

Dear Dr. Carvalho,

Please find a revised version of “Environmental DNA provides quantitative estimates of Pacific hake abundance and distribution in the open ocean.” (RSPB-2021-2613) accompanying this letter. We would like to thank you and two reviewers for their thoughtful and substantive reviews to the manuscript. Most of the reviewer comments were focused on providing clarity and improving the readability of the manuscript and we have adopted many of the reviewer suggestions. Specifically, we have highlighted the link between our work and its broader fisheries implications and added relevant methodological details for the qPCR analysis to the manuscript. We provide detailed responses to individual comments below (author responses in red) and have attached a pdf showing the changes between the current and previous version of the manuscript. We believe the manuscript is much improved following revision and that it is now appropriate for publication in Proceeding of the Royal Society B.

Sincerely,

A.O. Shelton for the co-authors

Reviewer 1:**Overview:**

The manuscript details a large-scale qPCR-based eDNA survey of hake along the U.S. west coast. The authors collected eDNA samples across the region (186 stations, 6 depths) and simultaneously gathered hake biomass data using acoustic and trawl methods. Using collected data, they separately modeled eDNA and acoustic-trawl distributions in this large area, allowing for quantitative method comparisons at management-level scales. This study represents an impressive and novel fish eDNA dataset in the ocean and draws some important conclusions regarding the use of non-invasive qPCR methods to survey fisheries-important taxa, like hake.

This manuscript should be considered for publication in Proceedings B and I think it will be of interest to a wide audience. While I do have a lot of suggestions to improve the manuscript, most of these are minor and I believe all of them can be addressed in a timely manner in a revised draft.

General comments:

I would consider changing the title of the manuscript to better reflect the specific target organism and scope of the study. For instance, something like "Environmental DNA provides quantitative estimates of Pacific hake abundance and distribution in the coastal ocean". You may or may not want to include the scientific name as well.

Response: We have modified the title of the manuscript to be more specific.

There are some missing details that may strengthen the abstract. It would be nice to highlight some quantitative results. For example, is there some number or percent overlap that you can report based on the comparison of qPCR and acoustic-trawl data? What was the level of variation between methods? You could also report a correlation value (Pearson). There is also no reference of the models or eDNA index in the abstract. An important finding you touch on but may expand is that on a sample-to-sample basis, eDNA and other methods (acoustic-trawl) may not correlate due to inherent differences in the method (and in your case the differences in sampling time of day) but they do correlate nicely on large scales that are relevant for fisheries management.

Response: We think all of the points the reviewer brings up are entirely reasonable and worthy of inclusion in greater detail in the abstract. However, the abstract uses all of the 200 word limit allowed by the journal. We have made small editorial changes to the abstract to improve readability but have not added all of the detail suggested by the reviewer.

The introduction reads a bit unorganized and is redundant at times. I think you can tighten this up. I would suggest a paragraph that introduces eDNA, its benefits (e.g., non-invasive, sensitive, cost-effective), and applications to management strategies that require quantitative measurements. Important to note, is that applying eDNA to management of ocean resources is still new. Another paragraph could transition to our knowledge of traditional surveys (trawls, visual, acoustic, etc.) in the ocean vs. eDNA and include inherent differences in fish eDNA in the environment compared to a single discrete fish sample collected in a net tow. Then, you can have a paragraph or two that

zoom in on your study region, taxa, and question. Somewhere in the introduction, you may also distinguish amplicon metabarcoding vs. qPCR-based eDNA approaches and how they differ in the types of information gained (e.g., metabarcoding promotes monitoring though is only qualitative vs. qPCR that is more targeted yet quantitative). This is touched on in the discussion but I would mention it earlier. Lastly, I think the introduction would benefit by zooming in on fish and fisheries earlier, establishing the potential for qPCR to inform large scale management in the ocean.

Response: We appreciate the comments on the introduction. We are not sure what the reviewer reads as redundant. We have adopted specific suggestions from the reviewer in editing the introduction including linking the work to fisheries earlier (lines 58-60), emphasizing the importance of moving away from sample-to-sample comparisons (lines 71-76) and added additional species information to the intro (lines 77-87). We have not elaborated upon qPCR vs. metabarcoding as this distinction is not an emphasis of the research in this paper.

Regarding methods, more detail would be helpful in the main text. A lot of this detail is buried in the Supplemental file. I would pull some of this information into the manuscript and refer to the Supplement for additional details and figures. For instance, more information on your qPCR methods is needed in the main text. I know the emphasis is on the model(s) but it is important to clearly state qPCR eDNA methods, especially for reproducibility (see Langlois et al. 2021 – *Environmental DNA*). The reader may miss these key details. Some qPCR questions that arise: What samples did you use for positive hake controls? Did you use voucher specimens? What primers (and other reagents) did you use? Did you account for false positives or negatives? What about closely-related taxa? Did you measure bulk DNA concentrations in the samples (e.g., via Qubit)? Or did you just quantify hake-specific DNA concentration? I would also pull-out information on DNA contamination and washing error and touch on these in the methods or results section (see specific points below), referencing the Supplement for more details. Again, some readers may not look at the Supplemental file.

Response: While we appreciate all of these points, the eDNA collection, extraction and qPCR methods for this study were previously published in our companion paper in great detail, and include a detailed protocol (Ramon-Laca et al. 2021, citation 25), and are otherwise summarized in the Supplement. We now make clear that all of the topics listed in the reviewers questions are discussed in the companion paper (lines 102-105). We have also added several specific sentences in response to the points above that are particularly relevant to our analyses (e.g. wash errors and contamination) as they are directly linked to the models and analyses (see additional responses below).

I would be consistent with the way you reference “acoustic-trawl” surveys throughout the manuscript. It is important to emphasize that your biomass model is integrating both acoustic and trawl data, and so, referring to this as “acoustic-trawl” seems accurate.

Response: We now use acoustic-trawl throughout the manuscript.

More specific comments:

Line 23: Environmental DNA does not necessarily reflect species present in a given habitat. Something like, “...and collecting this environmental DNA can provide

information on species that may be present in a given habitat” may be more appropriate.

Response: While we appreciate the point of the reviewer, we think they are conflating the presence of DNA from a species in a particular location and the process of detecting that species in a particular sample or sampling regime. If a species is truly present in a habitat, its DNA will also be there. Said DNA may or may not be detected by a particular sampling regime and there are many complexities about sampling that are worthy of research and discussion. In the first line of the abstract, we are not trying to express the full complexity of the problems of detection and sampling. We only wish to note that species shed DNA and their DNA is in the environment. We disagree that adding conditional verbs indicating sampling uncertainty aids the readers understanding of our work. We have not changed this line.

Line 29: Spatially smooth signature? Sounds cryptic. Do you mean a wide spatial coverage? I would clarify here and in other parts where this is mentioned.

Response: We are uncertain what the reviewer means by “wide spatial coverage” here. However, we appreciate that the term spatially smooth may be unfamiliar to readers who are not accustomed to thinking about spatial analyses. As a result, we have removed this language and replaced it with more specific in line 30.

Line 71: You could also include information about distribution of eDNA from its source location due to movement by currents or tides. Though not a concern for your study, it is important to note that fish eDNA can also be patchy in coastal areas (see Kumar et al. 2021 – *Environmental DNA*). When you integrate over wide areas this becomes less of an issue. I would make this clear.

Response: We agree that the movement of water is an important factor in the ecology of eDNA. The citations within this paragraph (#5-10) each discuss eDNA and transportation extensively (lines 66-73). We also mention DNA transport in passing elsewhere in the introduction (line 45).

Line 75: I would be careful saying this is the most spatially extensive eDNA survey of the oceans to date. It may be more accurate to say “fish eDNA” survey. Though not sampling eDNA per se, the Tara Oceans expedition performed DNA barcoding of microbes-protists on a global scale. I am not diminishing your study; I would just clarify. You may also say “qPCR-based eDNA” survey. Both would be accurate.

Response: We have modified this line and refer to the survey as simply very large (line 77).

Line 76: I would remove the statement “an area of ocean approximately equivalent to Portugal’s land area”. This seems extraneous.

Response: We wrote this to provide people who are not familiar with this part of the world a rough mental reference for scale, but we have removed it in the new version (see line 77).

Line 81: You mention there are rich datasets of hake biomass, presumably from trawls. Have there been many studies targeting hake with qPCR-based eDNA methods? Have others conducted trawl vs. eDNA of important fisheries taxa? What have they found?

Response: There have been no comparable large-scale studies of hake or any other fisheries species using qPCR to our knowledge. We provide reference to the fisheries relevant literature (citations 10-12) but these are nearshore or riverine applications. We have added a sentence making the uniqueness of our application to fisheries more explicit (see lines 58-60)

Line 87: Has this type of model been used in terrestrial applications? Any papers you could cite?

Response: Spatial models of various flavors and structures have an enormous literature with many types of applications in all sorts of habitats - terrestrial, oceanic, riverine, etc. We now provide a few citations to some of the more general introductions to spatial models in ecological and management settings (see line 91-3)

Lines 96-99: I would move this material to the introduction.

Response: We have moved this information to the introduction. (see lines 77-87)

Line 101: You reference a related paper with more details. Still, I think it would be nice to point out lat-long range here or geographic sites that cover the study area.

Response: We have added relevant geographical coordinates to this section of the methods (line 102)

Line 114: Here, you could briefly mention that you found some DNA in the negative controls but it was mostly detected below xxx copies L . Then reference the Supplemental file for more information.

Response: We have added this information to the methods (see lines 122).

Lines 116-117: Is this the filtering protocol used for controls and discrete seawater samples? I would specify. How long were the filters stored before extraction?

Response: We have clarified that we filtered both field samples and controls using the same methodology (line 124-129). Filter storage time before extraction ranged from 6 to 14 months. We have added this information to the supplement (page 1 of the ESM).

Lines 119-120: Citation associated with the extraction method and the increase in yield?

Response: We have added the relevant citation (line 125)

Lines 120-123: Expand on qPCR details. Pull out from Supplemental file and reference additional details and figures.

Response: See the response to the general comments above.

Line 126: Confused by this sentence. What is the “process”?

Response: The notion of separating a model for a biological process and a model for the observations of that process are the foundation of the world of state-space models. In our case the process is the true, but unobserved distribution of hake DNA in space

and with depth. The observation model is how we deal with sample stochasticity, inhibition, wash errors, etc. We have added a sentence to elaborate on what we mean by a process model and provide several references to the general literature of state-space models (lines 133-137)

Line 134: I would refer to specific figures in the Supplemental file that are associated with model optimization mentioned in text.

Response: We don't show alternate model formulations in the supplement, we only describe our exploration in words (see page 8 of the ESM).

Line 140: I would briefly expand on the wash error and refer to the Supplemental material for relevant figures. Otherwise, the reader will have no idea why you included this in your model.

Response: We have added language about wash error earlier in the methods (see lines 129-131).

Line 147: Could say "indicates" instead of "indexes".

Line 153: Could say "simultaneously" instead of "contemporaneously".

Lines 159-161: This is a key aspect of the study. The reader should be aware of this somewhere in the introduction. Not necessarily the day vs. night differences between methods that preclude direct site level comparisons but rather that fact that you are focusing on broad-scale and not sample-to-sample comparisons.

Response: We agree that this is an important part of our study and that is why we devote a paragraph in the introduction describing how eDNA observation methods are likely to differ from other traditional sampling approaches (see lines 61-76). Like the reviewer, we think that day-night comparisons is an example of a much broader class of comparisons that we bring up in the introduction. As suggested by the reviewer, we have added language about sample-to-sample comparisons in the introduction to clarify our general point about comparisons among sampling methods. (see line 73-4)

Lines 216-217: Remove "genetic signatures of hake" and replace with "hake eDNA". When you mention detected hake eDNA, are you implying this is above your qPCR detection limit (e.g., 20 copies L⁻¹). I would mention this detection limit somewhere in the qPCR methods section. May be nice to report the range (and mean) of hake DNA concentrations you observed in the study.

Response: We have clarified our language in this section and specified that any signal of DNA is included in our detection of hake. We have added a reference to our detection thresholds as well as the range of hake concentrations estimated by our models (see lines 235)

Line 218: Spatial smoothing sounds odd. Maybe, spatial coverage?

Response: We have removed spatial smoothing and replaced it with language about the relative rates of detection for the two survey types (lines 230-232)

Lines 220-222: Give a range of values and average.

Response: We have added values describing the observed variation in hake DNA among individual water samples as well as among stations by depth (sampled with multiple individual water samples; lines 234-240).

Line 223: Results are vague. What is generally low? <50 copies L⁻¹? What is meant by higher and more heterogeneous? Give us some values (average and range).

Response: See previous two comments.

Lines 226-228: Confusing and seems like a discussion topic. Why do you think there was more variability between replicates at the surface? Just a more environmentally heterogeneous sample compared to deeper waters. Hake would be more abundant at the surface when you sampled eDNA at night, so this may reflect feeding behaviors on patchy prey. I wonder if including environmental variables (via CTD) would support your data or provide more context. Did you explore this?

Response: We agree that this is confusing. We think it is especially worth mentioning because we do not have a good explanation for why this pattern would occur. To improve readability, we have broken our writing into two short paragraphs, the first on mean DNA concentrations (233-239) and the second on variability in DNA concentrations (line 240-245). We have no data on hake abundance or distribution at night but there is literature that hake do vertically migrate. It certainly does motivate further examination and we hope to look into it in future work.

Line 229: Still not liking the sound of “spatially smooth”. Lines 231-232. Move to introduction.

Response: We have removed lines 231-232 and added comparable language to the introduction (lines 91-92)

Line 233: What is meant by strong spatial patterning? Line 235: Move some of this to discussion.

Response: Strong spatial patterning is simply substantial variation in hake biomass in the acoustic-trawl survey in space. We have left this sentence in the results as it is describing what we found.

Line 237: You present Pearson correlation data between the two indices but do not explain this in the methods. You should also report significance of the correlations (p-values).

Response: We have added a sentence describing the correlations in the methods (lines 225-227). We provide uncertainty intervals on the bounds of the correlation coefficient which clearly support the estimates being very far from and not including zero. As we work within the Bayesian paradigm for the entire paper, we generally disagree with the use of p-values in any context. We believe our points are made clearly without the use of a p-value.

Lines 247-248: This statement seems redundant after you provide strong Pearson correlation data.

Response: We have cut this sentence from the manuscript.

Lines 249-251: Redundant with information in methods.

Response: We agree that this information is mentioned in two places separately in the methods, but we think it is relevant to draw attention to the different efforts expended between the acoustic-trawl and the eDNA surveys (lines 259-262)

Lines 256-257: Move to discussion. You may consider combining the last two paragraphs of the results.

Response: We think the last two paragraphs are about different measures of spatial and depth distribution. We have retained the current structure.

Lines 259-263: I may be a bit more direct here in terms of what you did and the major findings. For instance, “ Our study assessed the efficacy of eDNA sampling to quantitatively estimate hake concentration and distribution along the US west coast, a well-established fisheries habitat for this species. We sampled eDNA at 186 stations, with depth and latitude, and used a model to estimate and compare with traditional acoustic-trawl data. We found these methods were comparable across wide geographic scales (Pearson of xx), despite less eDNA samples being collected compared to acoustic transects.”

Response: We appreciate the suggestion and have made modest changes to this paragraph (starting line 278).

Lines 271-272: Okay, so this type of model has been used elsewhere?

Response: Yes. Spatial models are used in many ecological contexts and applications including fisheries.

Lines 275-279: You hint at this but I would emphasize that with eDNA you are non-invasively collecting seawater and applying more sensitive methods to target hake. This can be applied to other fisheries-important species within the same samples and supplement existing fisheries operations and stock assessment data.

Response: We appreciate this suggestion and have presented this information in the introduction (lines 53-55) and at several points in the discussion (line 329-32).

Line 291: Do you report environmental metadata collected with the water samples (CTD data)? This may be nice to include, if not for the standard practice of reporting metadata. Including extra variables in the models is likely beyond the scope of this study, which was aimed at comparing sampling methods, but it may be useful to interpret spatial trends in hake eDNA.

Response: There is an enormous amount of data associated with this survey. All of that information is contained and made available elsewhere [citation 28]. This topic is beyond the scope of the manuscript.

Line 292: Typical to reference figures in the discussion for this journal?

Line 307: I agree it is worth mentioning barcoding here. I would expand a bit.

Incorporating both approaches would be useful, allowing for biomonitoring of the food web (barcoding), as well as more targeted approaches that inform fisheries (qPCR).

Both methods can be exploited from the same collected sample and DNA samples can be archived for long periods of time, preserved or frozen. There is thus a lot of potential and upside to collecting water samples in conjunction with traditional surveys and would not add too much more effort (e.g., filtering water and storing it).

Response: We agree and mention these positive aspects of eDNA sampling in both the introduction and discussion (lines 296, 325-329).

Lines 318-321: Add more detail here about what this information can provide to fisheries management and stakeholders. Expand or wrap this into a Conclusions section. What is the next step? Do you need to complete more eDNA surveys over different years/seasons and compare to trawl data? Are there other types of surveys we could compare with in this region, like camera-based monitoring? How do we translate a strong correlation between eDNA index and fish biomass to an actual fisheries-reliable metric (akin to landings or stock assessments)? Can we use this to model future populations? Add more variables to the model? Mentioning next steps would be helpful.

Response: Some of these questions are already answered in the discussion – e.g. the important role of time-series for making an index of abundance fisheries management is discussed in lines 316-318. One of the major points of the paper is that eDNA is not fundamentally all that different from existing survey methodologies. Logically eDNA can do similar things to existing methods. We have added a sentence to the final paragraph about future links to oceanography and estimates of distribution (paragraph starting line 341).

Lines 322-324: All files are on the GitHub page! I would think about adding more to the description (readme) to indicate what types of data files and R scripts are in the zip file. It will make it easier for those trying to access the data in the future.

Response: We have added information to the readme and archived this data to a permanent Dryad repository associated with this manuscript.

Figure 1: Overall, the figures look good. The bars in panel G are a bit hard to interpret.

Response: We have slightly modified the caption to figure 1

Figure 2: Panel letters are on the inside and smaller compared to Figure 1. I would try to be consistent. I would reference that latitudinal bins are denoted by dotted lines.

Response: We have added a reference about latitudinal bins to the figure caption.

Figure 4: Panel C is hard to interpret. The colors are faint in the graph and everything is blended. You may try faceting panel C based on bottom depth. This figure or parts of it could also go into the Supplemental section. I think Figures 1-3 are most central to the story.

Response: We think Figure 4 is important as it reports several aspects of distribution that are commonly used in the species distribution literature. We have retained it but adjusted the figure size to enhance readability and make it more comparable to the other figures in the manuscript.

Supplemental Information:

Standards: Include information on qPCR standards and controls in the main text.

Response: We have mentioned the range of information included in the accompanying supplement.

Investigation of contamination: Some of this information (sentence 2) is redundant with main text. I would briefly mention this in the methods/results section and refer to the Supplemental figures or additional details. Any speculation as to the sources of the contamination? I would move some of your explanation regarding the contamination to the main text. This is important to report.

Response: We have move reference to contamination to the main text (see responses above). We are uncertain about the source of contamination.

Inhibition: Why so much inhibition in surface samples?

Response: We suspect that the inhibition is related to phytoplankton collected in the samples, but we are uncertain the ultimate source of inhibition.

Ethanol wash error: Some of this should be in main text to provide context for the model.

Response: We have made these additions to the main text (see responses above).

eDNA model – Measures of fit and uncertainty: Regarding the 3m samples. Was the ships intake line properly sanitized or flushed? Was there a visual difference in the flow rate that may have caused DNA to shear?

Response: The intake line for surface waters is not sanitized. The entire system is continuously flushed with surface sea waters when in operation, and the local access point is flushed for 5-10 minutes prior to sample collection.

Referee: 2

Comments to the Author(s)

The paper is extremely relevant for conservation. It stems from the fact that eDNA has become an efficient method to assess species diversity and changes in community with the potential to greatly improve our understanding of natural communities while it remains unclear whether eDNA signals can provide quantitative metrics of abundance to support management. The study is based on the results of a large ocean survey (spanning 86,000 km² to depths of 500m) and is focused on the abundance and distribution of Pacific hake (*Merluccius productus*) along the west coast of the United States. The knowledge available for hake provides an opportunity to rigorously compare available information from traditional surveys with eDNA assessment. The paper is well written and suitable for the journal and it could be accepted on its present form. My only questions and suggested revisions are the following:

- among the most significant results there is the assessment of hake DNA variability in the study area which varied substantially with depth, with the highest concentrations between 100m and 300m depth (which I believe is consistent with the species preferred habitat) and concentrations lower and more homogeneous at depth than near the surface. I was wondering whether the fact that the most of water collection for eDNA occurred at night may have also an influence in this respect. Perhaps the authors may want to add this in their discussion;

Response: The reviewer presents an interesting point. Unfortunately we have no information to evaluate the effect of night vs. day sampling on eDNA. We have added a short note (line 339) suggesting that diel migration patterns may play a role in the observations.

- about the e DNA index that was created for the purpose of the spatial analysis, the authors explain that they have generated a depth-integrated index of hake DNA summing the values across all depths and not integrating values across the entire water column or multiplying by the total water volume within each grid cell so that the absolute value of the index depends upon the number of discrete depths at each location. I have two questions in this respect (same questions that I would expect also the readers may have): why the authors decided to use this index instead of the posterior predictions at each depth provided at 200, 250, 350, 400, and 450m for each 5km grid cell, and secondly, given that some locations spatial locations had depths lower than 500m, why they did not standardise the index to a depth common to all the locations?

Response: The reviewer raises several interesting points. As far as this question : "why the authors decided to use this index instead of the posterior predictions at each depth provided 200, 250, 350, 400, and 450m" We did this because we have no observations at those depths, and therefore no estimated model or posterior predictions for those depth. We think of the model as providing a predicted spatial surface at the depths we have observations (3m, 50m, 100m, 150m, 300m, 500m). An independent spatial (latitude-longitude) smooth is estimated for each depth which allows for different

depths to have different spatial patterns and the flexibility to follow the observations at each depth. This can be observed in the spatial maps presented in Fig. 1. However, this statistical form does mean that there is no direct connection between the predictions at a particular spatial location outside of the covariates that are shared due to location (in this model, the bottom depth covariate effect is shared among water depths). As a result, there are no posterior predictions at other depths to sum. We have added a sentence to the section about creating an eDNA index to emphasize that we do not have predictions to other, unobserved depths (line 212).

As far as why we do not sum to deeper depths, we have no information from eDNA below 500m even if the bottom depth is more than 500m. Additionally, the acoustic survey only includes information between 50m and 500m. So we developed an index for between 50m and 500m to make a legitimate comparison between the two methods.

Finally, we do not standardize the index to a common depth, because the intention of the survey is to provide an estimate of total abundance not an estimate at a particular depth. Thus we are interested in an index of the sum across depths, not an index associated with a particular depth.

¹ Environmental DNA provides quantitative estimates of Pacific hake abundance
² and distribution in the open ocean.

³ Andrew Olaf Shelton^{1*}, Ana Ramón-Laca², Abigail Wells³, Julia Clemons⁴, Dezhong Chu⁴, Blake E. Feist¹,
⁴ Ryan P. Kelly⁵, Sandra L. Parker-Stetter^{4,6}, Rebecca Thomas⁴, Krista M. Nichols¹, Linda Park¹

⁵ ¹Conservation Biology Division, Northwest Fisheries Science Center, National Marine Fisheries Service,
⁶ National Oceanic and Atmospheric Administration, 2725 Montlake Blvd. E, Seattle, WA 98112, U.S.A.

⁷ ²Cooperative Institute for Climate, Ocean, and Ecosystem Studies, University of Washington at Northwest
⁸ Fisheries Science Center, National Marine Fisheries Service, Seattle, WA 98112, U.S.A.

⁹ ³Lynker Technologies, Under contract to Northwest Fisheries Science Center, National Marine Fisheries
¹⁰ Service, National Oceanic and Atmospheric Administration, 2725 Montlake Blvd. E, Seattle, WA 98112,
¹¹ U.S.A.

¹² ⁴Fisheries Resource Analysis and Monitoring Division, Northwest Fisheries Science Center, National Marine
¹³ Fisheries Service, National Oceanic and Atmospheric Administration, 2725 Montlake Blvd. E, Seattle, WA
¹⁴ 98112, U.S.A.

¹⁵ ⁵University of Washington, School of Marine and Environmental Affairs, 3707 Brooklyn Ave NE, Seattle, WA
¹⁶ 98105, U.S.A.

¹⁷ ⁶Resource Assessment & Conservation Engineering Division, Alaska Fisheries Science Center, National Marine
¹⁸ Fisheries Service, National Oceanic and Atmospheric Administration, 7600 Sand Point Way NE, Seattle, WA
¹⁹ 98115, WA

²⁰ *corresponding author: ole.shelton@noaa.gov

²¹ *Keywords:* environmental DNA, species distributions, fisheries, ocean surveys

²² **Abstract**

²³ All species inevitably leave genetic traces in their environments, and the resulting environmental DNA
²⁴ (eDNA) reflects the species present in a given habitat. It remains unclear whether eDNA signals can provide
²⁵ quantitative metrics of abundance on which human livelihoods or conservation successes depend. Here, we
²⁶ report the results of a large eDNA ocean survey (spanning 86,000 km² to depths of 500m) to understand
²⁷ the abundance and distribution of Pacific hake (*Merluccius productus*), the target of the largest finfish
²⁸ fishery along the west coast of the United States. We sampled eDNA in parallel with ~~traditional survey~~
²⁹ ~~methods and show how eDNA provides a spatially smooth signature of hake relative to traditional acoustic~~
³⁰ ~~surveys~~a traditional acoustic-trawl survey to assess the value of eDNA surveys at a scale relevant to fisheries
³¹ management. Despite local differences, ~~at management-relevant scales~~ the two methods yield comparable
³² information about the broad-scale spatial distribution and abundance ~~of hake~~. Furthermore, we find depth and
³³ spatial patterns of eDNA closely correspond to ~~acoustic-acoustic-trawl~~ estimates for hake. We demonstrate
³⁴ the power and efficacy of eDNA sampling for ~~estimation of estimating~~ abundance and distribution and move
³⁵ the analysis eDNA data beyond sample-to-sample comparisons to management relevant scales. We posit that
³⁶ eDNA methods are capable of providing general quantitative applications that will prove especially valuable
³⁷ in data- or resource-limited contexts.

38 **Introduction**

39 Environmental DNA, the DNA from target organisms collected from an environmental medium (e.g. soil
40 or water), can reflect species in a wide range of terrestrial, aquatic, and marine habitats [1]. eDNA has
41 the potential to revolutionize our understanding of natural communities by enabling rapid and accurate
42 surveys of many species simultaneously ~~without capturing individuals~~ [1]. At present, eDNA can efficiently
43 survey species diversity and changes in community membership [2–4]. However, many natural resource
44 questions depend upon quantitative estimates of abundance (e.g., fisheries or managing species of conservation
45 concern) ~~and for these topics,~~ ~~so~~ eDNA must provide such information in order to be most useful [5]. While
46 most studies find a positive relationship between eDNA concentrations and other survey methods [reviewed
47 by 6], uncertainty about the strength of the eDNA-abundance relationship due to the complexity of eDNA
48 generation, transport, degradation, and detection have limited the application of eDNA in many quantitative
49 applications [5, 7, 8]. While the use of eDNA methods has grown exponentially from tens of publications
50 in 2010 to ~~over 600~~ many hundreds in 2020 [6, 9], reflecting widespread adoption of eDNA technologies,
51 basic questions about the characteristics of eDNA limit its practical application and slow its adoption in
52 environmental management.

53 Rigorous, well designed surveys underlie the successful management and conservation of wild populations.
54 But field surveys are expensive – ~~for example,~~ open-ocean surveys involve ship time costing tens of thousands
55 of dollars per day – and are typically tailored to one or a few species. eDNA methods are appealing for
56 open-ocean or other difficult-to-sample locations because sampling can be fast, standardized, non-lethal, and
57 detect many species simultaneously; sampling involves only the collection and processing of environmental
58 samples [1]. Even modest improvements in sampling efficiency from current surveys can reduce the duration
59 of surveys, yield substantial cost savings for focal species surveys, and free survey time to be reallocated
60 to other understudied communities. However, such broad-scale implementation depends upon providing
61 eDNA-based estimates of abundance at management-relevant scales [10–12]. To date, there have been no
62 eDNA surveys conducted at sufficiently large scale to inform ocean fishery management, a field with many
63 potential eDNA applications.

64 Observations of eDNA differ from observations derived from traditional methods (e.g. visual [13–14], capture
65 ~~and~~ [15–16], or acoustic [12] surveys) and the degree of agreement between individual samples of eDNA and
66 traditional methods collected simultaneously often determines whether eDNA-based methods are viewed as
67 successful or not [12,16]. However, eDNA observations arise from fundamentally different processes than
68 observations from these traditional survey methods – most dramatically, by exponential amplification of
69 DNA molecules in an environmental sample [18, 19], but also because the distribution of eDNA itself in the

70 environment is not identical to the distribution of its source organisms [5,7,8]. In the case of microbial eDNA,
71 this distributional distinction is negligible, but for larger animals – such as fishes or marine mammals – it is not.
72 Conceptually, fish are discrete, while the DNA traces they leave in the water are relatively continuous, blurring
73 their environmental fingerprint over space and time [10]. For example, acoustic surveys of pelagic fishes
74 reflect the patchy ~~and skewed~~ distribution of schooling fishes [20]. By comparison, we expect the associated
75 eDNA to be distributed more evenly as a result of fish movement~~and~~^{and} the lag between shedding and decay
76 processes, and water movement [7,8,11]. Thus, simple sample-level comparisons between eDNA and other
77 survey methodologies are a poor method for determining the usefulness of eDNA surveys. Understanding
78 the ecology of eDNA [7] makes possible an honest assessment of the potential uses and limitations of eDNA
79 for applied environmental problems, and allows each data stream to be used to its best advantage.

80 Here, we leverage ~~the most~~^a spatially extensive eDNA survey of the oceans ~~to date~~ – spanning over 86,000 km²
81 ~~, an area of ocean approximately equivalent to Portugal's land area,~~ and to depths of 500m – to document the
82 empirical patterns of eDNA for a commercially important and abundant fish species, Pacific hake (*Merluccius*
83 *productus*). Hake is a semi-pelagic schooling species and is among the most abundant fish species in the
84 California Current Ecosystem [21, 22], ~~supporting~~. They support a large and important fishery along the
85 Pacific coasts of US and Canada with coastwide catches in excess of 400,000mt annually and ex-vessel value
86 in excess of \$60 million in recent years [22]. Hake are a key component of the California Current ecosystem
87 as both predator and prey, migrating to the surface at night and back to mid-water depths during the day
88 [23]. Seasonally, adults migrate between southern spawning areas and northern foraging areas [21, 23, 24].
89 The rich datasets available for hake provide an opportunity to rigorously compare available information from
90 traditional surveys with eDNA using parallel statistical models that relate observations from each data type
91 to quantitative indices of abundance.

92 We investigate large-scale and depth-specific spatial patterns of hake DNA in the open ocean using a
93 quantitative PCR assay targeting the 12S mitochondrial gene region [23²⁵]. We show how eDNA can be
94 aggregated to provide a depth-integrated index of hake abundance comparable to acoustic-trawl survey results
95 used for fisheries stock assessments [22]. The spatial-statistical model we use is a first for eDNA in the ocean
96 and makes results for eDNA surveys comparable to other methods used in quantitative natural-resources
97 management [26, 27]. We derive metrics of the species' spatial and depth distribution and investigate the
98 relative precision of the eDNA and ~~acoustic~~^a acoustic-trawl surveys. Our results show that eDNA analyses
99 can provide important information about species abundance and distribution at management-relevant scales,
100 provide relatively straightforward opportunities for supplementing existing surveys, and open the door for
101 providing quantitative information for additional species that are currently un- or under-studied.

102 **Materials and Methods**

103 Pacific hake are a schooling, semi-pelagic fish that are a key component of the California Current ecosystem
104 as both predator and prey, migrating to the surface at night and back to mid-water depths during the day 25.
105 They are targeted by major fisheries in both the United States and Canada 22and adults perform seasonal
106 migrations between southern spawning areas and northern foraging areas 21, 24, 25.

107 **Field sampling and processing for eDNA**

108 We collected eDNA samples during the 2019 U.S.-Canada Integrated Ecosystem & Acoustic-Trawl Survey
109 for Pacific hake aboard the NOAA Ship *Bell M. Shimada* from July 2 to August 19 [2628] including waters
110 from 38.3°N to 48.6°N along the Pacific coast of the United States (123°W to 126.5°W longitude). Detailed
111 collection protocols and laboratory analyses for all all laboratory analyses including information on sample
112 preservation and extraction, primers (12S primer description, specificity and sensitivity testing, and other
113 aspects), qPCR protocols, voucher specimens, and all other steps are provided in Ramón-Laca *et al.* [2325].
114 We briefly summarize those protocols here.

115 We collected seawater from up to six depths (3, 50, 100, 150, 300, and 500m) at 186 stations where a
116 ConductivityTemperature, Temperature, and Depth (CTD) rosette was deployed. These stations were
117 spread across 36 acoustic transects (Fig. 2). We included 1,769 individual water samples collected at 892
118 depth-station combinations (a small number of samples were contaminated or lost during processing). 710
119 depth-stations were collected at 50m deep or deeper. Two replicates of 2.5L of seawater were collected at each
120 depth and station from independent Niskin bottles attached to a CTD rosette. Water samples from 3m were
121 collected from the ship's saltwater intake line but processed identically to Niskin samples. Nearly
122 all CTD casts, and therefore water collection, for eDNA occurred at night while acoustic-trawl sampling (see
123 below) took place during daylight hours.

124 To account for possible contamination, negative sampling controls were collected routinely by filtering 2L of
125 distilled water from either the onboard evaporator or from distilled water brought from the laboratory for this
126 purpose (N=49 in total). The water was Both Niskin collected and control samples were filtered immediately
127 using 47mm diameter mixed cellulose ester sterile filters with a 1μm pore size using a vacuum pump. The
128 filters were stored at room temperature in Longmire's buffer until DNA extraction [2729]. We detected low
129 levels of hake contamination in control samples with most negative controls having average estimated DNA
130 concentrations of 44 copiesL⁻¹ which is slightly larger than the detection threshold of 20 copiesL⁻¹ (see
131 ESM: Figs. S3, S4), but below most estimated hake DNA concentrations from field samples.

132 The DNA was extracted using a modified phenol:chloroform method with a phase lock to increase the
133 throughput and yield [25]. Quantification of Pacific hake was performed by qPCR using a specific TaqMan
134 assay on a QuanStudio 6 (Applied Biosystems) that included an internal positive control (IPC) of the
135 reaction to account for PCR inhibition. Any delay of more than 0.5 cycles from the IPC at the non-template
136 controls of the PCR was considered inhibition. For inhibited samples, we used a 1:5 dilution in subsequent
137 analyses. A subset of samples had a final wash with an incorrect concentration ethanol (30% ethanol instead
138 of 70% ethanol). These samples had reduced hake DNA concentrations and we accounted for samples with
139 the improper wash in our statistical model below (see ESM for more details).

140 Spatial eDNA Model

141 We developed a Bayesian state-space framework for modeling DNA concentration in the coastal ocean. State-
142 space models separate the true biological process from the methods used to observe the ~~process~~biological
143 process [see 30, 31]. In our case, the biological process of interest is the spatial- and depth-specific pattern of
144 hake eDNA. We use a relatively simple process model. Let D_{xyd} be the true, but unobserved, concentration
145 of hake DNA ($\text{DNA copies } L^{-1}$) present at spatial coordinates $\{x, y\}$ (northing and eastings, respectively,
146 in km) and sample depth d (meters). We model the DNA concentration as a spatially smooth process at
147 each depth sampled ($d = 3, 50, 100, 150, 300, \text{ or } 500\text{m}$) and linear on the \log_{10} scale,

$$\log_{10} D_{xyd} = \gamma_d + s(b) + t_d(x, y) \quad (1)$$

148 where γ_d is the spatial intercept for each depth, $s(b)$ indicates a smoothing spline as a function of bottom
149 depth in meters (b), and $t_d(x, y)$ is a tensor-product smooth that provides an independent spatial smooth for
150 each depth. We use cubic regression splines for both univariate and tensor-product smoothes. We investigated
151 a range of knot densities for smoothes in preliminary investigations (see *ESM*).

152 From the process model in eq. 1, we construct a multi-level observation model. First, we model the DNA
153 concentration in each Niskin bottle i , as a random deviation from the true DNA concentration at that depth
154 and location and include three offsets to account for variation in the processing of eDNA extracted from
155 Niskin bottles.

$$\log_{10} E_i = \log_{10} D_{xyd} + \delta_i + \log_{10} V_i + \log_{10} I_i + \mathbf{I}\omega \quad (2)$$

$$\delta_i \sim Normal(0, \tau_d) \quad (3)$$

where V_i is the proportion of 2.5 L filtered from Niskin i (in nearly all cases $V_i = 1$), I_i is the known dilution used on sample i to eliminate PCR inhibition, and $\mathbf{I}\omega$ is an estimated offset for an ethanol wash error. Here, \mathbf{I} is an indicator variable where $\mathbf{I} = 1$ for affected samples and $\mathbf{I} = 0$ otherwise (see also *ESM*).

When using qPCR, we do not directly observe eDNA concentration, we observe the PCR cycle at which each sample can be detected (or if it was never detected). We use a hurdle model to account for the fact that there is a probabilistic detection threshold (the PCR cycle of amplification is detected $G = 1$ or is not observed $G = 0$). Conditional on being detected, we observe the PCR cycle (C) as a continuous variable that follows a t distribution,

$$G_{ijr} \sim Bernoulli(\phi_{0j} + \phi_{1j} \log_{10} E_i) \quad (4)$$

$$C_{ijr} \sim T(\nu, \beta_{0j} + \beta_{1j} \log_{10} E_i, \eta) \quad if G_{ijr} = 1 \quad (5)$$

Here j indexes the PCR plate on which sample i and replicate r were run. We conducted 3 PCR reactions for each E_i . We fix the degrees of freedom for the t-distribution ($\nu = 3$) to allow for heavy-tailed observations and the parameter η is a scale parameter that controls the dispersion of the distribution. Note that there are different intercept (ϕ_{0j}, β_{0j}) and slope (ϕ_{1j}, β_{1j}) parameters for each PCR plate to allow for among-plate variation in amplification. See the *ESM* for additional components of the statistical model. We use diffuse prior distributions for all parameters (Table S1).

Acoustic-trawl data

In parallel with water collection for eDNA, we incorporated data on hake biomass derived from the contemporaneously collected data [26, 28], consisting of 57 acoustic transects totaling 4,483 km in length. 45 midwater trawls were deployed that provide information on the age, size and therefore signal strength of hake [22, 26, 28]. Methods for converting raw acoustic and trawl data to biomass concentrations can be found in [21, 22] and references therein. We used derived estimates of biomass concentration ($mt\ km^{-2}$) for hake ages 2 and older that integrate the biomass in the water column between depths of 50 and 500m in all analyses. All acoustic

177 data and associated trawls were collected during daylight hours. Therefore there was a lag between collection
 178 of aeoustic-acoustic-trawl and eDNA data, though for nearly all cases were separated by less than 24 hours.
 179 The temporal separation of eDNA and aeoustic-acoustic-trawl sampling precluded direct comparisons at the
 180 single-sample level.

181 **Spatial acoustic-trawl model**

182 In parallel with the model for qPCR data, we estimated a spatial model for the hake biomass derived from
 183 the acoustic-trawl survey. The biomass index created from the acoustic-trawl data for the entire survey area
 184 (34.4°N to 54.7°N) is used in stock assessments that determine the allowable catch and allocation of hake
 185 catch for fleets from the United States, Canada, and Tribal Nations [22]. As the eDNA samples only cover
 186 a portion of this range (38.3°N to 48.6°N), we used the biomass observations within this latitudinal range
 187 to generate spatially smooth estimates of biomass. Acoustic transects are divided into 0.926km (0.5nm)
 188 segments and the biomass (age 2 and older) concentration within each segment is used as data [22].
 189 Unlike the eDNA data, age-specific biomass estimates are available only as a biomass integrated across the
 190 entire water column (from depths of 50 to 500m). We fit a Bayesian hurdle model using a form similar to the
 191 eDNA, modeling biomass concentration (F_{xy} ; units: mt km^{-2}) using two separate spatial submodels: a) the
 192 probability of occurrence and b) abundance conditional on the presence of hake. We model both components
 193 as a function of bottom depth (smooth) and a spatial smooth,

$$H_{xy} \sim \text{Bernoulli}(\text{logit}^{-1}(\zeta_H + s_H(b) + t_H(x, y))) \quad (6)$$

$$F_{xy} \sim \text{LogNormal}(\zeta_F + s_F(b) + t_F(x, y) - 0.5\kappa^2, \kappa) \quad \text{if } H_{xy} = 1 \quad (7)$$

194 where H_{xy} is 1 if the observed biomass concentration is non-zero and 0 otherwise. In this formulation, ζ is
 195 the spatial intercept for each model component, $s(b)$ indicates a smoothing spline of as a function of bottom
 196 depth in meters (b), and $t(x, y)$ is a tensor-product smooth over latitude and longitude. κ is the standard
 197 deviation of the positive observations on the log scale. Table S2 provides the prior distributions for this
 198 model.

199 **Model Estimation**

200 We implemented both the eDNA and acoustic-trawl models using the Stan language as implemented in R
 201 (*Rstan*). All relevant code and data are provided in the online supplement. For the eDNA model, we ran 4

202 MCMC chains using 1,500 warm up and 9,000 sampling iterations. For the acoustic-trawl model, we ran 4
203 MCMC chains using 1,200 warm up and 3,000 sampling iterations.

204 We used traceplots and \hat{R} diagnostics to confirm convergence ($\hat{R} < 1.01$ for all parameters) – there were no
205 divergent transitions in the sampling iterations. To generate design matrices necessary for estimating covariate
206 effects we used the R package *brms* [28, 29, 32, 33]. We use diffuse prior distributions for all parameters (Table
207 S1). Posterior summaries of parameters can be found in the *ESM*.

208 Coordinate systems, covariates, and spatial predictions

209 We generated 5km resolution gridded maps for both the acoustic-trawl and eDNA models to enable direct
210 comparisons between models. This vector-based grid was developed and used by others [30, 34] for interpolating
211 various spatial models and uses a custom coordinate reference system that conserves area and distance
212 reasonably well across the west coast of the United States (*ESM*) and was a suitable resolution for the
213 purposes of our analyses.

214 To create spatial predictions for both eDNA and acoustic-trawl models, we took 4,000 draws from the joint
215 posterior and generated predictions for the centroid of each grid cell. We calculated posterior means and
216 uncertainty bounds among posterior draws. For the eDNA model we made projections for D_{xyb} ; we do not
217 present results from including additional observation processes on top of the estimated DNA concentrations in
218 the main text. We generated posterior predictive distributions for other model diagnostics checks (see *ESM*).

219 Creating an eDNA index

220 Our model provides direct predictions for hake DNA concentration at depths of 50, 100, 150, 300, and 500m.
221 To-The model lacks a term to directly make predictions to water depths other than those that were observed.
222 Therefore, to produce an index spanning depths of 50 to 500m, we equally weighted depths between 50 and
223 500m using linear interpolation between the closest depths. We used posterior predictions at each depth
224 to provide predicted DNA densities at 200, 250, 350, 400, and 450m for each 5km grid cell. Because some
225 spatial locations have depths of less than 500m, we only include predicted DNA concentrations to a depth
226 appropriate for the bathymetry (e.g. a location with a depth of 180m only includes values from 50, 100 and
227 150m). We sum across all depths (between 50 and up to 500m) to generate a depth-integrated index of hake
228 DNA. This index will be proportional to the hake DNA found in the water column. However, as we are only
229 summing across discrete depths, not integrating values across the entire water column nor multiplying by the
230 total water volume within each grid cell, the absolute value of the index will depend upon the number of
231 discrete depths we use. We refer to this as an eDNA abundance index to differentiate it from predictions to

232 for specific depths.

233 We compare estimates from the acoustic-trawl with the eDNA index using Pearson product-moment
234 correlations. We compare predictions from the methods at the scale of 25km^2 grid cells and after
235 aggregating estimates from each method into one degree latitudinal bins (for a total of 11 bins; Fig. 2).

236 Results

237 Hake were detected throughout the survey region (Fig. 1, 2), but hake DNA was far more commonly detected
238 than the ~~acoustic~~ acoustic-trawl signature of hake. As expected for a patchily distributed species, ~~acoustic~~
239 ~~acoustic-trawl~~ sampling identified hake biomass in a minority of 0.983km long transect segments (1764 of
240 4841; 36%). By contrast, ~~genetic signatures of hake were nearly ubiquitous, hake eDNA was~~ detected in 94%
241 of water samples (~~non-zero concentrations of hake DNA were quantified in~~ 1670 of 1769 2.5L samples) and
242 98% of sampling stations (875 of 892 stations), reflecting ~~considerable spatial smoothing of the~~ a considerable
243 ~~increase in detection of~~ eDNA signal relative to the ~~acoustic~~ acoustic-trawl detections.

244 ~~Distributions of hake~~ Hake DNA in the study area varied substantially ~~(estimated DNA concentration of~~
245 ~~individual samples ranged from below detection (< 20 copiesL⁻¹) to greater than 40,000 copiesL⁻¹)~~. DNA
246 ~~concentrations at stations – there are two water samples at each depth-station – varied strongly~~ with depth,
247 ~~with highest concentration between 100m~~ high estimated DNA concentrations at 150m (grand mean[range]
248 = 377[36 - 1,701] ~~copiesL⁻¹~~) and 300m depth (306[69 - 1,567]). DNA concentration at stations declined
249 ~~at both shallower (e.g. 50m: 180[44 - 535] copiesL⁻¹) and deeper depths (500m: 144[46 - 382] copiesL⁻¹)~~.
250 ~~Hake DNA showed notable spatial patterns, peaking~~ along the continental shelf break and south of the
251 Oregon-California border at 42°N (Fig. 1).

252 ~~There were also striking patterns in the variation in~~ Hake DNA concentration ~~was far more homogeneous~~
253 ~~at depth than near the surface: concentrations at 500m were generally low and showed limited spatial~~
254 ~~variation, while the near-surface layers (3m with depth (Fig. 1)). Specifically, deep stations (100m, 50m)~~
255 ~~showed generally higher but more heterogeneous concentrations (Figs. 4, 150m, S8) among individual sample~~
256 ~~bottles. The 300m and 500m) had relatively low uncertainty with regard to hake concentration (median~~
257 ~~coefficient of variation (CV) was larger than 1 of approximately 0.3), whereas the median CV for both 3m~~
258 ~~and 50m depths but only about 0.3 for depths 100m and deeper (Fig. 1G). Such uncertainty reflects large~~
259 ~~observed differences was larger than 1. Large CVs indicate both large bottle to bottle variation in DNA~~
260 ~~concentration between replicate samples taken at the same sample location and substantial differences among~~
261 ~~proximate within a station and substantial variation in hake eDNA concentration among nearby~~ sampling

262 locations (see *ESM*Figs. 1, S8).

263 We combined DNA information between 50 and 500m to produce both a spatially smooth, depth-integrated
264 estimate of hake DNA concentration (Fig. 2B). Separately, we generated a spatially smooth estimate of age
265 2+ biomass from the acoustic-trawl survey (Fig. 2C). ~~While spatial analyses for acoustic data are commonly~~
266 ~~used in assessing pelagic species worldwide, this is a first application of these methods with eDNA data. The~~
267 ~~eDNA biomass~~ ~~The eDNA abundance~~ index showed strong spatial patterning with highest values along the
268 continental shelf break with notable peaks in central ~~California~~ California and Oregon waters. In contrast,
269 acoustic-trawl observations were highly spatially variable – a common feature observed in acoustic surveys
270 [3135] – with high hake density and others with very low density in close proximity (see also Fig. S17). At the
271 scale of individual 25km^2 grid cells, eDNA and ~~acoustic-acoustic-trawl~~ surveys were modestly correlated ($\rho =$
272 $0.55[0.53, 0.57]$, Pearson product-moment correlation on posterior mean prediction [90% CI]; Fig. 3) but there
273 is considerable scatter in the relationship. Large eDNA values never occurred at locations which had very low
274 ~~acoustic-acoustic-trawl~~ biomass, but very high ~~acoustic-acoustic-trawl~~ estimates corresponded to moderate
275 values of eDNA. Notably, ~~acoustic-acoustic-trawl~~ biomass estimates had a very right-skewed distribution
276 across the 3,455 25km^2 ocean cells considered – most values were near zero with very few high values –
277 while eDNA values were decidedly less skewed (Fig. 3). Taken together, these observations again suggest a
278 smoother distribution of eDNA information relative to the patchier ~~acoustic-acoustic-trawl~~ detections.

279 When aggregated to one degree latitude bins, the correlation between eDNA and ~~acousties-acoustic-trawl~~
280 increased substantially ($\rho = 0.88[0.65, 0.96]$; Fig. 3) with ~~acousties-acoustic-trawl~~ and eDNA scaling
281 approximately linearly. Such increased correlation is not dependant upon the spatial groupings in Fig. 3 (see
282 Fig. S15 and S16 for results from an alternate spatial grouping). At this scale, eDNA and ~~acoustic-trawls~~
283 ~~acoustic-trawl~~ provide nearly equivalent information about relative biomass. At a coast-wide scale, the
284 uncertainties (CVs) of the acoustic-trawl estimate and eDNA index were nearly identical (both 0.09). This
285 similarity occurred despite the eDNA only being collected at 186 locations, whereas the acoustic-trawl data
286 includes 4,841 acoustic transect segments and 45 mid-water trawls to determine age- and length-structure of
287 the hake.

288 Finally, the two methods produced nearly identical latitudinal distributional estimates as measured by center
289 of gravity (median value within the projection range) and cumulative distribution (90% CIs overlapping for the
290 entire latitudinal range; Fig. 4). Furthermore, averaged across space, hake DNA concentrations were highest
291 along the continental shelf break (bottom depths between 125 and 400m) and at water depths between 150m
292 and 300m (Fig. 4C). All of these observations are consistent with published descriptions of hake depth and
293 habitat preferences [21, 22, ~~32, 33~~36].

294 **Discussion**

295 Ocean surveys are often used to generate large-scale, quantitative indices of species' abundances. At the
296 spatial scale relevant to management for hake along the U.S. west coast – our survey region encompasses the
297 majority of habitat for the Pacific hake stock – analysis of ~~a limited number of discrete water samples~~ ~~water~~
298 ~~samples taken~~ for eDNA provides comparable indices of hake biomass to acoustic-trawl surveys despite far
299 fewer eDNA observations. While other efforts have developed quantitative methods for eDNA within rivers
300 [10], lakes [15, ~~33~~37], estuaries [11], and nearshore marine habitats [12], we produce a large-scale study that
301 can serve as a template for using eDNA to determine abundance and species distributions with clear practical
302 applications to both conservation and fisheries. Importantly, ~~we push beyond our analysis demonstrated~~
303 ~~the value in analyses that push beyond simple~~ sample-to-sample comparisons ~~of~~ ~~between eDNA and other~~
304 alternate sampling methods ~~and make comparisons to make inferences~~ at the population-scale. The spatial
305 scale investigated here (~~on the order of~~ tens of thousands of km^2 ~~or more~~) is roughly comparable to the scale
306 at which most large ocean fisheries are managed both in the United States and internationally, suggesting
307 eDNA approaches can begin to be broadly adopted for that purpose.

308 The kind of spatial-statistical model we report here brings eDNA analysis inline with the methods currently
309 used in quantitative natural-resources management (~~[e.g. –~~ ~~27, 38~~34, 35]. Despite the clear differences in
310 biological processes producing eDNA signals versus acoustic trawl signals, these distinct data sets are both
311 subject to rigorous analytical methods. We emphasize that eDNA data here are processed independently
312 from acoustic-trawl data; no information from the ~~acoustics~~ ~~acoustic-trawl~~ informs eDNA or vice versa. Thus,
313 the implementation of eDNA surveys provides a second survey of abundance for hake without requiring
314 any additional days at sea, and should provide improved precision for estimated fish abundance when the
315 two indices are incorporated into a stock assessment. ~~Ultimately, eDNA~~ Additionally, the eDNA samples
316 are archived and can be used to investigate other species in future analyses. ~~eDNA~~ holds unprecedented
317 potential for improving the precision of abundance surveys, particularly when conducted in concert with
318 existing surveys.

319 For determining an index of abundance over a very large area, we assert that eDNA works well because
320 the concerns about the impact of DNA transport, degradation, and other processes [5, 7, 8] are negligible
321 for our application (providing an index of abundance on large spatial scales). Hake DNA present within
322 our survey boundaries was generated by hake present within the survey area; oceanographic processes like
323 currents or upwelling are not of sufficient magnitude to transport meaningful amounts of water into or out
324 of the survey domain on the time scale at which eDNA degrades [8]. Similarly, rates of DNA degradation
325 are expected to be consistent across our sampling domain – cool, offshore, oceanic waters below 50m with

326 relatively little among-sample variation in temperature, salinity, and other covariates identified as important
327 for degradation [37, 38, 39, 40]. Such population closure and constant rate assumptions are reasonable [see
328 also 12] and allow us to treat eDNA observations as analogous to other traditional sampling methods. We
329 note that our modeling framework provides the flexibility to directly include relevant covariates into the
330 observation model to account for relevant DNA processes if and when such information becomes available
331 (see *Materials and Methods* and *ESM*). For hake, our eDNA results match available geospatial (Figs. 2, 4)
332 and depth-specific patterns of hake abundance [21, 26] (Fig. 4) from other methods, strongly suggesting our
333 assumptions are reasonable and justified. eDNA approaches may be less effective in applications focused on
334 smaller temporal and spatial scales such as detailed habitat-association studies where the precise locations of
335 individuals are required [but see 39, 40, 41, 42].

336 Many challenges to implementing eDNA surveys remain. Surveys are primarily valuable because they inform
337 temporal trends; most surveys, particularly those of marine species, are not used as measures of absolute
338 abundance but as indices of abundance relative to previous years [22, 34, 38]. It will accordingly require
339 years to accumulate the kinds of eDNA-based time series that parallel those used in current management
340 and can be used in a stock assessment context. Furthermore, there are additional data streams needed for
341 management applications that are not currently possible from eDNA. For example, physical specimens are
342 needed to document age, size, sex, and condition, all of which cannot be extracted from eDNA at present,
343 though these are active areas of research [41, 42, 43, 44]. At present, eDNA approaches should be regarded as
344 supplementing existing surveys, not replacing them.

345 Despite these limitations, the characteristics of eDNA surveys have several advantages. First, the samples
346 collected and analyzed here for hake can be re-analyzed for other species. Analyses using species-specific
347 qPCR should provide similar quantitative data for additional species. DNA metabarcoding approaches
348 can detect many species simultaneously [1], but metabarcoding results are difficult to link to abundance or
349 biomass [18, 19]. Second, surveys of eDNA provide the potential for large-scale replication and high precision
350 because they only involve collecting water; as many replicate samples as desired can be collected, enabling
351 researchers to target and achieve a desired level of precision. Such replication is often not possible for other
352 sampling methods that involve capturing individuals. For example, repeatedly trawling a particular location
353 will deplete the fish present, and therefore such repeated sampling is generally not helpful for estimating
354 abundance. In theory there are few limits on replication using eDNA and our results indicate that the
355 amount of small-scale variation between water samples declines with depth (Fig. 1G, Fig. S9), suggesting
356 that the amount of statistical noise and therefore the amount of sampling needed may vary concomitantly.
357 It is wholly unknown if other marine species will exhibit similar depth-specific patterns of variability to

358 those observed in hake, though we hypothesize that the patterns observed may be related to the diel vertical
359 migration patterns of hake.

360 We developed and applied our eDNA approach to Pacific hake because of its broad geographic range, economic
361 importance, and decades of associated survey information. The ability of eDNA to provide similar indices of
362 abundance and distribution to existing surveys lend strong support for the applicability of eDNA methods to
363 the unstudied majority of species in ocean ecosystems. We believe eDNA will be particularly valuable for
364 understanding future changes in distribution of hake as well as other species, and future work will connect
365 eDNA surveys and oceanographic variables to understand shifts in species distributions.

366 Data accessibility

367 Additional methodology and analysis are provided in the electronic supplementary material and in the a Dryad
368 online data and code repository [4345]. <https://doi.org/10.5061/github.com/nwfsc-eb/eDNA-Hake-public/releases/tag/v1.0.2>
369 [dryad.n2z34tmzf](#)

370 Acknowledgements

371 Funding for this project was provided by the NMFS Genomic Strategic Initiative. Special thanks to S. Allen,
372 B. Dewees, J. Witmer and J. Davis for assistance during sample collection and the captain and crew of the
373 NOAA ship *Bell M. Shimada* for overall support during the survey cruise. A. Billings, J. Pohl, and members
374 of the Fisheries Engineering and Acoustic Technologies team provided logistical and analytical support for
375 both the acoustic and eDNA components of the survey. E. Iwamoto and members of the molecular genetics
376 lab at the Northwest Fisheries Science Center supported laboratory work. A. Berger, M.J. Ford, and J.F.
377 Samhour, and two anonymous reviewers provided helpful comments on earlier versions of the manuscript.

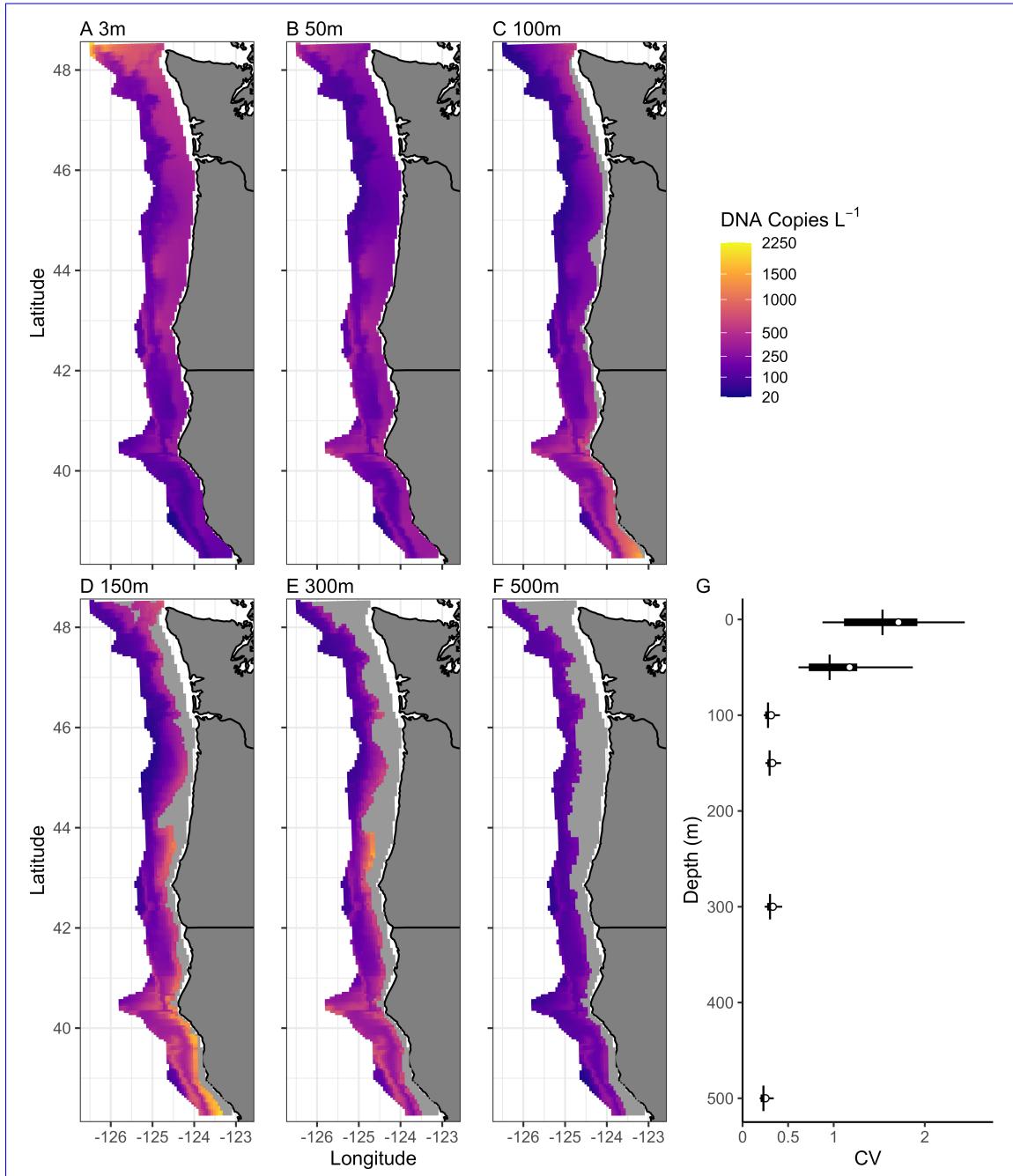


Figure 1: Predicted DNA concentration for six water depths shows clear spatial patterning in DNA concentration (A-F; posterior mean). G) Uncertainty around the posterior mean for each water depth as measured by the coefficient of variation. The distribution of CV among all projected 25km^2 grid cells are shown (mean(circles), median(vertical line), interquartile 50% and 90% CI shown).

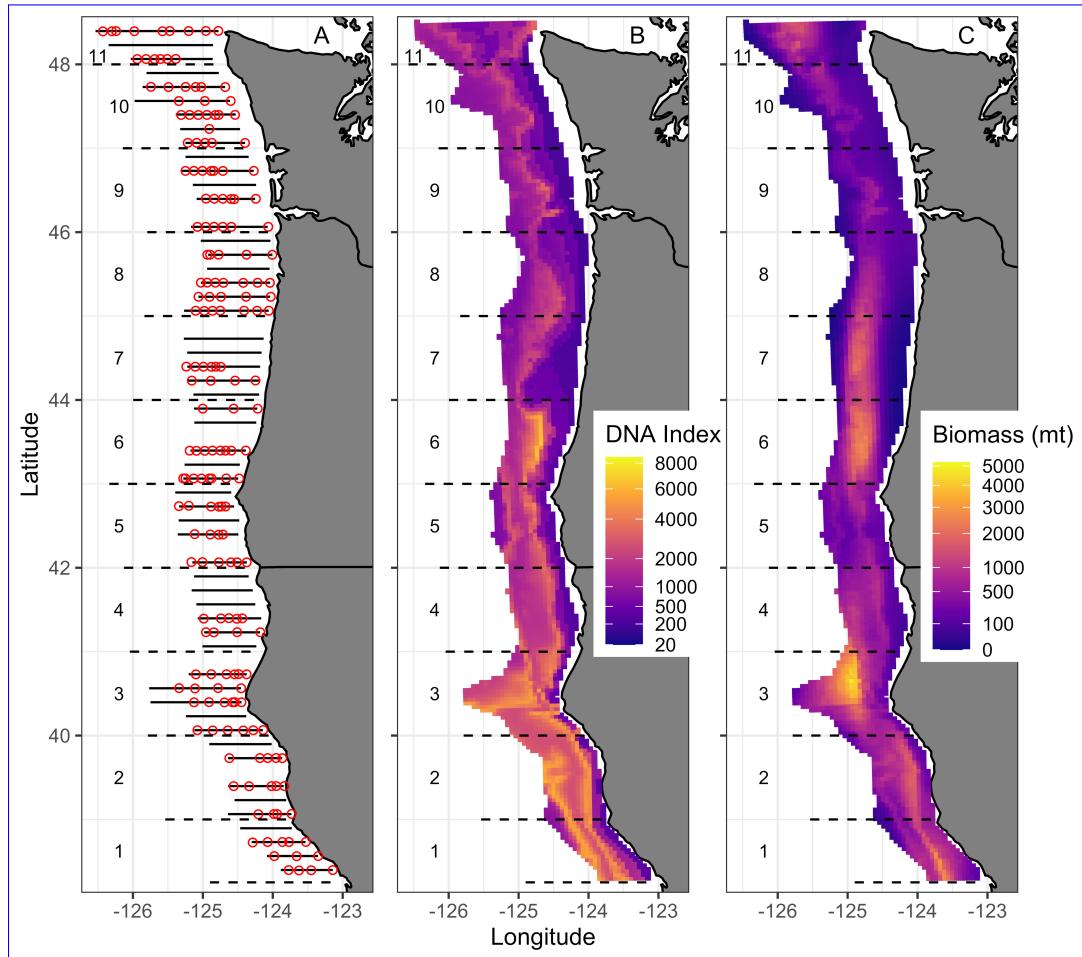
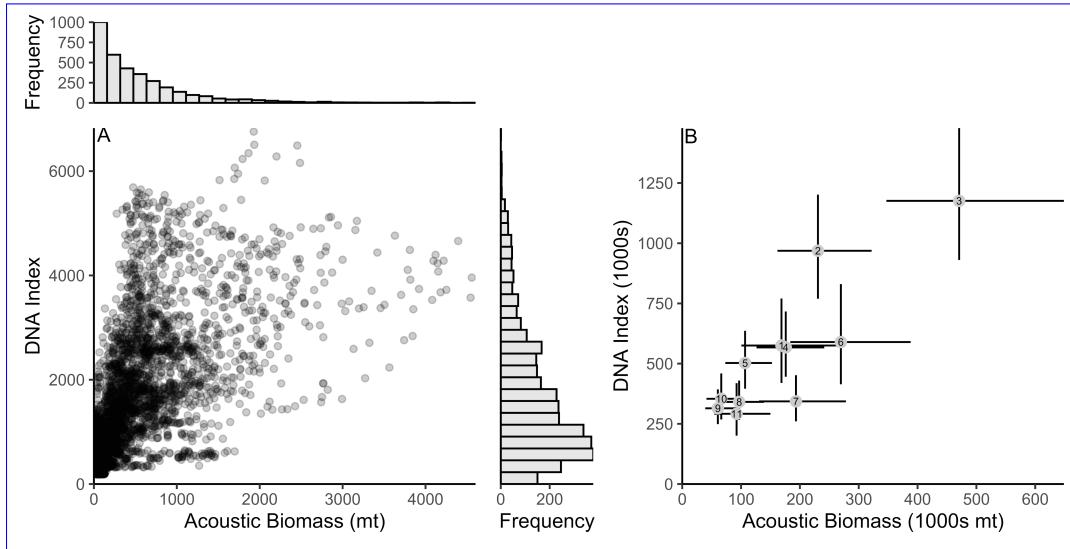


Figure 2: 2019 survey locations (A; red circles show eDNA sampling locations, lines show acoustic transects), depth-integrated index of hake DNA (B) and hake biomass from acoustic surveys (C). Both DNA and acoustic estimates are mean predicted values projected to a 5km grid and include information between 50 and 500m deep. All panels show one degree latitudinal bins ([numbered](#); [separated by dashed lines](#)) used to aggregate abundance estimates over larger spatial scales (see Fig. 3).



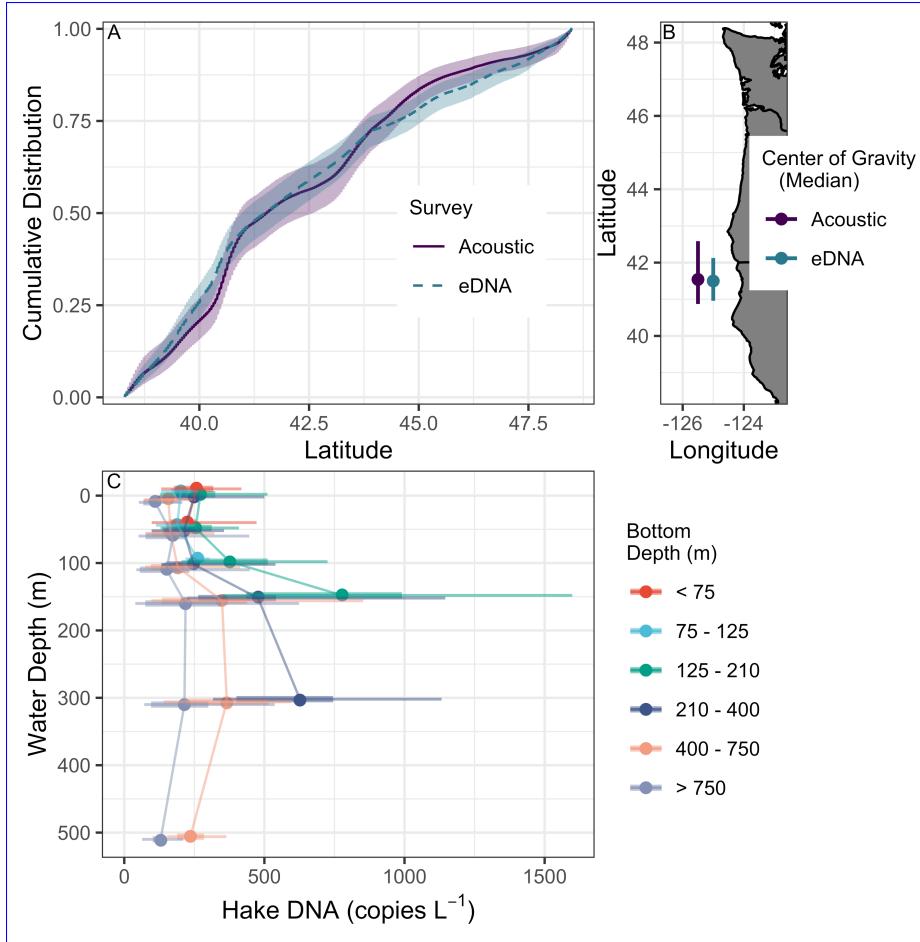


Figure 4: Estimates of distribution of Pacific hake. A) Cumulative distribution between 38.3 and 48.6°N (posterior means, 90% CI). B) Center of gravity (median of distribution) for each method (posterior means and 90% CI; only areas within the projection grid are included in this calculation ; see Figs. 1, 2). C) Posterior estimates of hake DNA concentration at each station-depth combination by the water depth sampled and categories of the depth of the bottom. The distribution of mean DNA concentration among station-depths (mean, interquartile range, and 90% CI among station-depths). Bottles at a sample location become increasingly similar at deeper sampling depths

378 **Citations**

- 379 1. Thomsen PF, Willerslev E. 2015 Environmental DNA—an emerging tool in conservation for monitoring
380 past and present biodiversity. *Biological Conservation* **183**, 4–18.
- 381 2. Boussarie G *et al.* 2018 Environmental DNA illuminates the dark diversity of sharks. *Science Advances* **4**,
382 eaap9661. (doi:10.1126/sciadv.aap9661)
- 383 3. Leempoel K, Hebert T, Hadly EA. 2020 A comparison of eDNA to camera trapping for assessment of
384 terrestrial mammal diversity. *Proceedings of the Royal Society B* **287**, 20192353.
- 385 4. Nguyen BN *et al.* 2020 Environmental DNA survey captures patterns of fish and invertebrate diversity
386 across a tropical seascape. *Scientific reports* **10**, 1–14.
- 387 5. Hansen BK, Bekkevold D, Clausen LW, Nielsen EE. 2018 The sceptical optimist: Challenges and
388 perspectives for the application of environmental DNA in marine fisheries. *Fish and Fisheries* **19**, 751–768.
- 389 6. Rourke ML, Fowler AM, Hughes JM, Broadhurst MK, DiBattista JD, Fielder S, Wilkes Walburn J, Furlan
390 EM. 2021 Environmental ~~dna~~-DNA (eDNA) as a tool for assessing fish biomass: A review of approaches and
391 future considerations for resource surveys. *Environmental DNA*
- 392 7. Barnes MA, Turner CR. 2016 The ecology of environmental DNA and implications for conservation
393 genetics. *Conservation genetics* **17**, 1–17.
- 394 8. Harrison JB, Sunday JM, Rogers SM. 2019 Predicting the fate of eDNA in the environment and implications
395 for studying biodiversity. *Proceedings of the Royal Society B* **286**, 20191409.
- 396 9. Rodríguez-Ezpeleta N *et al.* 2021 Biodiversity monitoring using environmental DNA. *Molecular Ecology*
397 *Resources* **21**, 1405–1409.
- 398 10. Carraro L, Hartikainen H, Jokela J, Bertuzzo E, Rinaldo A. 2018 Estimating species distribution and
399 abundance in river networks using environmental DNA. *Proceedings of the National Academy of Sciences*
400 **115**, 11724–11729.
- 401 11. Shelton AO, Kelly RP, O'Donnell JL, Park L, Schwenke P, Greene C, Henderson RA, Beamer EM. 2019
402 Environmental DNA provides quantitative estimates of a threatened salmon species. *Biological Conservation*
403 **237**, 383–391.
- 404 12. Fukaya K *et al.* 2021 Estimating fish population abundance by integrating quantitative data on
405 environmental DNA and hydrodynamic modelling. *Molecular ecology* **30**, 3057–3067.
- 406 13. Port JA, O'Donnell JL, Romero-Maraccini OC, Leary PR, Litvin SY, Nickols KJ, Yamahara KM, Kelly

- 407 RP. 2016 Assessing vertebrate biodiversity in a kelp forest ecosystem using environmental DNA. *Molecular*
408 *Ecology* **25**, 527–541.
- 409 14. Tillotson MD, Kelly RP, Duda JJ, Hoy M, Kralj J, Quinn TP. 2018 Concentrations of environmental
410 DNA (eDNA) reflect spawning salmon abundance at fine spatial and temporal scales. *Biological Conservation*
411 **220**, 1–11.
- 412 15. Hänfling B, Lawson Handley L, Read DS, Hahn C, Li J, Nichols P, Blackman RC, Oliver A, Winfield IJ.
413 2016 Environmental DNA metabarcoding of lake fish communities reflects long-term data from established
414 survey methods. *Molecular ecology* **25**, 3101–3119.
- 415 16. Stoeckle MY, Adolf J, Charlop-Powers Z, Dunton KJ, Hinks G, VanMorter SM. 2021 Trawl and eDNA
416 assessment of marine fish diversity, seasonality, and relative abundance in coastal New Jersey, USA. *ICES*
417 *Journal of Marine Science* **78**, 293–304.
- 418 17. Knudsen SW *et al.* 2019 Species-specific detection and quantification of environmental ~~dna~~DN from
419 marine fishes in the baltic sea. *Journal of Experimental Marine Biology and Ecology* **510**, 31–45.
- 420 18. Kelly RP, Shelton AO, Gallego R. 2019 Understanding PCR processes to draw meaningful conclusions
421 from environmental DNA studies. *Scientific Reports* **9**, 1–14.
- 422 19. McLaren MR, Willis AD, Callahan BJ. 2019 Consistent and correctable bias in metagenomic sequencing
423 experiments. *Elife* **8**, e46923.
- 424 20. Zwolinski J, Fernandes PG, Marques V, Stratoudakis Y. 2009 Estimating fish abundance from acoustic
425 surveys: calculating variance due to acoustic backscatter and length distribution error. *Canadian Journal of*
426 *Fisheries and Aquatic Sciences* **66**, 2081–2095.
- 427 21. Malick MJ, Hunsicker ME, Haltuch MA, Parker-Stetter SL, Berger AM, Marshall KN. 2020 Relationships
428 between temperature and Pacific hake distribution vary across latitude and life-history stage. *Marine Ecology*
429 *Progress Series* **639**, 185–197.
- 430 22. Grandin CJ, Johnson KF, Edwards AW, Berger AM. 2020 Status of the Pacific hake (whiting) stock in
431 U.S. and Canadian waters in 2020., 273.
- 432 23. Ressler PH, Holmes JA, Fleischer GW, Thomas RE, Cooke KC. 2007 Pacific hake, *Merluccius productus*,
433 autecology: A timely review. *Marine Fisheries Review* **69**, 1–24.
- 434 24. Methot RD, Dorn MW. 1995 Biology and fisheries of north Pacific hake (*M. productus*). In *Hake*, pp.
435 389–414. Springer.

- 436 25. Ramón-Laca A, Wells A, Park L. 2021 A workflow for the relative quantification of multiple fish species
437 from oceanic water samples using environmental DNA (eDNA) to support large-scale 2021. *PLoS ONE* **16**,
438 e0257773.
- 439 24. Methot RD, Dorn MW. 1995 Biology and fisheries of north Pacific hake (*M. productus*). In *Hake*, pp.
440 389–414. Springer.
- 441 26. Cressie N, Wikle CK. 2015 Statistics for spatio-temporal data. John Wiley & Sons.
- 442 25. Ressler PH, Holmes JA, Fleischer GW, Thomas RE, Cooke KC. 2007 Pacific hake, *Merluccius productus*,
443 autecology: A timely review. 27. Thorson JT, Shelton AO, Ward EJ, Skaug HJ. 2015 Geostatistical
444 delta-generalized linear mixed models improve precision for estimated abundance indices for west coast
groundfishes. *ICES Journal of Marine Fisheries Review* **69** 72, 1–241297–1310.
- 445 26–28. de Blois S. 2020 The 2019 joint U.S.–Canada integrated ecosystem and Pacific hake acoustic-trawl
446 survey: Cruise report sh-19-06. U.S. Department of Commerce, NOAA Processed Report NMFS-NWFSC-
447 PR-2020-03
- 448 27–29. Renshaw MA, Olds BP, Jerde CL, McVeigh MM, Lodge DM. 2015 The room temperature preservation
449 of filtered environmental DNA samples and assimilation into a phenol–chloroform–isoamyl alcohol DNA
450 extraction. *Molecular Ecology Resources* **15**, 168–176.
- 451 30. Clark JS, Bjørnstad ON. 2004 Population time series: process variability, observation errors, missing
452 values, lags, and hidden states. *Ecology* **85** 3140–3150.
- 453 31. Hilborn R, Mangel, M. 1997 The ecological detective: confronting models with data. Princeton University
454 Press, Princeton, New Jersey.
- 455 28–32. Bürkner P-C. 2017 brms: An R package for Bayesian multilevel models using Stan. *Journal of
456 Statistical Software* **80**, 1–28. (doi:10.18637/jss.v080.i01)
- 457 29–33. Bürkner P-C. 2018 Advanced Bayesian multilevel modeling with the R package brms. *The R Journal*
458 **10**, 395–411. (doi:10.32614/RJ-2018-017)
- 459 30–34. Feist BE, Samhouri JF, Forney KA, Saez LE. 2021 Footprints of fixed-gear fisheries in relation to
460 rising whale entanglements on the US West Coast. *Fisheries Management and Ecology* **28**, 283–294.
- 461 31–35. Rivoirard J, Simmonds J, Foote K, Fernandes P, Bez N. 2008 *Geostatistics for estimating fish
462 abundance*. John Wiley & Sons.
- 463 32–36. Agostini VN, Francis RC, Hollowed AB, Pierce SD, Wilson C, Hendrix AN. 2006 The relationship
464 between Pacific hake (*Merluccius productus*) distribution and poleward subsurface flow in the California

- 465 Current System. *Canadian Journal of Fisheries and Aquatic Sciences* **63**, 2648–2659.
- 466 33.–37. Buxton AS, Groombridge JJ, Zakaria NB, Griffiths RA. 2017 Seasonal variation in environmental
467 DNA in relation to population size and environmental factors. *Scientific reports* **7**, 1–9.
- 468 34.–38. Maunder MN, Punt AE. 2004 Standardizing catch and effort data: A review of recent approaches.
469 *Fisheries research* **70**, 141–159.
- 470 35. Thorson JT, Shelton AO, Ward EJ, Skaug HJ. 2015 Geostatistical delta-generalized linear mixed models
471 improve precision for estimated abundance indices for west coast groundfishes. *ICES Journal of Marine
472 Science* **72**, 1297–1310.
- 473 36. Murakami H *et al.* 2019 Dispersion and degradation of environmental dna from caged fish in a marine
474 environment. *Fisheries science* **85**, 327–337.
- 475 37.–39. Collins RA, Wangensteen OS, O’Gorman EJ, Mariani S, Sims DW, Genner MJ. 2018 Persistence of
476 environmental DNA in marine systems. *Communications Biology* **1**, 1–11.
- 477 38.–40. Andruszkiewicz Allan E, Zhang WG, C Lavery A, F Govindarajan A. 2021 Environmental dna-DNA
478 shedding and decay rates from diverse animal forms and thermal regimes. *Environmental DNA* **3**, 492–514.
- 479 39.–41. Pont D *et al.* 2018 Environmental DNA reveals quantitative patterns of fish biodiversity in large
480 rivers despite its downstream transportation. *Scientific Reports* **8**, 10361.
- 481 40.–42. Jeunen G-J, Knapp M, Spencer HG, Lamare MD, Taylor HR, Stat M, Bunce M, Gemmell NJ.
482 2019 Environmental dna-DNA (eDNA) metabarcoding reveals strong discrimination among diverse marine
483 habitats connected by water movement. *Molecular Ecology Resources* **19**, 426–438.
- 484 41.–43. Adams CI, Knapp M, Gemmell NJ, Jeunen G-J, Bunce M, Lamare MD, Taylor HR. 2019 Beyond
485 biodiversity: Can environmental DNA (eDNA) cut it as a population genetics tool? *Genes* **10**, 192.
- 486 42.–44. Mayne B, Korbie D, Kenchington L, Ezzy B, Berry O, Jarman S. 2020 A DNA methylation age
487 predictor for zebrafish. *Aging* **12**, 24817–24835. (doi:10.18632/aging.202400)
- 488 43.–45. Shelton AO, Ramón-Laca A, Wells A, Clemons J, Chu D, Feist BE, Kelly RP, Parker-
489 Stetter SL, Thomas R, Nichols K, Park L. 2021 Environmental DNA provides quantitative esti-
490 mates of Pacific hake abundance and distribution in the open ocean. Available at :—https:Dryad:
491 doi:10.5061/github.com/nwfsc-eb/eDNA-Hake-public/releases/tag/v1.0.2 dryad.n2z34tmzf