

Mapping the marine distribution of eulachon (*Thaleichthys pacificus*) in the Northeast Pacific using environmental DNA

Received: 6 June 2025

Accepted: 9 February 2026

Cite this article as: Liu, O.R., Shelton, A.O., Ramón-Laca, A. *et al.* Mapping the marine distribution of eulachon (*Thaleichthys pacificus*) in the Northeast Pacific using environmental DNA. *Commun Biol* (2026). <https://doi.org/10.1038/s42003-026-09733-5>

Owen R. Liu, Andrew O. Shelton, Ana Ramón-Laca, Krista M. Nichols, Eric J. Ward, Elizabeth M. Phillips, Jeannette E. Zamon, Abigail Wells & Ryan P. Kelly

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

Title

Mapping the marine distribution of eulachon (*Thaleichthys pacificus*) in the Northeast Pacific using environmental DNA

Authors

Owen R. Liu^{1,2*}, Andrew O. Shelton³, Ana Ramón-Laca⁴, Krista M. Nichols³, Eric J. Ward³, Elizabeth M. Phillips⁵, Jeannette E. Zamon⁶, Abigail Wells³, Ryan P. Kelly¹

¹School of Marine and Environmental Affairs, University of Washington, 3707 Brooklyn Ave NE, Seattle, WA 98195 USA

²School of Aquatic and Fishery Sciences, University of Washington, 1122 NE Boat St Box 355020, Seattle, WA 98195 USA

³Conservation Biology Division, Northwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, 2725 Montlake Blvd E, Seattle, Washington, 98112 USA

⁴Museo Nacional de Ciencias Naturales (CSIC), Madrid 28006, Spain

⁵Fishery Resource Analysis and Monitoring Division, Northwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, 2725 Montlake Blvd E, Seattle, Washington, 98112 USA

⁶Fish Ecology Division, Northwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, , 2725 Montlake Blvd E, Seattle, Washington, 98112 USA

*corresponding author: owenrlu@.uw.edu

Keywords: environmental DNA, eulachon, species distribution model, endangered species

Abstract:

Rare species are difficult to observe in the wild, particularly in the ocean where large spatial scales and accessibility hinder effective sampling. Environmental DNA (eDNA) is a non-destructive, scalable sampling method with the potential to inform the distribution of rare species in marine ecosystems. We sample eDNA within the California Current ecosystem to estimate the distribution of eulachon (*Thaleichthys pacificus*), a threatened anadromous smelt ranging along the coastal Northeast Pacific. We amplify eulachon DNA from thousands of water samples collected at night across two years and more than 200,000 square kilometers along the U.S. west coast. We then use spatiotemporal models to derive quantitative estimates of eulachon DNA across space, depth, and time relative to environmental covariates. We find that eulachon DNA has a distribution weighted towards the ocean surface, spatially associated with major river mouths and productive offshore banks. Temperature and prey density are key covariates, with eulachon more likely to be found in warmer waters with higher prey concentrations. We discuss how our results can augment the information currently used in eulachon recovery planning, and describe the wide applicability of our statistical models for estimating distribution and abundance for other species of conservation concern.

Introduction:

Threatened and endangered species are generally rare and therefore difficult to observe. These difficulties are exacerbated in environments that are remote or costly to access, or for species that occupy multiple habitats seasonally or ontogenetically. At the same time, such species are often precisely those for which we need the most information to ensure their conservation. This conflict is the “rare species modeling paradox”¹, when species requiring the greatest protection – and thus with the greatest information demand² – are those for which we have the least data, making modeling particularly challenging.

Resolving this paradox requires finding or designing tools that have high sensitivity—that is, a high chance of observing a rare species of interest, given its presence^{2,3}—but are minimally invasive or destructive. For marine species in the open ocean this requirement can be particularly daunting: access to direct observation is severely limited, and monitoring often relies on destructive sampling (e.g. fishing, which often involves bycatch) or other methods that require significant interaction with the species of interest such as tagging and biopsy sampling. These methods, while well-established, can lead to potential harm or mortality of threatened species.

In contrast to these traditional methods, environmental DNA (eDNA) analysis offers sensitive, species-specific resolution without destructive sampling, and can be cost-effective to scale up to large numbers of samples^{4,5,6}. Moreover, existing environmental samples can be queried many times for various species or communities of interest^{7,8}, and emerging statistical tools can use replicated sampling designs – common in eDNA studies – to reflect both process and observation variation in phenomena of interest⁹. However, eDNA sampling has still rarely been performed across multiple years (though see Sildever et al.¹⁰ for a time series example) and at an ecosystem scale simultaneously, and has seldom moved beyond presence-absence data into the realm of robust quantitative estimates for threatened or endangered species.

Here, we address this gap by using quantitative measures of eDNA concentration collected over a large spatial scale to characterize the marine distribution of eulachon (*Thaleichthys pacificus*) along the west coast of the USA, under the reasonable assumption that eDNA concentration is positively correlated with eulachon abundance⁶. Eulachon is a semelparous anadromous smelt endemic to the northeast Pacific Ocean and distributed from Alaska to northern California. South of the Nass River in British Columbia, Canada, the species is listed as threatened under the U.S. Endangered Species Act¹¹ largely due to substantial declines in the returns of the species to known spawning rivers^{12,13}.

Though eulachon spend more than 95% of their life in the ocean, their marine distribution remains poorly characterized^{13,14}. To inform future conservation planning and recovery efforts, we present the first eDNA-derived, spatiotemporal distribution model for eulachon. The model is trained on an eDNA dataset derived from thousands of discrete water samples collected across two sampling years, ten degrees of latitude, and approximately 200,000 km² of open ocean. Our

analysis is an example of the integration of spatial statistics and eDNA observations at an ecosystem scale, and the approach is both scalable and extensible to other species and ecosystems.

Results:

To characterize the ocean distribution of eulachon eDNA, we collected water samples during a large-scale effort aboard the NOAA Ship *Bell M. Shimada* between July 2 and August 19, 2019 and between July 2 and September 21, 2021 (Figure 1). The sampling was performed at multiple stations along cross-shore transects (one to nine stations per transect), and two replicate water samples (biological replicates) were collected for eDNA analysis at up to six depths at each station. In the lab, biological samples were analyzed using at least three technical replicates and processed for eulachon using a eulachon-specific quantitative PCR (qPCR) assay¹⁵ (Table 1). Negative controls were collected and analyzed at multiple steps along the pipeline to identify any possible issues with contamination. For model fitting, we chose to use data only from the surface, 50m, and 150m depths to describe the 3-dimensional distribution of eulachon DNA. The selected data include water samples from 908 unique location-depth combinations in 2019, and 899 unique location-depth combinations in 2021. Altogether—including biological replicates and technical replicates produced in the lab—we produced 4175 and 3328 statistical samples for 2019 and 2021, respectively. Of these, 459 and 367 samples (approximately 11.0% of samples in both years) had observed amplification of eulachon DNA (Fig. 1, Fig. S1).

We also collected information on environmental covariates hypothesized to be important in shaping eulachon life history and distribution. These included water temperature and salinity from an ocean model¹⁶, freshwater discharge volumes from major US west coast rivers¹⁷, and estimates of euphausiid abundance¹⁸. Also known as krill, euphausiids are a primary prey item for adult eulachon in this region, with the main species being *Euphausia pacifica* and *Thysanoessa spinifera*^{12,19}. Using these covariates, we built a species distribution model (SDM) to describe the distribution of eulachon DNA across space (34.5 to 48.5 degrees N), depth (surface to 150m), and time (2019 and 2021). For details on covariate construction and model selection, see Methods.

Even though eulachon DNA was detected infrequently among the collected samples, (3-17% of samples depending upon year and depth category), we were able to successfully build an SDM describing eulachon DNA concentration and distribution. We estimate that eulachon are relatively widespread along the U.S. west coast, but this distribution is not uniform (Figure 2). In general, higher eulachon DNA abundances are found towards the northern extent of our study domain, particularly off the coast of central Oregon and Washington (Figure S1). Interestingly, however, there were a small number of amplified samples from the southern portion of the study domain between Point Conception (34.5°N) and the southernmost known eulachon spawning river (Mad River, 40.5°N) (Figures S1, S10). This suggests eulachon could be distributed further south in the California Current Ecosystem than previously thought. Additionally, nighttime eulachon distribution varies by depth, and eulachon DNA has a much more widespread

distribution at the surface than in deeper samples. By projecting our fitted model onto a regular grid, we produced an index of total eulachon DNA abundance by depth (Figure 3). The total estimated abundance of eulachon DNA was highest at the surface and lowest at 50m. It also varied between 2019 and 2021: 2021 had greater estimated eulachon DNA at all depths and across the coast, except for surface waters in Oregon which had higher estimated eDNA in 2019.

We expected to find a relationship between the distribution of eulachon and major rivers because larval and juvenile eulachon enter the ocean from riverine habitats and occur in surface waters during their early marine residence^{20,21}. We constructed a spatially-interpolated measure of river outflow to test as a covariate in the SDMs, but this metric was ultimately not among the environmental predictors selected in our final model (see Methods). However, there is a clear association along the coast between some major river mouths and the latitudinal distribution of eulachon DNA (Figure 4). Our highest estimates of eulachon DNA concentration are in areas just north of the mouth of the Columbia River (46°N), which is the largest river by volume on the west coast and hosts the largest existing spawning runs of eulachon in the study area^{12,13}. Other, smaller latitudinal peaks in eulachon DNA abundance occur between the mouths of the Eel and Klamath rivers around 41°N. The Klamath River historically supported large spawning runs prior to the 1980s²². The second largest latitudinal peak in eulachon DNA around 44-45°N is likely also associated with the Columbia River, which frequently produces a bidirectional plume extending both north and south of the river mouth²³, depending on the time of year and prevailing, alongshore downwelling- or upwelling-favorable winds.

Our selected eulachon distribution model included temperature (Figure S2) and a measure of relative krill abundance (Figure S3) as key covariates. Eulachon DNA concentration increases with temperature above approximately 11°C. This result mirrors the observed pattern of greater eulachon DNA concentration at the surface, because temperature is inversely related to depth and the samples were matched to depth-specific temperature values from the ocean model (Figure 5a). As expected from previous studies¹², eulachon also has a positive, but saturating, relationship with estimated euphausiid abundance (Figure 5b). Although the two species could simply be responding to similar environmental gradients, it is likely that they interact, since euphausiids are a known and strongly preferred prey for adult eulachon. This trophic relationship has been shown previously only through stomach-content analyses^{19,21}, but the overlap of their respective spatial distributions here provides additional evidence of strong ecological interactions between eulachon and euphausiids.

Eulachon DNA is also associated with well-known bathymetric and oceanographic features in the region that are often cited as areas of concentrated biological productivity. Across the study domain, we estimate high concentrations of eulachon DNA near the 150m isobath, at both the surface and at depth (Figure 2). We also estimated elevated eulachon DNA concentrations at the northern boundary of the study domain at 48°N, near the western entrance of the Strait of Juan de Fuca. This is an area of consistently high biological productivity due to outflow of estuarine waters from Puget Sound mixing with entrained upwelled nutrients from the deep, productive waters of the Juan de Fuca Canyon into the persistent Juan de Fuca eddy²⁴. In

addition, we estimate high concentrations of eulachon DNA around Heceta and Stonewall Banks in central Oregon, and just north of Astoria Canyon, near the mouth of the Columbia River. Heceta Bank and Stonewall Bank are part of a known upwelling region and hydrographic retention area²⁵ that concentrates biological productivity. The region has been identified as an important foraging habitat for a number of species, including krill²⁶ and rorqual whales^{27,28}. Although the high biological productivity found here would make it a suitable habitat for eulachon, Heceta and Stonewall Banks have not been previously identified as a eulachon hotspot. This may be partially due to the fact that the only previous analysis to estimate eulachon ocean hotspots¹⁴ relied on fishery-dependent data (eulachon bycatch in a shrimp fishery) to estimate the distribution, and Heceta Bank is closed to commercial fishing. Regardless, the significant association between bathymetric features, known productivity hotspots, and eulachon deserves further study to validate our estimates.

Discussion:

We have shown that even for rare taxa across hundreds of thousands of square kilometers in the open ocean, eDNA sampling can be used in conjunction with spatiotemporal modeling to provide new information on the likely distribution of species of concern. To the best of our knowledge, this is the first study to successfully use environmental DNA to identify and locate eulachon in the ocean, providing significant advances in our understanding of their marine distribution and life history. Eulachon face many challenges associated with human impacts across their anadromous life history, and are listed as threatened in the southern portion of their range under the U.S. Endangered Species Act. While much of the information used in eulachon recovery planning and regulations comes from monitoring efforts during annual spawning runs in freshwater ecosystems, there has been a paucity of information available concerning eulachon in the marine habitats where they spend the majority of their lives. Analysis of eDNA helps to address the rare species modeling paradox for eulachon because it is non-destructive, does not require direct human observation of the species of concern, and is sensitive enough to reveal a three-dimensional, time-varying spatial distribution. Through our analytical and modeling approach, not only did we detect eulachon—which is non-trivial given the rarity of eulachon, lack of existing knowledge on eulachon distribution, and the scale of the study area—but we were also able to provide a preliminary characterization of the species' preferred marine habitats relative to environmental covariates.

Our eDNA-based SDM suggests that eulachon on the US west coast are associated with river mouths, offshore productive banks, and biological retention zones. These findings are coherent with, but expand upon, current best available scientific knowledge of the species¹³. The potential hotspots of eulachon abundance in the ocean—the Juan de Fuca Eddy, Heceta and Stonewall Banks, and the dynamic Columbia River plume—may benefit from more dedicated study. Confirmation of these areas as important marine habitats for eulachon could lead to more informed conservation efforts for the species, potentially including further fisheries bycatch mitigation measures in areas with high eulachon concentrations, or as areas within which to conduct directed studies to further develop our understanding of eulachon marine life history.

While our study was not designed to explicitly estimate overall eulachon population size, our derived eDNA index (Figure 3) compares favorably with other indirect indices of eulachon abundance for 2019 and 2021. Gustafson et al.²² compiled data on eulachon run size and bycatch in ocean fisheries (primarily the pink shrimp and groundfish fisheries), finding that since a low in 2018, indices of eulachon abundance have increased steadily. Our results agree with this assessment, with larger eDNA index estimates in 2021 compared to 2019 for all state and depth combinations except for surface samples from Oregon, which were higher in 2019. Even this pattern, though, is qualitatively similar to the finding of Gustafson et al.²² that bycatch of eulachon was greater in Oregon in 2019 than in 2021 in both an absolute sense, and relative to total shrimp catch (i.e., a bycatch ratio).

There was also a difference in depth distribution of eulachon DNA between the two sample years. The reason behind this interannual variability in depth distribution is unclear—and difficult to characterize with only two years of data—but could be due to changes in age structure²⁹ or different sampling periods (many 2021 samples were collected approximately one month later than 2019 samples). In both years, however, the highest concentrations of eulachon DNA were found in surface samples. This result was surprising because eulachon is generally considered a demersal (i.e., bottom-dwelling) species in the adult stage. However, juvenile and adult eulachon have been observed in near-surface waters at night, including near the Columbia River mouth^{30,31}. These observations, combined with our own nighttime collection of eDNA, suggests the existence of a diel vertical migration behavior in eulachon, likely tracking a well-documented diel migration of their euphausiid prey³².

Alternatively, it is possible that our observation of eulachon eDNA near the surface could be due to genetic material from eggs and larvae, or movement of free DNA in the water column separate from the organisms themselves. Eulachon eggs and larvae are thought to be distributed more epipelagically than adults, but the timing of our survey (July to September) likely precludes the possibility that we were sampling DNA from those life stages. Eulachon spawning is generally finished by May in this region¹², and given the relatively brief period of eDNA detectability after shedding^{33,34}, and the fact that our surveys never sampled before July 2, it is unlikely that we were capturing eDNA transported by river discharge. Another potential alternative explanation is that, once shed, eulachon eDNA moves vertically in the water column. However, vertical mixing likely has little influence on observed eDNA stratification when placed in the context of other processes such as decay and horizontal advection³⁵. It is also plausible that some of the eulachon DNA in surface waters was derived from the waste products of its predators, who may capture eulachon at depth but excrete their DNA at a different depth. While this is a possible mechanism of eDNA transport, it is extremely unlikely that this phenomenon is a primary driver of the observed pattern across the entire dataset—after all, eulachon DNA concentration was consistently higher at the surface across the study domain, not just in one or two samples.

Additional years of eDNA data, beyond the two sample years considered here, could allow a more comprehensive comparison between eDNA and other time series, such as estimates of eulachon distribution and abundance derived from fisheries bycatch data¹⁴, and from

observations of eulachon spawning runs in major west coast rivers²². From a threatened species monitoring perspective, additional data could also allow for the study of optimal sampling, to find the minimum required number of environmental samples for robust distribution estimation. Similarly, more within-year sampling across seasons could help distinguish any seasonal distribution changes from interannual shifts. At the same time, our results could be further strengthened and validated in the future by ground-truthing the eDNA with *in situ* observations of eulachon at different depths.

There are also other sampling and laboratory approaches that could help further our knowledge of the marine distribution and ecology of eulachon. Future sampling schemes, for instance, could include collecting eDNA at multiple times of day in the same location to determine whether there truly is a diel pattern in eulachon feeding behavior – i.e., whether surface eDNA concentrations are greater in the same location during the night than during the day. Further, eDNA tools can in theory be used to extract genetic information to ascertain the riverine origin of eulachon populations and determine whether these populations form mixed schools in the ocean. Studies of genetic stock structure using eulachon tissue samples have been performed previously³⁶, but further exploration is warranted to explore the idea of utilizing eDNA in this way. Currently, it is difficult to do stock analysis with eDNA in part because while tissue samples come from known individuals, a given eDNA sample can potentially reflect multiple genetic compositions due to varying mixtures of many individual fish. However, new laboratory and statistical methods are being developed to determine age or life stage information directly from eDNA or eRNA, which could provide richer, size- or age-structured eDNA data to expand our scope of inference^{37,38}.

As a generalized statistical approach using publicly available software, our approach to eDNA data analysis is broadly applicable to many other species and habitats where there is a need to characterize species' spatial distributions. This is not limited to only rare species: the same water samples used in this study are being separately analyzed to predict the distribution and abundance of Pacific hake, one of the most abundant fish species in the California Current, and the target of a \$60 million fishery annually⁶. As statistical models for eDNA continue to improve, the possibilities for their application to conservation problems will only expand. Moreover, because of the nature of our data collection (filtered water samples), eDNA analysis is free of some of the challenges and biases associated with traditional fisheries data collection like the need to kill fish in order to sample them. Although eDNA collection and processing comes with its own set of challenges—including, but not limited to, designing an appropriate spatial sampling scheme, ensuring appropriate primer design, risk of sample contamination, and the need for PCR amplification and significant quantitative processing to translate from raw data to abundance estimates—we have provided a modeling framework that can address these challenges to simultaneously estimate presence/absence and positive abundance of eDNA in a statistically robust way. When appropriately processed and modeled, eDNA data can provide quantitative and novel information for threatened and endangered species monitoring and assessment that is complementary to existing tools.

Methods:

We used environmental DNA collected in the field in 2019 and 2021 to build a spatial model of eulachon (*Thaleichthys pacificus*) distribution along the Pacific coast of the United States.

Field sampling and processing of eDNA

eDNA samples were collected during the U.S.-Canada Integrated Ecosystem & Acoustic-Trawl Survey in 2019 and 2021. The survey took place aboard the NOAA Ship *Bell M. Shimada* from July 2 to August 19, 2019 and July 2 to September 21, 2021. Sampling took place at 187 stations along 37 transects between 38.0-48.5 N in 2019, and 202 stations along 42 transects between 34.5-48.5 N in 2021 (Figure 1). Depending on the maximum water depth at a given survey station, samples were consistently collected at six different depths—surface, 50, 100, 150, 300, and 500m. To avoid conflict with other survey activities occurring on the *Shimada*, sampling took place during the night, from sunset to sunrise. Additionally, to identify and reduce the risk of contamination from these other activities, negative controls were collected and analyzed at multiple stages along our analysis pipeline, including at the biological sampling, DNA extraction, and PCR steps.

At each station and depth, two replicate 2.5L seawater samples were collected from two Niskin bottles on a CTD rosette, except for surface samples, which were collected from the *Shimada*'s saltwater intake line. Negative controls were collected by filtering 2L of either distilled water from the onboard evaporator or deionized laboratory water (N=36 evaporator and 17 laboratory water samples, respectively, in 2019, and N=80 evaporator and 44 laboratory water samples in 2021). These negative controls were taken at random times of night during the collection and filtering of field samples. In addition, 50 “air samples” were taken in 2021 by filtering air only, to control for potential ambient DNA in the lab. Both samples and controls were filtered immediately following collection, using 47mm mixed cellulose ester sterile filters (1 μ m pore size), and preserved with Longmire's buffer at room temperature. If filters became clogged two or three filters (in the same extraction) were used to filter the 2.5 liters successfully. DNA was extracted using a phenol:chloroform method modified with a phase lock to increase throughput of the large number of samples.

We used a eulachon-specific TaqMan assay for the qPCR to quantify eulachon DNA in triplicate for each sample¹⁵ (Table 1). The assay includes an internal positive control (IPC) of synthetic DNA that is spiked into each sample and non-template control (NTC) to account for possible inhibition of the PCR reaction and minimize false negative results. If the difference in amplification between an eDNA sample+IPC and the NTC+IPC was greater than 0.5 PCR cycles, the sample was considered inhibited. For inhibited samples (approximately 20 percent of samples), we diluted them 1:5 for subsequent analysis to mitigate the inhibition. Finally, a small number of samples from the 2019 survey were incorrectly washed with 30% ethanol instead of 70% ethanol, resulting in lower DNA concentrations than those samples would have otherwise. This affected only about 5 percent of all samples, and was corrected for by using an offset in our statistical analysis (see section *Spatiotemporal modeling of eDNA*). For a

comprehensive description of the eDNA data collection and validation of the workflow, see Ramón-Laca et al.¹⁵ and Shelton et al.⁶.

The number of samples (and samples that were positive for eulachon) varied by depth (Figure 1). To ensure enough data availability for model fitting, we further filtered the eDNA dataset such that there were at least 30 positive eulachon samples for each depth category in both survey years. Additionally, although water samples were collected at 100m in 2019, this sampling depth was removed from the *Shimada's* survey protocol in 2021. As a result, we chose to use only the 0, 50, and 150m depth categories from the eDNA data for both 2019 and 2021. The selected data include water samples from 908 unique location-depth combinations in 2019, and 899 unique location-depth combinations in 2021, producing a total (including biological and technical replicates) of 4175 and 3328 statistical samples for 2019 and 2021, respectively. Of these, we observed PCR amplification of eulachon DNA in 459 samples from 2019 and 367 samples in 2021. Our series of negative controls showed no evidence of contamination (i.e., they did not amplify), confirming that our results are not driven by erroneous sampling or incorrect sample processing.

Environmental covariate data

We collected data on environmental variables that were hypothesized to influence eulachon distribution. Three-dimensional (latitude, longitude, depth) seawater temperature and salinity were extracted from the GLORYS Global Ocean Physics Reanalysis model product from the Copernicus Marine Service¹⁶. Data were downloaded covering the spatial and temporal extent of the eDNA dataset, at a monthly, 1/12° (~9km) resolution. Temperature and salinity values were then matched from the GLORYS grid to the eDNA observations using nearest neighbors for latitude, longitude, and depth. Bottom depth (bathymetry) was also extracted as a covariate, using the *marmap* package in R (version 1.0.10, doi.org/10.32614/CRAN.package.marmap).

Because eulachon are anadromous, we also included data on the annual mean discharge of freshwater from major U.S. west coast rivers. River discharge data from the RC4USCoast dataset were downloaded from the National Centers for Environmental Information¹⁷. The data include a long-term climatology of river outflow for 35 U.S. west coast rivers, with mean monthly discharge ranging between 1 and 8000 m³ s⁻¹ (for the Elder and Columbia Rivers, respectively). We used the coordinates of the river mouths to project these discharge values onto a 5 km-resolution spatial grid, where the value in each grid cell i relative to each river mouth was calculated with exponential distance weighting:

$$R_i = \sum_{k=1}^k F_k e^{-\frac{\delta_{ik}}{10}}$$

(Equation 1)

Where F_k is mean discharge of river k in m³ s⁻¹, and δ_{ik} is the distance between grid cell i and river mouth k . After calculating this integrated river discharge metric for all grid cells, we matched values to the eDNA samples in a similar way as other covariates, using nearest neighbors.

Finally, we collected data on estimated euphausiid distribution and abundance, because euphausiids (krill) are a known preferred prey item of subadult and adult eulachon^{12,19}. Echosounder data on the two most common California Current euphausiid species, *Euphausia pacifica* and *Thysanoessa spinifera*, were collected on the same acoustic-trawl survey as the eDNA samples, and so are well-matched in both space and date to our observations. Euphausiids were measured using calibrated EK80 echosounders operating at 38 and 120 kHz, and reported as nautical-area backscattering coefficient (NASC), as detailed in Phillips et al.¹⁸. For the purposes of this study, the NASC values from each acoustic transect were summarized in different ways to obtain five alternative metrics to test in our models of eulachon distribution. As a sixth possible euphausiid metric, we obtained the smoothed kernel density estimates from Phillips et al.¹⁸, representing a depth-integrated and spatially smoothed index of euphausiid abundance.

Spatiotemporal modeling of eDNA

We developed a state-space statistical model of eulachon DNA in the ocean, separating our representation of the true biological process from the process we use to observe the data. The biological process we wish to represent is the spatial distribution and concentration of eulachon eDNA. Let $D_{t(x,y,d)}$ be the true concentration of eulachon DNA (in copies per liter), in year t at location x, y and depth d . We model the generic biological process as

$$\log_e D_{t(x,y,d)} = \alpha + \mathbf{X}\boldsymbol{\beta} + \omega \quad (\text{Equation 2})$$

where α is a simple intercept, \mathbf{X} is a design matrix of log-transformed environmental covariates with corresponding coefficients denoted by $\boldsymbol{\beta}$, and $\omega \sim MVN(0, \Sigma_\omega)$ is a spatial random field. The general biological process model in Eq. 2 has many possible variations, including choices about how to specify intercepts and spatial fields. These choices were explored in a model selection exercise detailed in the Supplementary Information. Ultimately, a model with separate intercepts by year (2019 and 2021) but a common spatial field across depths and years, was chosen as the most parsimonious model through an out-of-sample cross-validation procedure.

Regardless of the exact form of the biological process model, we observe this process through the lens of our eDNA samples, which are subject to various modifications throughout data collection. We use a multi-level observation model to represent this process.

First, we model the concentration of eulachon DNA in each Niskin bottle (i.e., each 2.5L water sample), E_i , as corresponding to the true concentration at a given spatial location, subject to offsets that account for modifications in the sample processing across eDNA samples:

$$\log_e E_i = \log_e D_{xyd} + O_i + \beta_W W_i \quad (\text{Equation 3})$$

Here, W_i is an indicator variable with estimated coefficient β_W denoting whether a sample was ($W=1$) or was not ($W=0$) affected by the ethanol washing error described above.

O stands for offsets, and is a log-transformed variable without a coefficient

$$O_i = \log_e V_i + \log_e I_i + \log_e(0.05) \quad (\text{Equation 4})$$

V_i is the proportion of each 2.5-liter Niskin that was filtered (almost always, $V_i = 1$). I_i is the dilution that was used on sample i to eliminate PCR inhibition. Finally, we include an expansion factor of 20 to convert our sample estimates to the standard measure of eDNA copies/L (or inversely, a reduction factor of $1/20=0.05$ to convert from true concentration to our observation). We derived this assuming a 2.5L water sample eluted in 100uL Longmire buffer, with 2 μ L used in each PCR reaction. Therefore, each 2 μ L of sample corresponds to 2% (2 μ L of 100 μ L) of the total water sample and is thus equivalent to the copies of DNA contained in 50mL of water. (i.e., 2% of a 2.5L Niskin = 50 ml). To convert to copies per liter, we multiply by 20. In simple terms, the offsets O ensure that we are comparing equivalent volumes across samples.

Because our samples are processed through qPCR, we do not directly observe the concentration of eDNA, but rather the PCR cycle C at which each sample fluoresces above some detection threshold, indicating successful DNA amplification. Samples with no DNA, or with DNA concentrations below the detection threshold, will not fluoresce. The limit of detection of our assay at the 95% threshold is below one copy per μ l as per validation of the assay experiments¹⁵. Conditional on detection, samples with a greater concentration of eulachon DNA should amplify faster and fluoresce at earlier cycles. We represent this qPCR process as a hurdle model indexed by PCR plate j , for replicate r of eDNA sample i ,

$$\begin{aligned} G_{ijr} &\sim \text{Bernoulli}(p) \text{ where } \text{logit}(p) = \phi_{0j} + \phi_{1j} E_{ij} \\ C_{ijr} &\sim \text{Normal}(b_{0j} + b_{1j} \log_e E_{ij}, \sigma_C(E_{ij})) \text{ if } G_{ijr} = 1 \end{aligned} \quad (\text{Equation 5})$$

As with all qPCR analyses, we use samples of known concentration on plate j (K_j), in our case ranging from 1 to 100,000 copies, to calibrate this relationship with standard curves

$$\begin{aligned} G_{ijr} &\sim \text{Bernoulli}(p) \text{ where } \text{logit}(p) = \phi_{0j} + \phi_{1j} K_j \\ C_{ijr} &\sim \text{Normal}(b_{0j} + b_{1j} \log_e K_j, \sigma_C(K_j)) \text{ if } G_{ijr} = 1 \end{aligned} \quad (\text{Equation 6})$$

Note the terms in Eqs. 5 and 6 for the intercept (b_{0j}, ϕ_{0j}) and slope (b_{1j}, ϕ_{1j}), which are shared between the PCR standards and field samples. These shared terms allow us to use the standard curves to estimate DNA concentration in the field samples, E_{ij} . We index b and ϕ by PCR plate to allow for plate-level variation in DNA amplification. These slope and intercept terms, as well as σ_C , are constrained by designating them as PCR-plate level random effects.

Statistics and Reproducibility

Computational burden can hinder the estimation of large, multi-level spatiotemporal models like ours, in part because of the need in spatial models to invert large matrices tracking covariance between data points. An effective solution to this problem is to use the Stochastic Partial Differential Equation (SPDE) model to approximate the latent spatial process and its variance through the use of a triangulated spatial mesh³⁹, which we calculated with the `fmesher` R package⁴⁰([10.32614/CRAN.package.fmesher](https://cran.r-project.org/web/packages/fmesher/index.html)). The SPDE approach drastically reduces computation time by using a sparse covariance matrix (dense precision matrix) to approximate Gaussian Markov Random Fields. Moreover, estimation of this latent spatial process provides interpretable and useful information on spatial (Matérn) covariance in the data—in particular, the magnitude of that variance and the spatial range, which captures the effective spatial decorrelation distance. Projections are then made to the locations of the data using distance-weighting among the vertices of the SPDE mesh. After testing multiple meshes with varying degrees of complexity, we selected a final mesh that maximized the log-likelihood of our fitted model while retaining environmental covariates (Figure S8). The final mesh has 118 vertices, with a minimum distance between vertices of 40km (Figure S9). Additional details of the mesh construction are given in the Supplementary Information.

We tested various model structures and environmental covariates to find the most likely model. In particular, for model structure we tested whether the scalar intercept(s) α and spatial field(s) ω in Eq. 2 should be common across all data, or indexed by sample depth category (i.e., 0, 50, 150m depths) and/or sample year (2019, 2021). For abiotic environmental covariates— \mathbf{X} in Eq. 2—we tested all combinations of bottom depth (bathymetry), temperature, salinity, and our river input metric R_i (Eq. 1). We additionally evaluated the inclusion of the euphausiid data as a predictor, but only allowed one of the six candidate euphausiid metrics to be tested at a time. Environmental predictors were modeled with low-dimensional penalized splines (Eilers and Marx 1996) with a maximum of 3 basis functions (i.e., $k=3$).

Each candidate model was fit to the data by maximum marginal likelihood, utilizing the `sdmTMB` package in R⁴¹ ([10.32614/CRAN.package.sdmTMB](https://cran.r-project.org/web/packages/sdmTMB/index.html)), itself an application of Template Model Builder (Kristensen et al. 2016). The data for fitting included water samples from 1807 unique locations. Biological (i.e., collection of pairs of water samples) and technical replication in the lab produced a total of 7503 statistical samples for model fitting.

We used cross-validation—whereby 10% of the data were predicted using a model fit on the remaining 90%—to compare the likelihood of each alternative model given the data, using the same cross validation folds for testing each model. The final model included a common spatial field with separate intercepts by year (for fits of the chosen model to the data, see Figures S4-S7). Temperature and a depth-integrated, spatially smoothed euphausiid metric were selected as the covariates. After fitting the chosen model, we projected it onto a regular, 5km resolution grid for mapping, as well as for the calculation of an eDNA index. Because grid cells of the projection grid have consistent area, we can define the eDNA index I for any spatial domain \mathbf{A} containing k grid cells as a simple sum of model predictions,

$$I_A = \sum_{k=1}^k D_k$$

(Equation 7)

where D_k is the predicted eulachon eDNA concentration in cell k . We calculated the eDNA index for different \mathbf{A} as one way to summarize our results, including across states and depth categories (Figure 3), and across latitude bins with integration across depths (Figure 4).

A full description of model and variable selection is provided in the Supplementary Information and the Zenodo archive⁴² (<https://doi.org/10.5281/zenodo.14721413>).

Acknowledgements:

The authors gratefully acknowledge the survey expertise and support of NOAA's Pacific Hake survey team and the NOAA Ship Bell M. Shimada personnel for supporting the collection of eDNA samples on the West Coast Hake Survey 2019-2021. In addition, the authors acknowledge L. Park and R. Gustafson for their leadership in the early stages of the development of this study; without their support this work would not have been possible. This work was supported in part by the NOAA Fisheries Omics Strategic Initiative (2019-2023).

Data Availability Statement:

All data and code required to reproduce our findings are archived and publicly available on Zenodo⁴² (<https://doi.org/10.5281/zenodo.14721413>). The source data behind the figures in the paper are available in the Zenodo archive. No genetic sequence data were generated in this study. All genetic analyses were performed via quantitative PCR (qPCR) using established, previously published primer sequences in Ramón-Laca et al.¹⁵ (also in Table 1) to quantify the expression of known targets, resulting in numerical data rather than producing sequence reads. Any remaining information can be obtained from the corresponding author upon reasonable request.

Code Availability Statement:

The R code required to reproduce our findings are archived with the study data on Zenodo⁴² (<https://doi.org/10.5281/zenodo.14721413>).

Author Contributions:

Conceptualization: O.R.L., A.O.S., K.M.N., R.P.K.

Methodology: O.R.L., A.O.S., R.P.K., E.J.W.

Data Analysis: O.R.L., E.J.W.

Data curation: A.R.-L., A.O.S., K.M.N., R.P.K., E.M.P., A.W.

Writing and Editing: O.R.L., A.O.S., A.R.-L., K.M.N., E.J.W., E.M.P., J.E.Z., A.W., R.P.K.

Competing Interests Statement:

The authors declare no competing interests

References:

1. Lomba, A. *et al.* Overcoming the rare species modelling paradox: A novel hierarchical framework applied to an Iberian endemic plant. *Biological Conservation* 143, 2647–2657 (2010).
2. Jeliaskov, A. *et al.* Sampling and modelling rare species: Conceptual guidelines for the neglected majority. *Global Change Biology* 28, 3754–3777 (2022).
3. Green, R. H. & Young, R. C. Sampling to Detect Rare Species. *Ecological Applications* 3, 351–356 (1993).
4. Duarte, S., Simões, L. & Costa, F. O. Current status and topical issues on the use of eDNA-based targeted detection of rare animal species. *Science of The Total Environment* 904, 166675 (2023).
5. Lafferty, K. Metabarcoding is (usually) more cost effective than seining or qPCR for detecting tidewater gobies and other estuarine fishes. *PeerJ* 12, e16847 (2024).
6. Shelton, A. O. *et al.* Environmental DNA provides quantitative estimates of Pacific hake abundance and distribution in the open ocean. *Proc. R. Soc. B.* 289, 20212613 (2022).
7. Jarman, S. N., Berry, O. & Bunce, M. The value of environmental DNA biobanking for long-term biomonitoring. *Nat Ecol Evol* 2, 1192–1193 (2018).
8. Zizka, V. M. A., Koschorreck, J., Khan, C. C. & Astrin, J. J. Long-term archival of environmental samples empowers biodiversity monitoring and ecological research. *Environ Sci Eur* 34, 40 (2022).
9. Keller, A. G., Grason, E. W., McDonald, P. S., Ramón-Laca, A. & Kelly, R. P. Tracking an invasion front with environmental DNA. *Ecological Applications* 32, e2561 (2022).
10. Sildever, S. *et al.* Eight years of weekly eDNA monitoring in the North-Western Pacific. *Environmental DNA* 5, 1202–1215 (2023).
11. Endangered and Threatened Wildlife and Plants: Threatened Status for Southern Distinct Population Segment of Eulachon. *Fed. Reg. 50 CFR Part 223* 13012–13024 (2010).
12. Gustafson, R., Ford, M., Teel, D. & Drake, J. Status Review of Eulachon (*Thaleichthys pacificus*) in Washington, Oregon, and California. NOAA Technical Memorandum NMFS-NWFSC-105 (2010).
13. Gustafson, R. G. *et al.* Conservation status of eulachon in the California Current. *Fish and Fisheries* 13, 121–138 (2012).
14. Ward, E. J. *et al.* Using spatiotemporal species distribution models to identify temporally evolving hotspots of species co-occurrence. *Ecological Applications* 25, 2198–2209 (2015).
15. Ramón-Laca, A., Wells, A. & Park, L. A workflow for the relative quantification of multiple fish species from oceanic water samples using environmental DNA (eDNA) to support large-scale fishery surveys. *PLOS ONE* 16, e0257773 (2021).

16. Lellouche, J.-M. et al. Recent updates to the Copernicus Marine Service global ocean monitoring and forecasting real-time $1/12^\circ$ high-resolution system. *Ocean Science* **14**, 1093–1126 (2018).
17. Gomez, F. A. et al. RC4USCoast: a river chemistry dataset for regional ocean model applications in the US East Coast, Gulf of Mexico, and US West Coast. *Earth Syst. Sci. Data* **15**, 2223–2234 (2023).
18. Phillips, E. M. et al. Spatiotemporal variability of euphausiids in the California Current Ecosystem: insights from a recently developed time series. *ICES Journal of Marine Science* **79**, 1312–1326 (2022).
19. Yang, M.-S., Dodd, K., Hibpshman, R., & Whitehouse, A. Food Habits of Groundfishes in the Gulf of Alaska in 1999 and 2001. U.S. Dep. Commer., NOAA Tech. Memo. NMFS-AFSC-164 (2006).
20. Barraclough, W. E. Contribution to the Marine Life History of the Eulachon *Thaleichthys pacificus*. *J. Fish. Res. Bd. Can.* **21**, 1333–1337 (1964).
21. Hay, D. & McCarter, P. B. Status of the eulachon *Thaleichthys pacificus* in Canada. *Canadian Science Advisory Secretariat Research Document* 2000/145, (2000).
22. Gustafson, R., Richerson, K. E., Somers, K. A., Tuttle, V. J. & McVeigh, J. T. Observed and Estimated Bycatch of Eulachon in the 2002-21 U.S. West Coast Fisheries. [10.25923/3611-EY18](#) (2023).
23. Hickey, B., Geier, S., Kachel, N. & MacFadyen, A. A bi-directional river plume: The Columbia in summer. *Continental Shelf Research* **25**, 1631–1656 (2005).
24. Hickey, B. & Banas, N. Why is the Northern End of the California Current System So Productive? *Oceanography*. **21**, 90–107 (2008).
25. Barth, J. A., Pierce, S. D. & Castelao, R. M. Time-dependent, wind-driven flow over a shallow midshelf submarine bank. *Journal of Geophysical Research: Oceans* **110**, (2005).
26. Ressler, P. H. et al. The spatial distribution of euphausiid aggregations in the Northern California Current during August 2000. *Deep Sea Research Part II: Topical Studies in Oceanography* **52**, 89–108 (2005).
27. Calambokidis, J. et al. 4. Biologically Important Areas for Selected Cetaceans Within U.S. Waters – West Coast Region. *Aquat Mamm* **41**, 39–53 (2015).
28. Derville, S., Barlow, D. R., Hayslip, C. & Torres, L. G. Seasonal, Annual, and Decadal Distribution of Three Rorqual Whale Species Relative to Dynamic Ocean Conditions Off Oregon, USA. *Front. Mar. Sci.* **9**, (2022).
29. Anderson, C. N. K. et al. Why fishing magnifies fluctuations in fish abundance. *Nature* **452**, 835–839 (2008).
30. Emmett, R. L. Ecology of marine predatory and prey fishes off the Columbia River, 1998 and 1999. *NOAA Technical Memorandum NMFS-NWFSC-51*.
31. Litz, M. N. C., Emmett, R. L., Bentley, P. J., Claiborne, A. M. & Barceló, C. Biotic and abiotic factors influencing forage fish and pelagic nekton community in the Columbia River plume (USA) throughout the upwelling season 1999–2009. *ICES Journal of Marine Science* **71**, 5–18 (2014).
32. Bollens, S., Frost, B. & Lin, T. Recruitment, growth, and diel vertical migration of *Euphausia pacifica* in a temperate fjord. *Marine Biology* **114**, 219–228 (1992).

33. Harrison, J. B., Sunday, J. M. & Rogers, S. M. Predicting the fate of eDNA in the environment and implications for studying biodiversity. *Proc. R. Soc. B.* **286**, 20191409 (2019).
34. Hansen, B. K., Bekkevold, D., Clausen, L. W. & Nielsen, E. E. The sceptical optimist: challenges and perspectives for the application of environmental DNA in marine fisheries. *Fish and Fisheries* **19**, 751–768 (2018).
35. Andruszkiewicz, E. A. *et al.* Modeling Environmental DNA Transport in the Coastal Ocean Using Lagrangian Particle Tracking. *Front. Mar. Sci.* **6**, (2019).
36. Sutherland, B. J. G. *et al.* Population structure of eulachon (*Thaleichthys pacificus*) from Northern California to Alaska using single nucleotide polymorphisms from direct amplicon sequencing. *Can. J. Fish. Aquat. Sci.* **78**, 78–89 (2021).
37. Hirayama, I. T., Wu, L. & Minamoto, T. Stability of environmental DNA methylation and its utility in tracing spawning in fish. *Molecular Ecology Resources* **24**, e14011 (2024).
38. Piferrer, F. & Anastasiadi, D. Age estimation in fishes using epigenetic clocks: Applications to fisheries management and conservation biology. *Front. Mar. Sci.* **10**, (2023).
39. Lindgren, F., Rue, H. & Lindström, J. An Explicit Link between Gaussian Fields and Gaussian Markov Random Fields: The Stochastic Partial Differential Equation Approach. *Journal of the Royal Statistical Society Series B: Statistical Methodology* **73**, 423–498 (2011).
40. Lindgren, F. *et al.* inlabru: software for fitting latent Gaussian models with non-linear predictors. Preprint at <https://doi.org/10.48550/arXiv.2407.00791> (2024).
41. Anderson, S. C., Ward, E. J., English, P. A. & Barnett, L. A. K. sdmTMB: an R package for fast, flexible, and user-friendly generalized linear mixed effects models with spatial and spatiotemporal random fields. *bioRxiv* **2022.03.24.485545**, (2022).
42. Liu, O. R., Shelton, A. O., Ramón-Laca, A., Nichols, K. M., Ward, E. J., Phillips, E.M., Zamon, J. E., Wells, A., Kelly, R. P. (2025). Dataset: Mapping the marine distribution of eulachon (*Thaleichthys pacificus*) in the Northeast Pacific using environmental DNA. <https://doi.org/10.5281/zenodo.17676472>

Tables

Gene	Name	Sequence 5´- 3´	Direction	Length (bp)
COI	Tp-COI-F	CCTTATTCGTCTGAGCCGTCCTG	forward	23
COI	Tp-COI-P2	ABY-GGCCGTTCTTCTCCTCCTTTCCCTCCCAGTTT-QSY	forward (probe)	32
COI	Tp-COI-R	GTTAAGATTTCCGGTCTGTTAGAAGCATA	reverse	28

Table 1. Species-specific qPCR primers used in the eulachon assay. A detailed description of the eulachon assay is provided in Ramon-Laca *et al.*¹⁵.

Figure Captions

Figure 1. **eDNA sample location and amplification**

Map of (a) locations and (b) total number of samples collected between July and September of 2019 and 2021, from depths of 0, 50, and 150m, that did or did not amplify (indicating presence or absence, respectively) with the eulachon assay. In (a), all three depth categories are included, which is why some locations have samples that both did and did not amplify in the qPCR assay.

Figure 2. **Predicted eulachon eDNA distribution**

Predicted eulachon eDNA at the surface, 50m, and 150m in Washington and Oregon coastal waters, including (a) estimated concentration in log(copies/L) and (b) all areas with at least a 95% probability of presence of eulachon eDNA. Bathymetric lines indicate the 150m isobath.

Figure 3. **Index of eulachon eDNA**

Index of eulachon eDNA (Equation 7) by year and depth, organized by coastline segment from Washington waters in the north to California in the south. Values represent the sum of gridded model predictions across the study domain for each depth category.

Figure 4. **eDNA and river discharge**

(a) Index of eulachon eDNA by year and latitude. Values represent the sum of model predictions in each 0.125-degree latitude bin, integrated across depth (Equation 7). (b) Climatological mean (1950-2022) monthly discharge from 14 major U.S. West coast rivers¹⁷. Inset map provided for reference.

Figure 5. **Effect of temperature and krill abundance on eDNA concentration.**

Marginal effect (± 1 SD) of temperature (a) and index of krill abundance (b) on estimated eDNA concentration in the selected eulachon model. The krill index is a measure of depth-integrated krill abundance chosen from a suite of candidate predictors (see Methods). Marginal effects are calculated by making predictions of eDNA concentration at different values of the focal predictor, while fixing all other random and fixed effects at their posterior mean values. Parameter uncertainty is characterized by drawing sample parameter values from the model's posterior joint precision matrix ($n=100$ predictions of eDNA concentration for each value of the predictor).

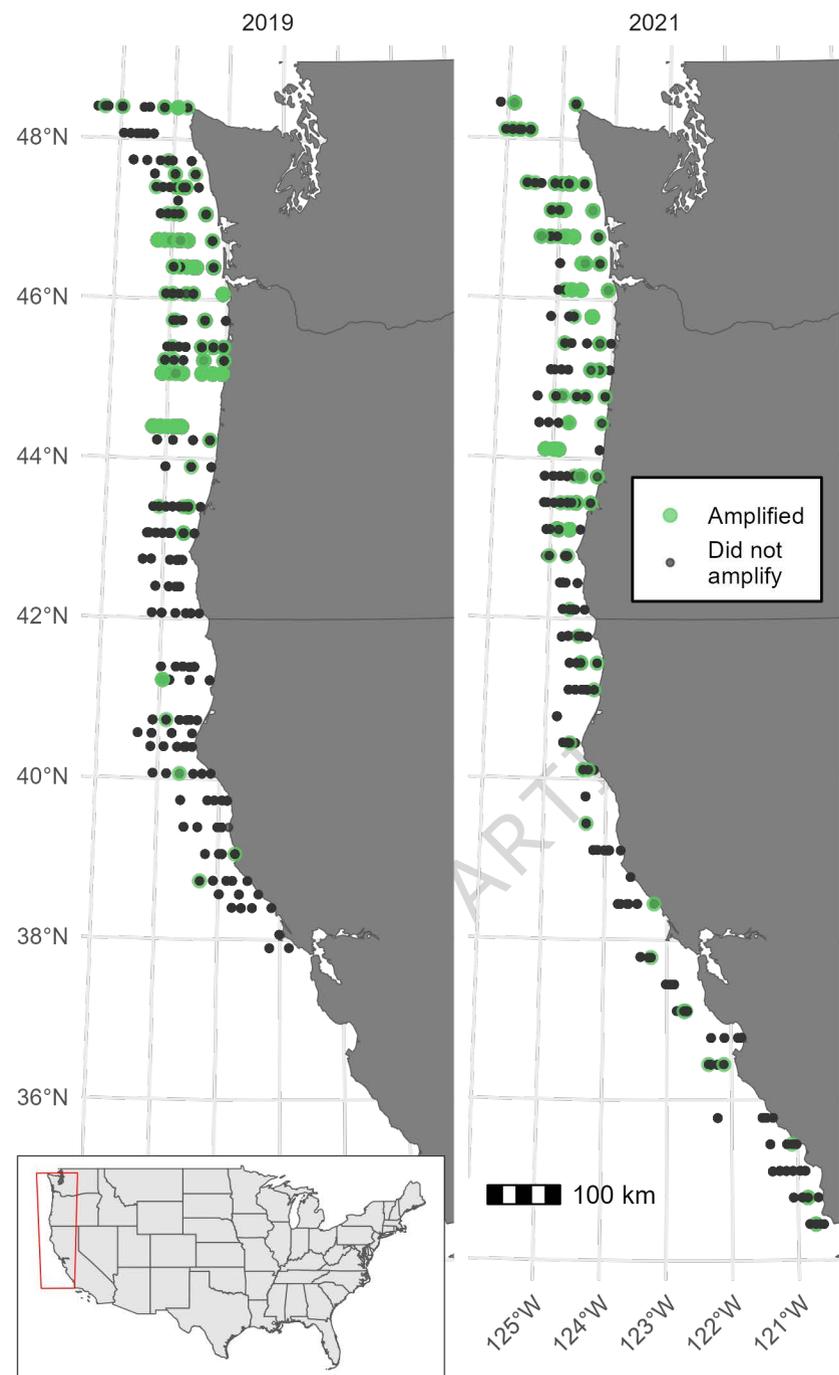
Editor's Summary

This manuscript combines eDNA data with a statistical model to describe the spatial distribution of eulachon, a threatened smelt. DNA was positively associated with river mouths, offshore banks, warmer waters, and areas with high prey concentration.

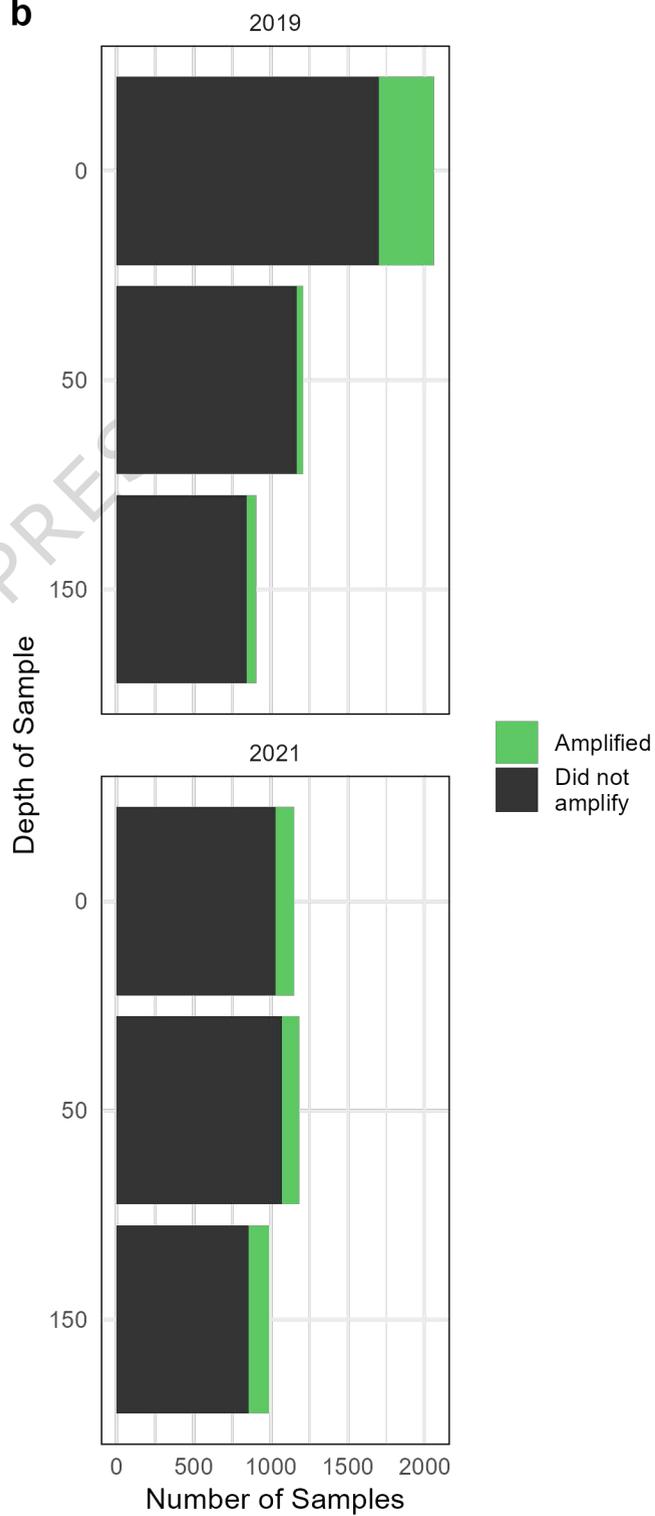
Peer Review Information

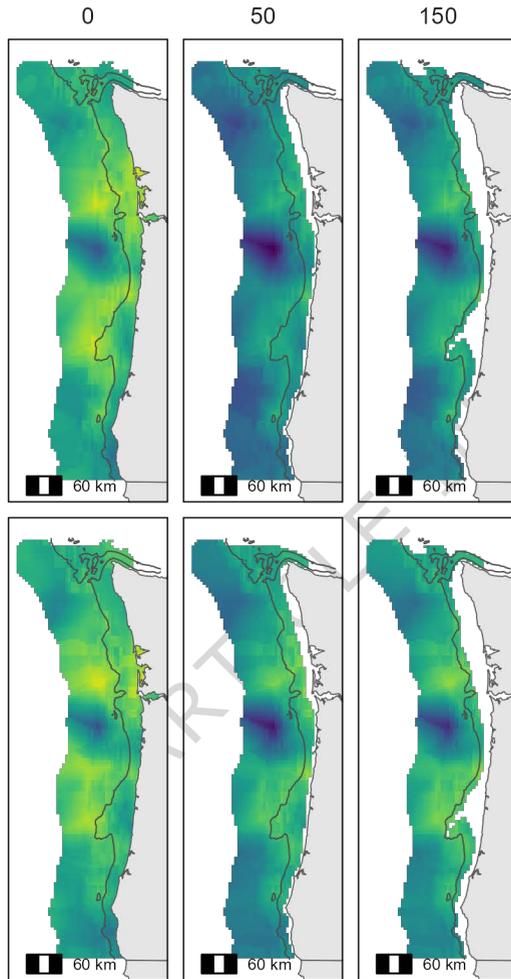
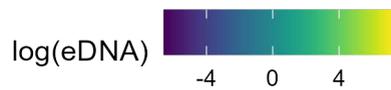
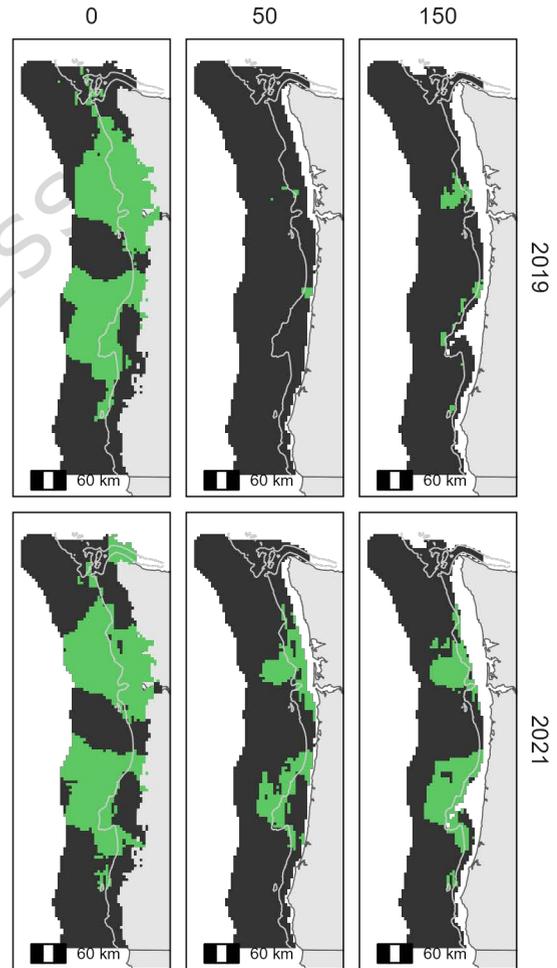
Communications Biology thanks Mads Reinholdt Jensen, Cristina Claver and the other, anonymous, reviewers for their contribution to the peer review of this work. Primary Handling Editors: Linn Hoffmann, Aylin Bircan, and Rupali Sathe. A peer review file is available.

a



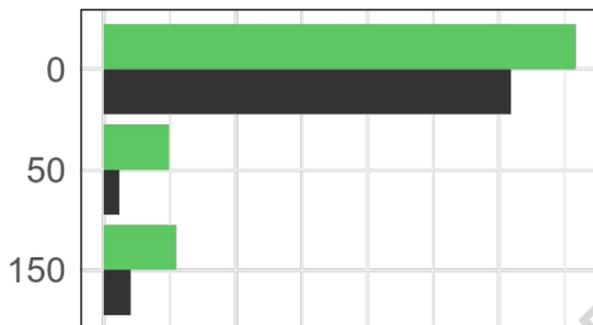
b



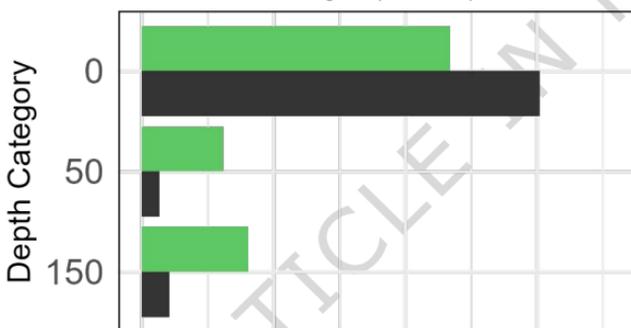
**a****b**

95% p(presence) FALSE TRUE

Washington (>46N)



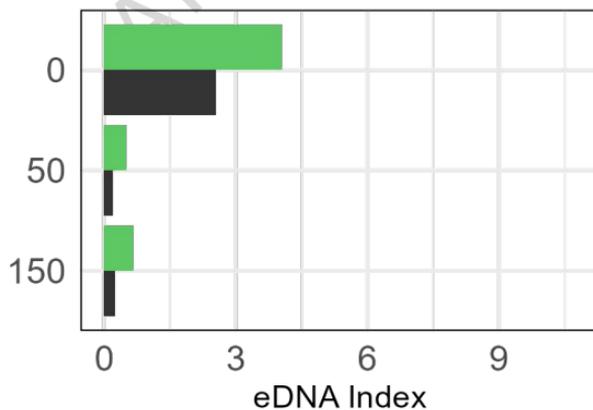
Oregon (42-46N)

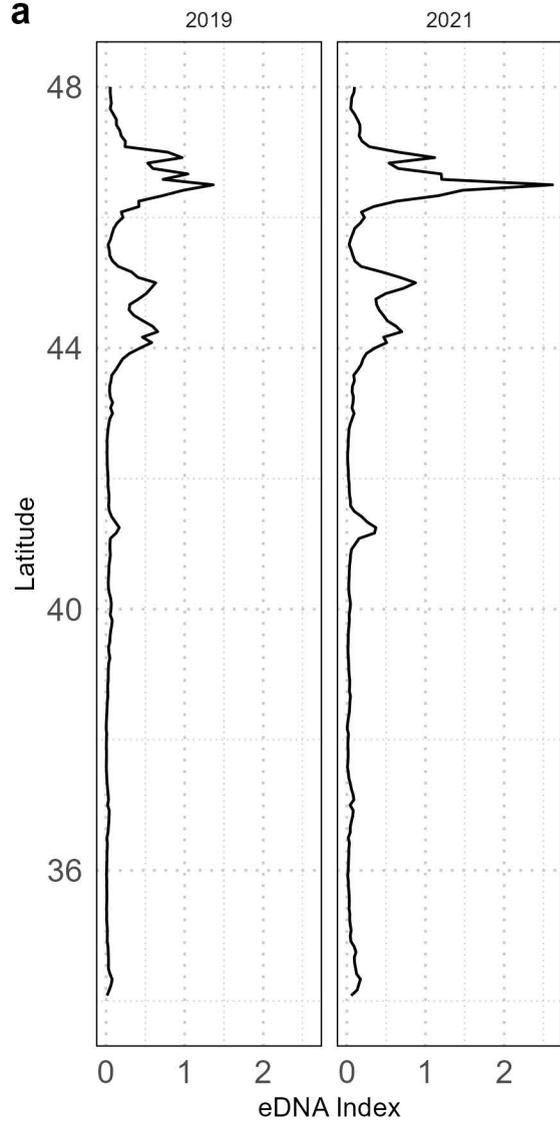


Year



California (<42N)



a**b**