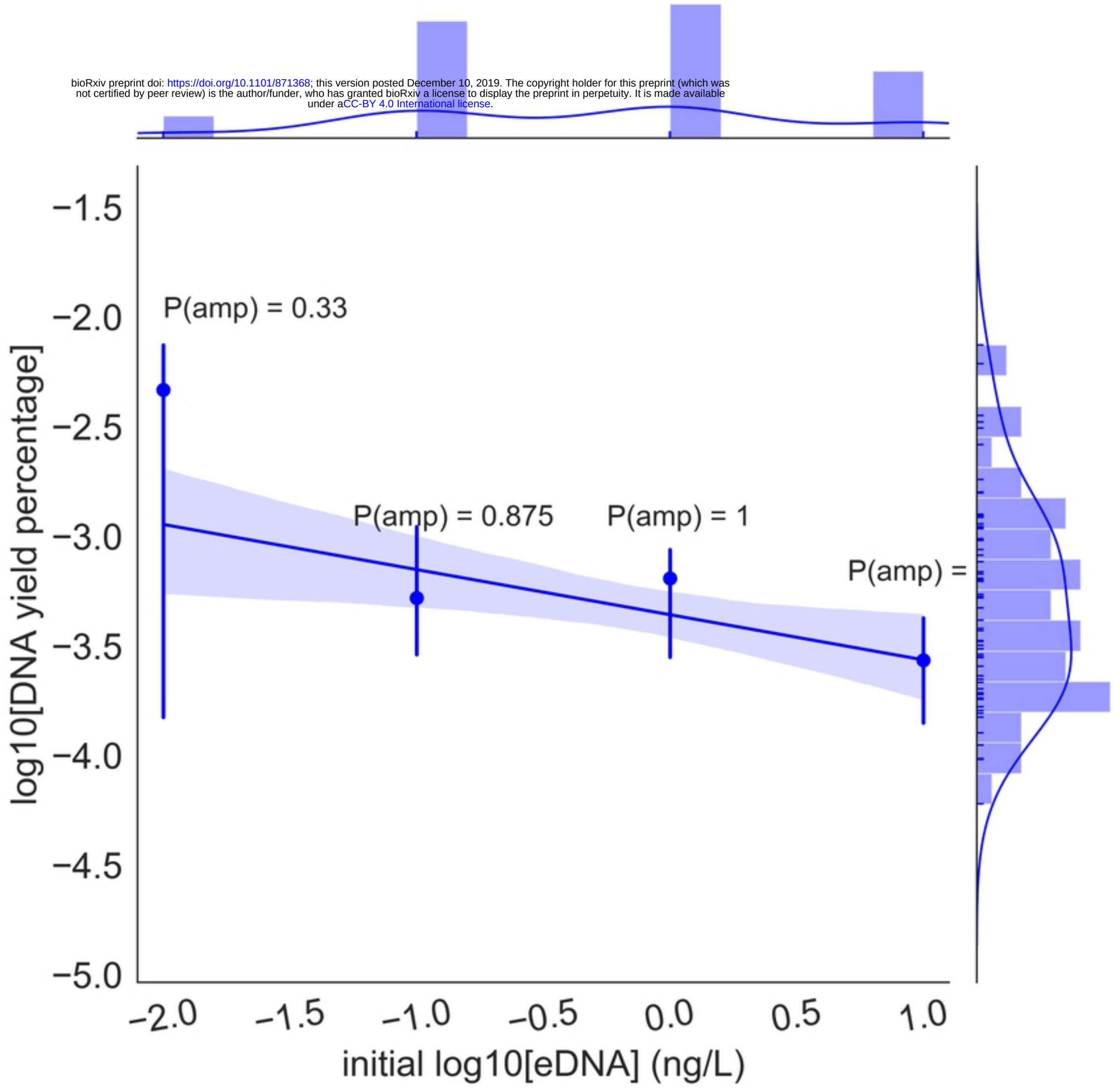
# Supporting information

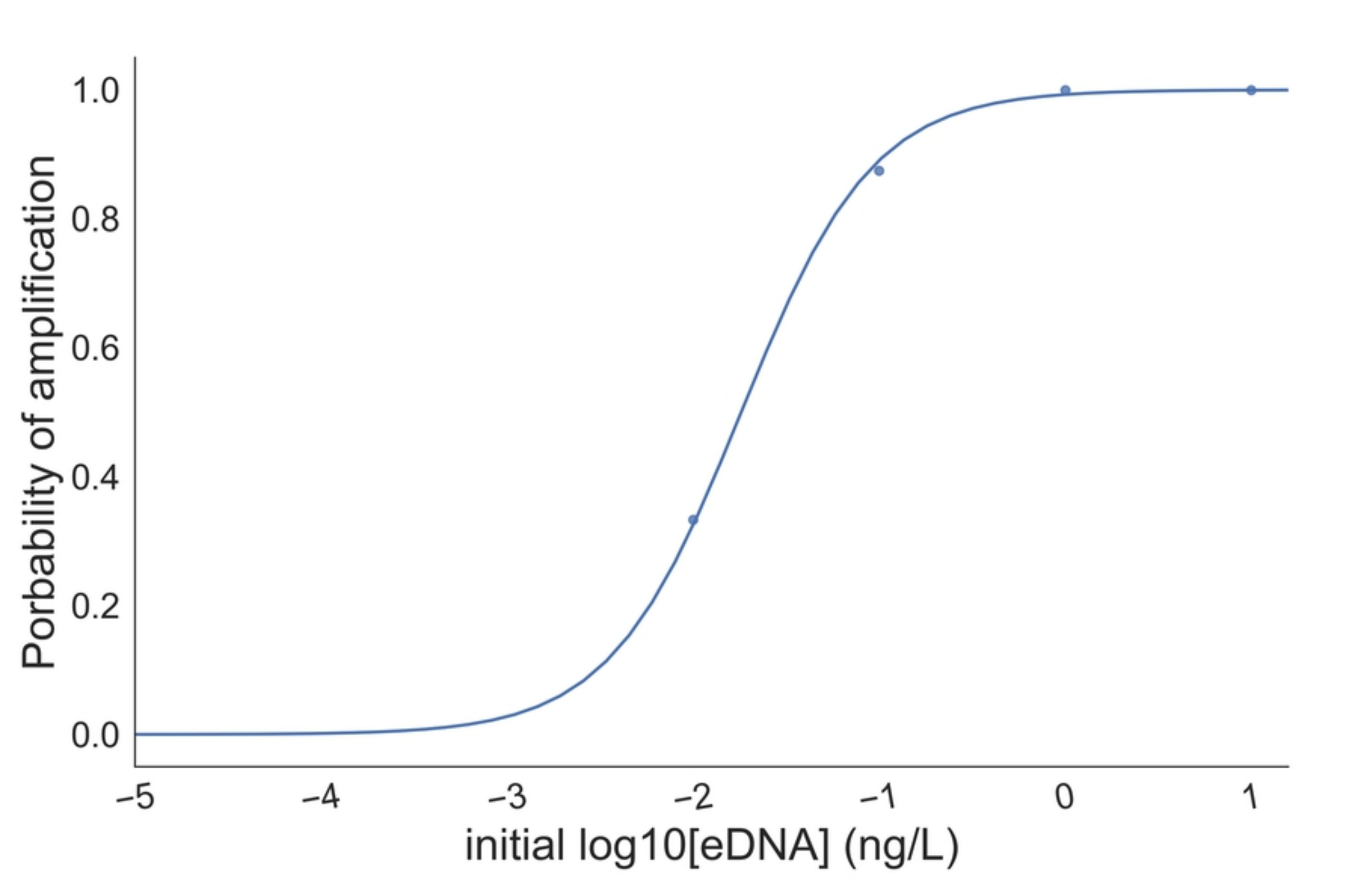
**Figure S1: Correlation between DNA yield and initial eDNA concentration.** Blue dots - median value; vertical lines - 95% CI; horizontal line - linear regression between protocol DNA yield and input DNA, with both axes being represented in $\log_{10}$ scale. P(amp) - probability of amplification.

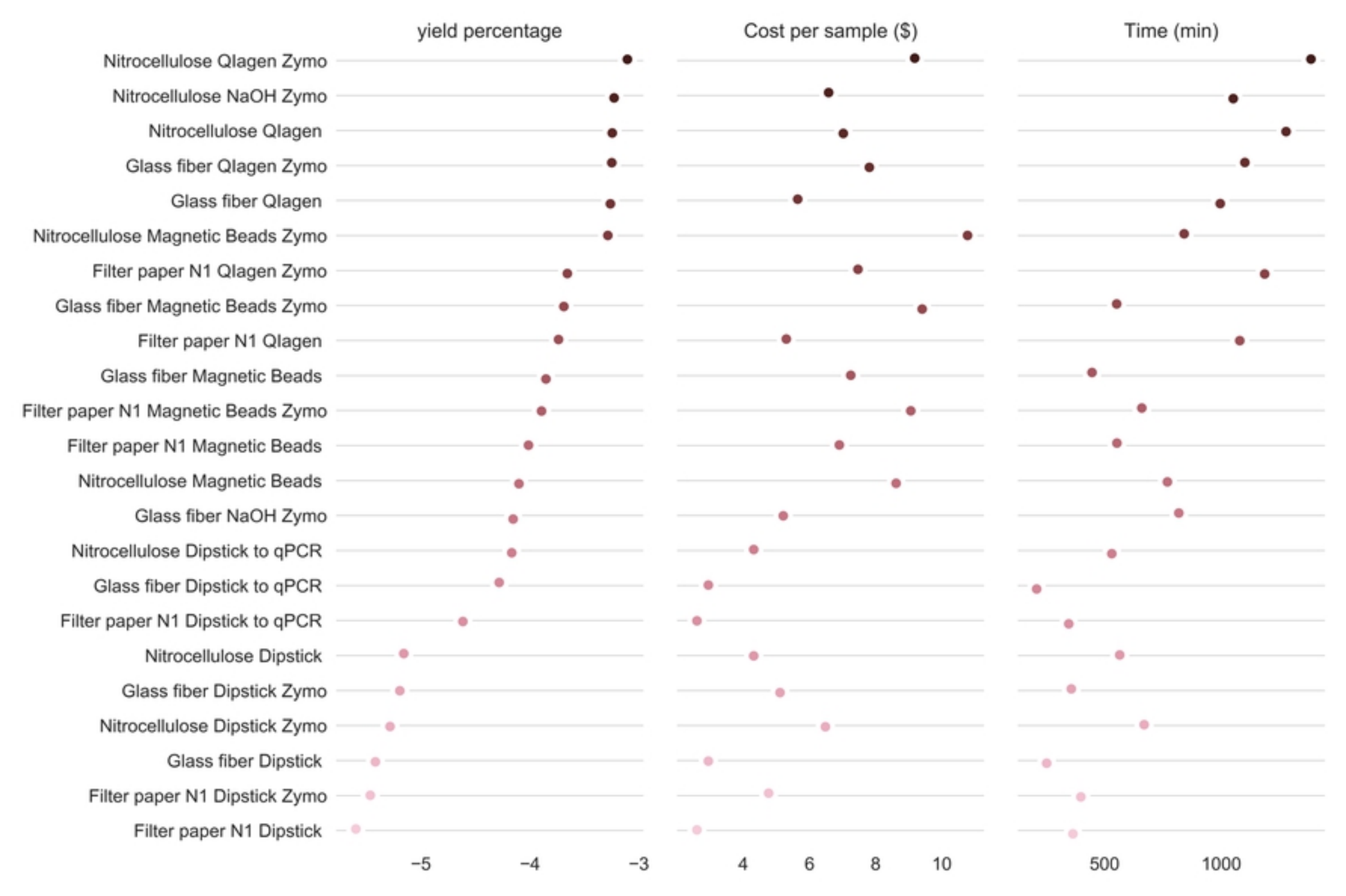
**Figure S2: Probability of amplification as a function of the input DNA concentration.** Dots - probability of amplification from DNA spiking experiment; line - logistic fit to data points.

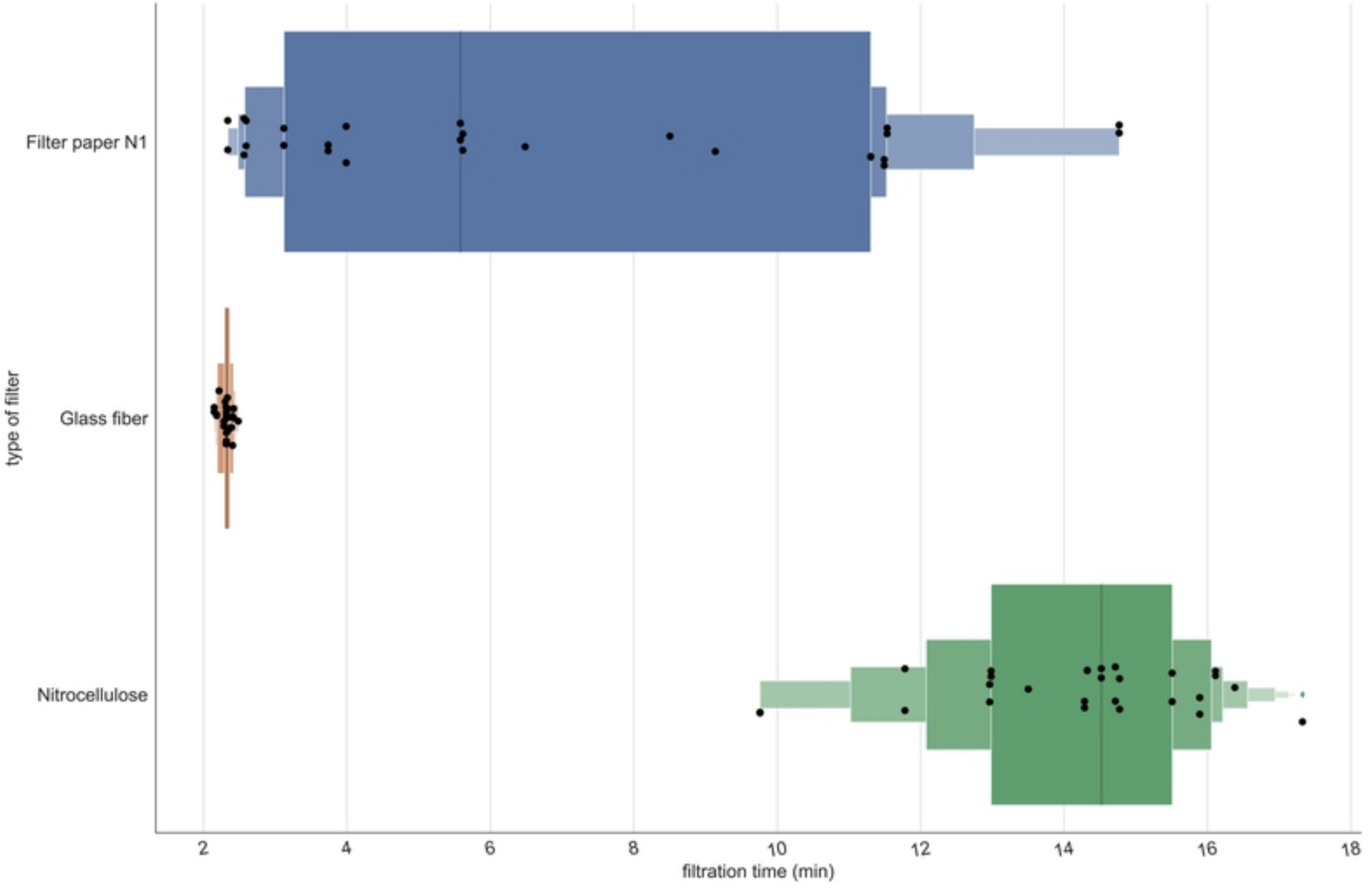

Filtration
Extraction
Inhibitor Removal
Amplification
Glass fiber
$0.44
2m20s
Nitrocellulose
$1.80
14m30s
Whatman Paper n°1
$0.10
5m50s
Magnetic beads
$0.10
25min
Qlagen DNeasy
$2.80
77min
NaOH
$0.05
50min
Paper Dipstick
$0.10
9min
Direct to PCR Dipstick
$0.10
5min
OneStep PCR Inhibitor Removal
$2.16
106min
None
$0
0min
TaqMan Env. Mastermix 2.0
$2.40
180min

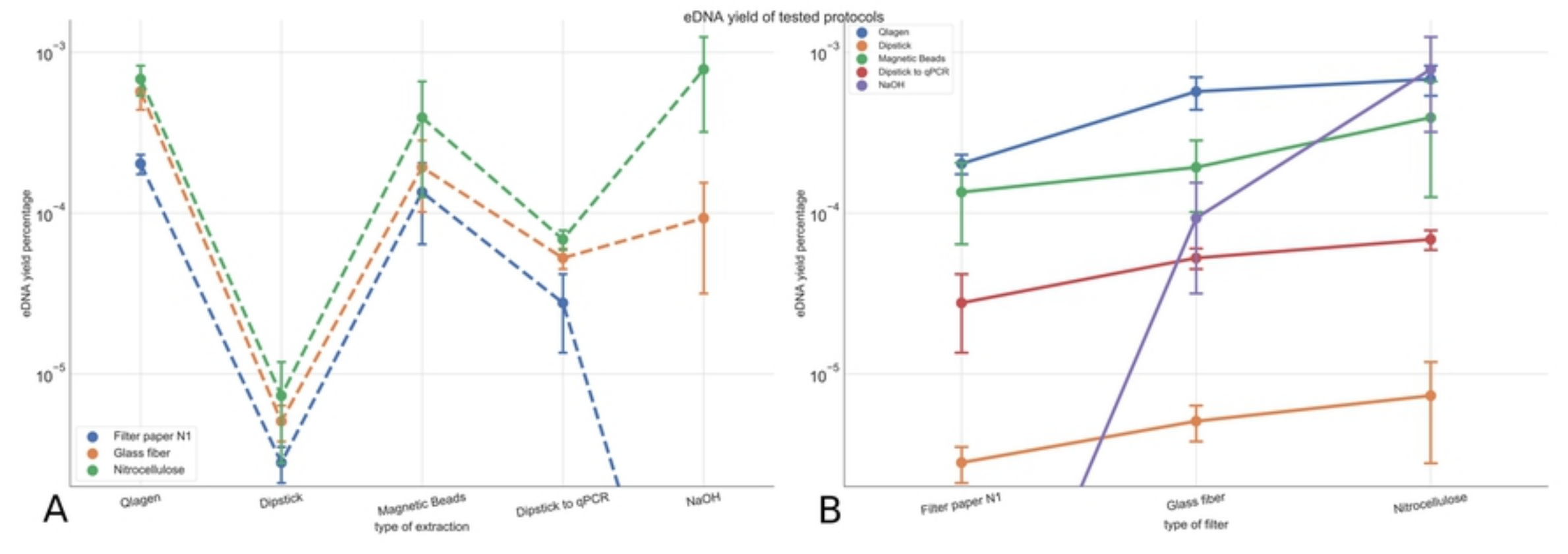

Does the interest species occupy a high percentage of the sampled sites?
Yes
No
Is time a major constraint for the project?
Is time a major constraint for the project?
Yes
No
Is cost a major constraint for the project?
Is cost a major constraint for the project?
Yes
No
Yes
Yes
No
No
Filter: Glass fiber Extraction: qPCR Dipstick
Filter: Glass fiber Extraction: NaOH
Filter: Glass fiber Extraction: Magnetic Beads
Filter: Nitrocellulose Extraction: Qiagen

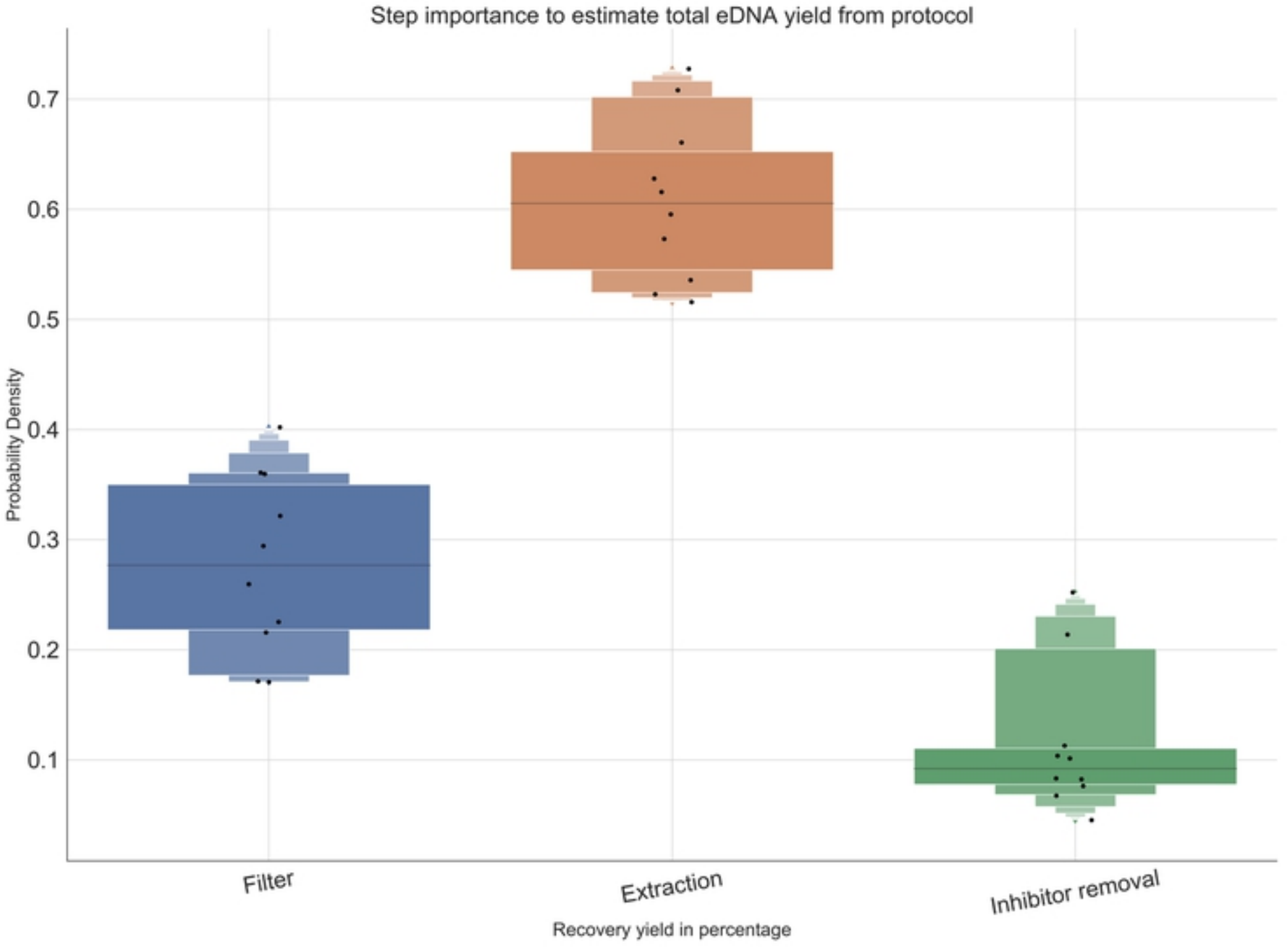


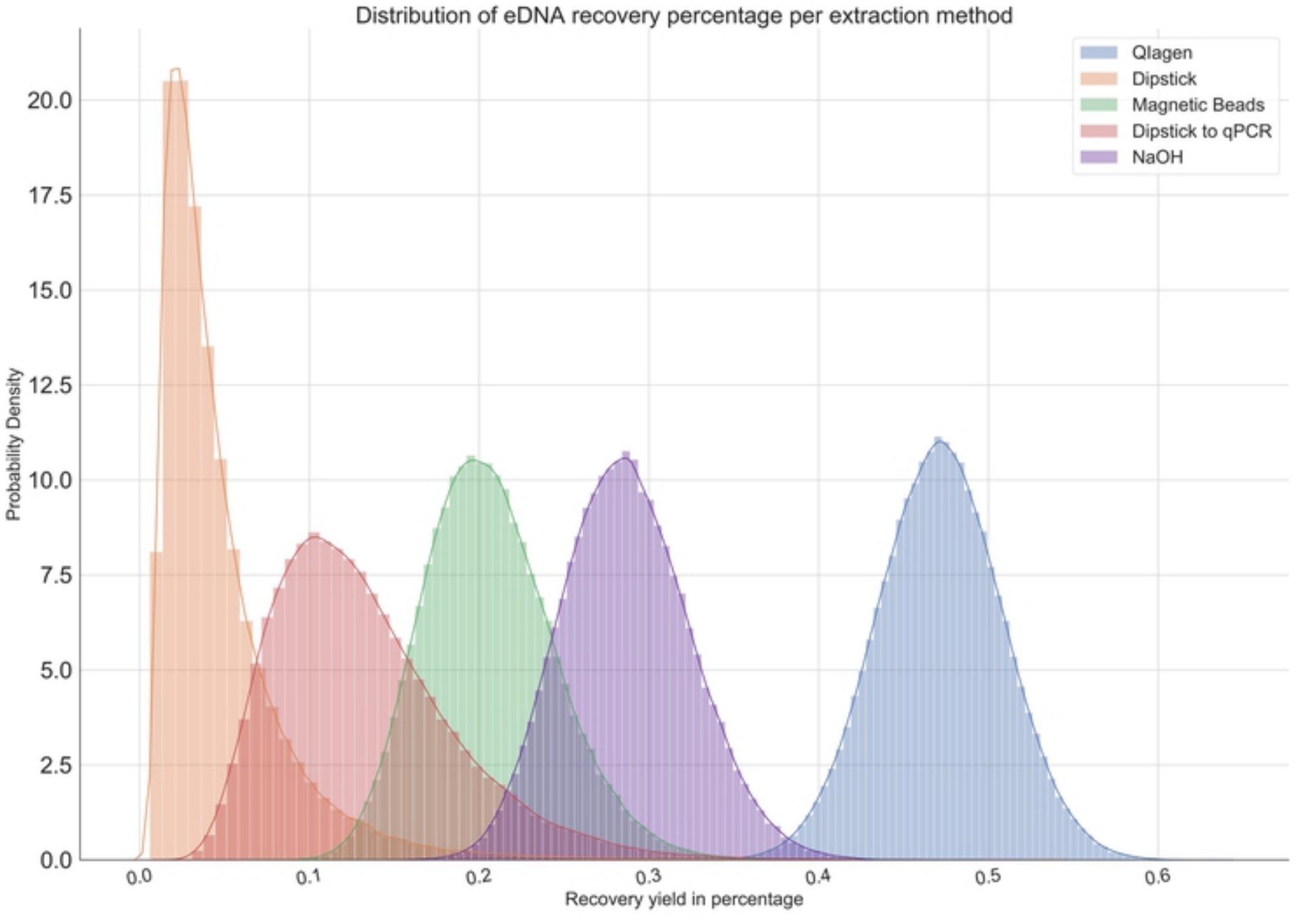

1.0
0.8
0.6
0.4
0.2
0.0
Porbability of amplification
−5
−4
−3
−2
−1
0
1
initial log10[eDNA] (ng/L)

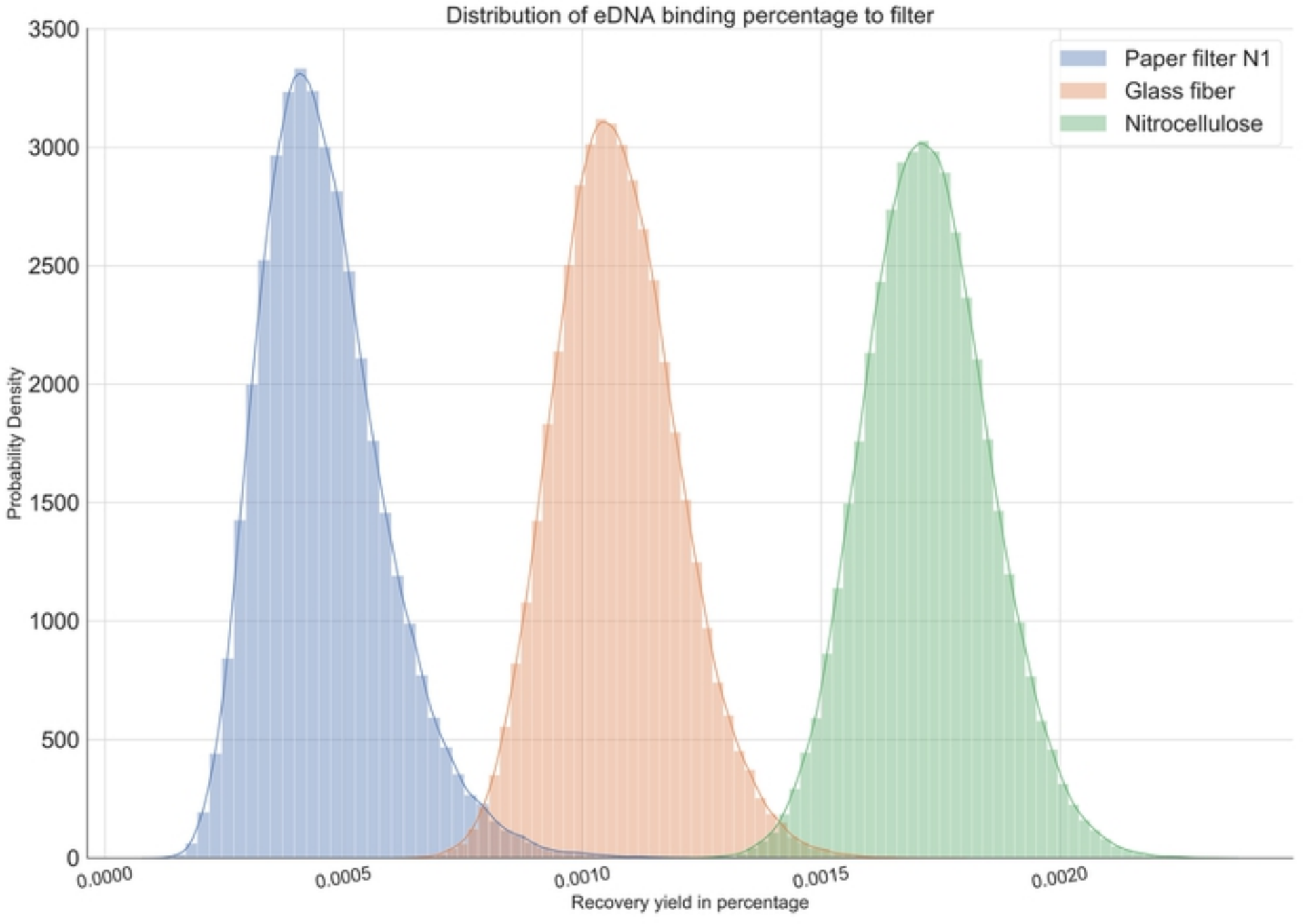

yield percentage
Cost per sample ($)
Time (min)
Nitrocellulose Qlagen Zymo
Nitrocellulose NaOH Zymo
Nitrocellulose Qlagen
Glass fiber Qlagen Zymo
Glass fiber Qlagen
Nitrocellulose Magnetic Beads Zymo
Filter paper N1 Qlagen Zymo
Glass fiber Magnetic Beads Zymo
Filter paper N1 Qlagen
Glass fiber Magnetic Beads
Filter paper N1 Magnetic Beads Zymo
Filter paper N1 Magnetic Beads
Nitrocellulose Magnetic Beads
Glass fiber NaOH Zymo
Nitrocellulose Dipstick to qPCR
Glass fiber Dipstick to qPCR
Filter paper N1 Dipstick to qPCR
Nitrocellulose Dipstick
Glass fiber Dipstick Zymo
Nitrocellulose Dipstick Zymo
Glass fiber Dipstick
Filter paper N1 Dipstick Zymo
Filter paper N1 Dipstick
−5
−4
−3
4
6
8
10
500
1000


Filter paper N1
Glass fiber
Nitrocellulose
type of filter
2
4
6
8
10
12
14
16
18
filtration time (min)


eDNA yield of tested protocols
eDNA yield percentage
$10^{-3}$
$10^{-4}$
$10^{-5}$
Filter paper N1
Glass fiber
Nitrocellulose
Qiagen
Dipstick
Magnetic Beads
Dipstick to qPCR
NaOH
type of extraction
A
Qiagen
Dipstick
Magnetic Beads
Dipstick to qPCR
NaOH
Filter paper N1
Glass fiber
Nitrocellulose
type of filter
B


Step importance to estimate total eDNA yield from protocol
Probability Density
0.7
0.6
0.5
0.4
0.3
0.2
0.1
Filter
Extraction
Inhibitor removal
Recovery yield in percentage


Distribution of eDNA recovery percentage per extraction method
Qlagen
Dipstick
Magnetic Beads
Dipstick to qPCR
NaOH
Probability Density
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
0.0
0.1
0.2
0.3
0.4
0.5
0.6
Recovery yield in percentage


Distribution of eDNA binding percentage to filter
Paper filter N1
Glass fiber
Nitrocellulose
Probability Density
3500
3000
2500
2000
1500
1000
500
0
0.0000
0.0005
0.0010
0.0015
0.0020
Recovery yield in percentage