**Response to Reviews**

ASSOCIATE EDITOR COMMENTS:  
  
Associate Editor  
Comments to the Author:  
Dear Dr Zachary Gold,  
  
We have now received two referee reports for your manuscript, both referees highlight the relevance of your work and recommend its publication. One referee provided valuable suggestions and I generally agree with the comments raised by this referee: it is not clear how much better the model combining morphological count data and metabarcoding data is compared to a model using only morphological count data or only metabarcoding data.

**We thank the associate editor for their comments and have made extensive revisions to the methods and discussion sections of the manuscript to better convey both the motivation and performance of our joint modeling efforts as well as highlight the importance of this research in establishing a framework for obtaining quantitative estimates from metabarcoding data (See Methods and Discussion). We realized from the review that our motivation and the significance of the work were not clear in the manuscript and thus have taken great strides to improve the clarity of our reseach.**

**The ultimate motivation for the approach taken here is to address the biased compositional nature of metabarcoding data. Because metabarcoding is compositional in nature it can only provide proportional abundance information at best. However, compounding this issue is that the observed proportions of sequences between taxa are impacted by amplification bias across species. Thus by only utilizing metabarcoding data for our analyses, we would be viewing a skewed observation of the fish larvae detected in each jar. Thus we felt it inappropriate to even conduct such analyses on the raw metabarcoding data alone.**

**We highlight in the methods and supplemental methods at length that metabarcoding data alone cannot solve the three-parameter PCR equation at the heart of our modeling framework because observed amplicon reads are a function of both starting DNA concentrations and amplification efficiencies. As discussed in the revitalized methods section, previous work has demonstrated success by utilizing mock communities to estimate amplification efficiencies. Here we took a different strategy and demonstrated that independent estimates of starting DNA concentration from morphological identifications can be used in a joint modeling framework to estimate amplification efficiencies for all species by constraining the possible starting DNA concentrations (See methods). Such a mechanistic modeling approach utilizes the non-linear relationship of DNA concentration (here assumed to be directly proportional to morphologically derived larval abundance) and amplicon sequences through the PCR equation to then estimate the starting DNA concentrations for all species observed by metabarcoding. This aspect of the model is key as metabarcoding is incredibly sensitive resulting in the detection of an additional 11 species of ichthyoplankton with metabarcoding data that were not identified by morphological methods. In addition, metabarcoding methods more frequently detected 48 species as compared to morphological counts.**

**By jointly modeling metabarcoding and morphological data we were able to derive quantitative abundance estimates for all 59 species while morphological identifications only identified 48 species (See Results). We should mention that we caught a minor mistake in our code that resulted in 3 species not originally being included into our model, this has been corrected and reflected in the manuscript and results throughout.**

**We recognize that the joint model outputs and morphological counts alone provide nearly identical overall ecological stories, particularly since the most interesting patterns in larval abundance shifts were captured by the morphological data. However, the joint modeling framework does ultimately provide greater information on the abundance of changes across a broader suite of fish larvae taxa.**

**Ultimately, the true value of this work is demonstrating the validity of the joint modeling framework. Such an approach can be leveraged by future work either to ease the burden of morphological identifications to only focus on conspicuous and easy-to-identify taxa or to provide an avenue for purely molecular-based quantitative abundance estimates through the joint application of qPCR/ddPCR assays and metabarcoding.**

In addition to the comments of the referee, I would also like to know whether the authors have an idea on how important the assumption on exponential amplification is for the model. The authors have provided a number of arguments which are very reasonable to me, but from the suppl info, roughly 50 PCR cycles were used for the amplification, I doubt the polymerase can work for so many cycles. If the authors could just add a few lines on what would happen in case this assumption would be violated, then this would be very helpful.

**We have included a discussion of the key assumptions of our model, including the assumption that the PCR reaction is exponential across all 49 PCR cycles employed here. From Line 509-523: “Previous work has suggested that PCR reactions are not perfectly exponential but saturate over time as reagents are consumed and enzymatic activity declines in efficiency (Boggy & Woolf, 2010; Chatterjee et al., 2012; Kubista et al., 2006). If the employed PCR reaction were to indeed follow a saturating function, then the number of successful PCR cycles would be effectively lower than the full 49 cycles employed here. Thus, the resulting amplification efficiency coefficients generated by our model would be underestimated. However, given the function of the PCR equation, such a bias would be applied uniformly as a scalar across all estimated amplification efficiencies and thus not affect the broad interpretation of our model results. Fortunately, accounting for any impact of PCR saturation can be easily remedied with the incorporation of a decay coefficient in the function of the underlying PCR equation in future modeling efforts (Boggy & Woolf, 2010; Chatterjee et al., 2012; Kubista et al., 2006). Future work understanding the fundamental mechanisms of PCR amplification and bias, particularly whether PCR follows an exponential or saturating function, is warranted given the importance of metabarcoding approaches to biological observations (Gold et al., 2022; M. R. McLaren et al., 2019; Shelton, Gold, Jensen, D’Agnese, Andruszkiewicz Allan, et al., 2022; Silverman et al., 2021).”**

I would also like to read how the species specific amplification efficiencies for the Mifish primerset have been determined.

**As described above, we clarified how species specific amplification efficiencies were estimated in the Methods section. From line 151-160: “To date, no work has successfully demonstrated the utility of the fourth strategy described in Shelton et al. 2022: employ the use of another independent set of observations of the same community (estimate a subset of ci) to constrain the total possible parameter space of amplification efficiency values. In such an approach, observed DNA concentrations and amplicon sequences from a subset of all observed taxa are used to constrain the parameter space of species-specific amplification efficiencies and thereby allow for the estimation of the starting DNA concentrations for all species observed from metabarcoding data. Thus, this study's objective is to demonstrate the validity of employing an independent set of abundance observations alongside metabarcoding data results to jointly estimate abundance for the entire amplicon observed biological community.”**

With respect to the use of the term eDNA raised by the referee: as I understand material and methods, you only used the ethanol and not the larvae/eggs from the ethanol fixed samples. Please clarify this in the text, so that it is clear that the use of eDNA is indeed correct.

**We thank the reviewer for their comment and have amended the methods section to clarify our use of the term eDNA: Line 103: “DNA metabarcoding was conducted on the ethanol in which 84 paired port side samples were preserved. Given that liquid preservative was used as the target substrate (as opposed to water, soil, or air) and the mass of larvae and eggs in the jar were not directly disturbed, we refer to this process as eDNA metabarcoding herein.”**  
  
We look forward receiving your revised version,  
Best regards,  
Sofie  
  
---  
REVIEWER(S)' COMMENTS:  
  
Reviewer: 1  
  
Comments to the Author  
Review  
  
The manuscript ’Message in a Bottle: Archived DNA Reveals Marine Heatwave-Associated Shifts in Fish Assemblages’ explores the use of DNA extraction from archived ichthyoplankton alcohol preserved samples to reliably monitor biodiversity and specifically monitor potential shifts in fish assemblies in response to climatic warming. This is done by comparing fish assemblies before and after the 2014-16 Pacific marine heatwave in connection to the California Current Large Marine Ecosystem.  
Overall, results are consistent with a shift in biodiversity and abundance where especially North Pacific Hake and Pacific Sardine decline while other more temperate species like Northern Anchovy seem to increase in biomass.  
While the investigation of climate associated shift is interesting, I think that the methods used in the paper are possibly of even broader interest. Here the authors use a Bayesian framework to integrate DNA and morphological data in a joint model to estimate species abundances that seem very interesting and which potentially has large implication for future studies. On the other hand, then it is somewhat unclear to me whether the DNA method also can be used alone without integration of morphological data. This would likely be of general interest given that such approach potential could reduce time/costs of the overall analysis. Moreover, in the same way it is also not clear how well DNA/morphology match, which also is crucial for evaluating the relevance of the method.  
Thus, while I find the study interesting, I have some comments and suggestions for improvement. Some of these relate directly to interpretation of the model, and potential discussion about optimization/validation. In this regard, I think it is fair to mention that I do not have a statistical background in such modeling approach. As such, I cannot review the model in depth. I also have some suggestions for the discussion on the warming effects on the observed ichthyoplankton where I sometimes find that the discussion would benefit from additional data. However, as I am not familiar with the study area, I do not know of relevant complementary studies.  
  
Overall, I find that the manuscript should be of general interest for the readership of Environmental DNA. However, I cannot accept the manuscript in its current form and advice resubmission after revision.

**We thank the reviewer for their comments and have taken extensive efforts to rewrite the manuscript to address their concerns. Please see the modified Methods, Results, and Discussion sections which include comparisons of metabarcoding and morphological results (Figure 2), avenues for applying our model framework to purely molecular based approaches, and a discussion of the strengths and limitations of our current modeling framework.**

Major comments  
  
1)      It is unclear to me how much better resolution one gets for using the integrated approach outlined in this study. As such, I think it would be very nice if the authors could put more emphasize on comparing the morphological and DNA method, with this integrated approach.

**We have included an extended motivation and discussion of modeling approaches in the methods and discussion section highlighting the advantages of the joint modeling approach. As described above, metabarcoding provides additional identification of 11 taxa that went undetected using morphological data, expanding our ecological analyses to include 59 taxa instead of the 48 observed by morphological methods. Furthermore, metabarcoding approaches more consistently detected 47 of the 48 jointly detected species, highlighting the sensitivity of this approach (Figure 2). Ultimately, the adoption of the joint model allows us to estimate abundances for these 11 species which we cannot obtain from metabarcoding data alone. Thus, the joint model provides us higher resolution ichthyoplankton assemblages with quantitative estimates for the full 59 species. We note that having the higher resolution ichthyoplankton assemblage data did not ultimately change the ecological story as the key species (Northern Anchovy, Pacific Sardine, etc.) were jointly detected by both morphological and metabarcoding identifications.**

Similarly, it would also be nice to get some more information about how well DNA/morphological data overlap. Some relevant question could be 1) Are there any taxonomic groups that seem better matched than others? Can anything be inferred about the detection accuracy of low abundant species in the sample for either method? Can anything be inferred about potential strength/limitations of one method over the other? I think a general discussion about this would be beneficial for the manuscript.

**We have included additional analyses to directly compare the overlap and mismatches between the metabarcoding and morphological methods (See Results and Figure 2). We focused these analyses on the 59 taxa that had sufficient representation across either the metabarcoding data set or larval data set to achieve model convergence (observed in >9 technical PCR replicates and >25 total larvae counted across the dataset). We note that these 59 taxa represent 99.0% of all larvae identified and reads sequenced. The results section notes the full description of all taxa identified. Metabarcoding more frequently detected 47 out of the 48 jointly observed species as well as identified an additional 11 taxa not observed by morphological identifications. In contrast, morphological identifications more consistently detected a single taxon and did not identify any additional taxa. We included Figure 2 which plots the co-detection of the 59 taxa analyzed in this study.**

**In addition, we explored the relationship between detection rate and abundance. We found that the probability of non-detections (zero observed sequence reads across technical PCR replicates from the same DNA extraction) was strongly associated with the abundance of larvae in the jar. We only observed non-detections in metabarcoding data when there were less than 10 larvae in the jar (Figure S2). We found a much weaker relationship between morphological non-detections and sequence read abundance which makes sense given our mechanistic framework and the compositional nature of metabarcoding. We include a detailed discussion of these results. Furthermore, we discuss at length the potential sources of these discrepancies, particularly highlighting a suite of drivers of non-detection in morphological identifications including the exclusion of eggs and unidentifiable tissue, stochastic sampling differences between the two cod ends of the bongo net tows, and taxonomic identification limitations of both methods.**

2)      I think that the method section on the modelling approach is a bit difficult to follow and too long. As such, I wonder if it is possible for the authors to reduce this section (e.g. by putting some information in the supporting material) and also try to clearly explain what the model estimates and what data is needed to run their model.

**We followed the reviewer’s advise and moved the majority of the modeling approach section into the supplemental methods. In it’s place we included a detailed description of our motivation and general strategy of our joint modeling approach (Lines 119-173). “To date, no work has successfully demonstrated the utility of the fourth strategy described in Shelton et al. 2022: employ the use of another independent set of observations of the same community (estimate a subset of c\_i) to constrain the total possible parameter space of amplification efficiency values. In such an approach, observed DNA concentrations and amplicon sequences from a subset of all observed taxa are used to constrain the parameter space of species-specific amplification efficiencies and thereby allow for the estimation of the starting DNA concentrations for all species observed from metabarcoding data. Thus, this study's objective is to demonstrate the validity of employing an independent set of abundance observations alongside metabarcoding data results to jointly estimate abundance for the entire amplicon observed biological community.**

**To achieve this, we link the sequencing data to the morphological ichthyoplankton counts from paired samples to constrain the species-specific starting concentrations of DNA in the ethanol jars, allowing for the estimation of abundance for all species observed in metabarcoding data (See supplemental methods).**

The authors state that (line 104-111) ‘We first model taxon sequence-read counts from metabarcoding to account for the PCR process, in which each taxon is subject to a different amplification efficiency based on the primer set used. Furthermore, we link the sequencing data to the morphological ichthyoplankton counts from paired samples to constrain the species-specific starting concentrations of DNA in the ethanol jars. The resulting integrated model leverages the taxonomic breadth and resolution (Gold et al., 2021; Miya et al., 2020) of amplicon sequencing, combining these with the power of morphological counts to yield species-specific quantitative abundance estimates.’

From this text it sounds like the morphological information is needed for estimating the abundances (as I understand it, it is generally used to calibrate the metabarcoding data like in Shelton et al. 2022 who uses mock samples). However, I think that many readers would find it of equal interest to understand if the model can be extrapolated to samples without morphological background data? Did the author do any test of this? Is this possible or do the model depend on morphological data for every sample? I think this would be of high interest, as a pure DNA-based method would reduce time and costs considerably.

**Yes, you are correct in that the morphological information is needed to estimate the amplification efficiencies by constrain the possible parameter space for all starting DNA concentrations. This then allows us to estimate abundances for all observed species from metabarcoding data, including the 11 species that were not observed by morphological identification.**

**We also whole heartedly recognize that many readers (ourselves most of all) are interested in the potential for a purely molecular based estimation of abundance data. Importantly, we included an additional discussion of the strengths and limitations of the model including highlight how our framework here provides a foundation for the development of such a purely molecular based approach. Our model framework employed here strongly suggests that any estimate of absolute abundance can be used to fit a joint model, including molecular based qPCR and ddPCR based approaches. We provide a detailed discussion of future work leveraging such an approach and the additional caveats and limitations future joint modeling efforts must address including subsampling processes (See Gold et al. 2022) and the inclusion of additional estimates of amplification efficiencies from mock communities.**

It looks like the authors have done a test of this (Figure 1) but unfortunately, the text is not clear about what data is actually being used here. It is at least very unclear to me whether the ‘predicted counts’ in Figure 1b are based on the output of the integrated model, and/or whether this data is used in a way so that the ‘predicted counts’ can be used as a truly independent variable? Some more detail in the result section and figure legend would be much appreciated here.

**We have updated the results and figure caption to better clarify our goodness of fit test of the model output. A goodness of fit is a common way of visualizing the accuracy of a model. For example, for a simple linear model a goodness of fit visualization plots the predicted values (typically in the form of a line) on top of the actual observed values. Here we took a similar approach, but instead of plotting non-linear functions for each individual species, we plotted the predicted value (sequence reads or counts) on the y axis and the observed value on the x axis. Here we can see that our non-linear model (given the estimated parameters) is performing well, particularly for abundant taxa. Line 274: “In contrast, model outputs (posterior means, given estimated amplification efficiency parameter values) of counts and sequence reads are much better reflections of observed data visual data as visualized by goodness of fit (Figure 3b, c). Our model predicted larval abundance with high accuracy, particularly for larvae with abundant amplicon and morphological and thus provided reliable quantitative estimates even for the 11 taxa which were only detected by eDNA metabarcoding. Given the performance of the joint model, all subsequent analyses are conducted using the model estimated abundances.”**

3)      An interesting point from the study is that DNA might reveal some insights about potential substructure of certain species. In the manuscript, the authors show an example of this for the Northern Lampfish that are represented by two ASVs (e.g. Line 511-515) that furthermore differ in frequency from before/after the warming period. This is certainly interesting; however, I think some more details is needed to better evaluate the possibility for substructure. Such information should include the exact genetic difference (percentage) between these linages and the observed frequencies over the analyzed period, as well as number of reads observed in the sample. If there were few sequences of both/one of the variants then one would expect an overall large degree of variation across time due to sampling uncertainty. This needs to be clarified. One way to do this would be to make a figure/table to show the actually frequencies and observed read counts for each variant across the sampling period.

**We have included additional information to support our claim that the observed substructure within Northern Lampfish is real. Lines 228-235: “Through these molecular taxonomic assignments, we identified two distinct morphologically indistinguishable lineages (ASVs) of the Northern Lanternfish (Stennobrachius leucopsarus). Although these two clusters of ASVs only differed by a single conserved base pair (99.5% sequence similarity), the two lanternfish lineages were repeatedly detected across samples and in high sequence read counts (variant 1: 6,145,100 total reads, 236 detections across technical replicates; variant 2: 259,989 total reads, 97 detections across technical replicates) and also exhibited dramatically different ecological patterns across the samples (Figures 2 and S5). Therefore, the two lineages were treated separately.”**

**Our results are inconsistent with sequencing errors in which a daughter ASV is only observed when the main ASV is in high proportional abundance (Figure S5). Although the genetic differentiation here is a single nucleotide representing ~0.5% of our ~185 bp length locus, this level of differentiation is consistent with many other bony fish species for this locus including nearly all of the salmonids (e.g. Sockeye and Chum salmon; Gold et al. 2021). Given the high rates of observation of both variants within the data set and the poor correlation in proportional abundances between both sequences, we argue that this variation is capturing real substructure within this species.**

Minor comments  
Line 4. I wonder if the authors can come up with another name than eDNA for their samples. Given that the DNA likely comes directly from the captured fish larvae (and eggs??), this is not what one normally thinks about when referring to eDNA. I would suggest to just call it DNA, but emphasize that the general method is the same as normally use when analyzing eDNA.

**We agree that sampling ethanol from the top of a museum specimen is not typically considered eDNA. We have included additional information justifying our use of the term here: Line 103: “DNA metabarcoding was conducted on the ethanol in which 84 paired port side samples were preserved. Given that liquid preservative was used as the target substrate (as opposed to water, soil, or air) and the mass of larvae and eggs in the jar were not directly disturbed, we refer to this process as eDNA metabarcoding herein.”. We also include an additional discussion of why this unique application of eDNA sampling is likely more amendable to quantitative abundance estimation than more traditional aquatic eDNa.**

Line 25. I think some papers showing climate-induced range shift might fit well in this part of the discussion in case there are any good examples from the studied region.

**We included citations to two manuscripts that highlighted range shifts of subtropical and tropical species within the region associated with the marine heatwave.**

Line 53: Just keep one bracket.

**Corrected.**

Line 58-59. The authors mention that traditional visual identification often is limited in terms of taxonomic resolution. I think it would be great with some examples here to exemplify how big of a problem this is.

**We included an example here (Line 58) : “For example, the larvae of only 3 of the 66 species of rockfishes in the genus Sebastes in the California Current can be identified morphologically, requiring genetic-based identification (Thompson, Chen, et al., 2017).”**

Line 61. Please state the name of the primers used.

**Line 62: “sequencing MiFish Universal Teleost 12S rRNA gene amplicons (Miya et al., 2015)”**

Line 67. The authors need to state the exact number of analyzed samples here divided on station, month and year. Without this information it is very difficult to assess the robustness of the results. It would further be nice to include the sampling map in the main paper.

**We have included a sampling map within the manuscript (Figure 1) as well as a description of the full scope of samples analyzed here (Line 116): “This resulted in a final data set of 84 jars across 4 stations and 23 years that were morphologically identified as well as 84 jars representing 90 unique DNA extractions and 244 unique PCR technical replicates that were sequenced.”**

Line 90. Please state where/how the alcohol samples were preserved. Were they stored frozen, or away from light etc?

**Line 93: “Ethanol samples were archived in the Pelagic Invertebrate Collection at Scripps Institution of Oceanography stored at room temperature and out of direct sunlight.”**

Line 93-95. I cannot see any idea of using the phrase ‘eDNA metabarcoding’. I suggest the authors to just call it metabarcoding, as it is not classical eDNA samples (see my comment above as well).

**We understand that this is a unique application of environmental DNA metabarcoding. We argue here that the ethanol is the environmental substrate in which our target DNA is residing. Although there are clear differences between this unique application of eDNA as opposed to classical eDNA samples from soil, water, and air, we argue that we meet the criteria of eDNA as we are not directly targeting any tissue samples, but the preservative in which the tissue is stored. We included a discussion of this unique sampling design and why this was likely more amenable to quantitative abundance estimates than traditional applications of eDNA as there are far fewer processes affecting the total amount of DNA within the preservative than in aquatic environments.**

Line 95. Please state the volume of ethanol used for DNA extraction.

**Line 107: Up to 125 mL (mean = 121.7 mL, n=6 jars with < 125 mL, min = 34 mL) of ethanol preservative was pipetted off of archived samples and filtered onto 0.2µm PVDF filters.**

Line 98-101. I think some more detail about the morphological identification is needed. Did the analysis include identification of the entire sample, or just a subsample? Did specialists perform identification? And which equipment was used etc.? This needs to be stated here and not in the supporting material. It would also be extremely valuable to get information about what the samples actually contained. Is it only fish larvae, or also eggs in the sample? As I understand it, then only larvae are identified here, so this could potentially lead to a bias between what is observed using two methods, as the DNA method likely would be able also to identify eggs.

**We included further information on the identification of the larval fish. We note that the reference here and supplemental methods provide extensive background on the morphological identification processes. Line 98: “Microscopy was conducted by a team of globally recognized expert ichthyologists to identify species abundance from 84 formaldehyde-preserved samples following standardized CalCOFI techniques in which all larvae are sorted from the other invertebrate zooplankton assemblages within each cod end and then subsequently identified to the lowest possible taxonomic rank (McClatchie, 2016). We note that eggs of a few species are typically sorted from each sample, but were not included in our analyses (See Discussion).” The only tools used are forceps and microscopes, along with incredibly keen eyes and decades of experience conducting these morphological identifications.**

Line 102: I think that the section about the modelling is very long. Although it is important then I believe that it can be reduced in length by moving some of the details to the supporting material.

Line 117-118. Can the authors elaborate a bit on how they actually estimate amplicon efficiency? Is this based on a mock sample, their morphological sample or other analyses? It is not clear to me.  
Line 277-282. As I understand it then the authors integrate in their final model both DNA and count data. This allows for higher species resolution and better quantification. However, I am a bit confused about how the authors use the morphological data. As I understand it, then they use it as a kind of mock, to basically establish a link between a biomass index (here number of individual) to metabarcoding read data. While this seems sensible to me, then I do wonder whether the authors tried other indices than just number. For example, from my own experience I know that surface area often is a better predictor of eDNA copy number compared to number of individuals. This relates to that DNA of dead organisms mainly is excreted/released from slime/scales etc from the surface. This could easily be the same for the alcohol-preserved samples. Here one could try to use the length to estimate surface of fish larvae. Is that something the authors have considered and/or tried during analysis of the data?

**We followed the reviewer’s advice and moved the details of the modeling into the supplemental methods. In its place we included a detailed description of the modeling strategy employed to better clarify how we were estimating amplification efficiencies and deriving quantitative abundance estimates for all species observed from metabarcoding. Our objective here was to demonstrate that the 3 parameter equation (input DNA concentrations, amplification efficiencies, and amplicon reads) can be solved using a joint model with information for both input DNA concentrations (assumed directly proportional to larval abundance) and amplicon sequence data. Importantly, we did not use any mock communities within this study. Instead, we used the morphological data from 48 taxa to constrain all possible parameter values to estimate amplification efficiencies and abundance estimates for all 59 species. Line 151-160:** “**To date, no work has successfully demonstrated the utility of the fourth strategy described in Shelton et al. 2022: employ the use of another independent set of observations of the same community (estimate a subset of ci) to constrain the total possible parameter space of amplification efficiency values. In such an approach, observed DNA concentrations and amplicon sequences from a subset of all observed taxa are used to constrain the parameter space of species-specific amplification efficiencies and thereby allow for the estimation of the starting DNA concentrations for all species observed from metabarcoding data. Thus, this study's objective is to demonstrate the validity of employing an independent set of abundance observations alongside metabarcoding data results to jointly estimate abundance for the entire amplicon observed biological community.**

**To achieve this, we link the sequencing data to the morphological ichthyoplankton counts from paired samples to constrain the species-specific starting concentrations of DNA in the ethanol jars, allowing for the estimation of abundance for all species observed in metabarcoding data (See supplemental methods).”**

**All models are wrong, but some are useful. We include a larger discussion of the limitations of our approach and the validity of our model assumptions in the discussion. We highlight that many of our model assumptions are likely incorrect, including that all larvae have equal shedding rates, that the paired cod ends from either side of the bongo net tow had identical ichthyoplankton abundances, and that PCR follows an exponential curve for 49 cycles. Despite the limitations of our approach outlined in detail in the discussion, we were able to capture the majority of variation explaining the relationship between metabarcoding and morphological results. Critically, we argue that this modeling framework provides a foundation for future mechanistic modeling efforts that account for the various processes including allometry, morphology, subsampling, origin, state, and fate and transport of eDNA. We highlight the future research avenues and questions to pursue to address many of these unresolved fundamental questions.**

Line 303-313. Did the test actually account for the potential very large variation across samples? I just wonder if the significant differences observed here might reflect the large sample variation more than the specific time period. As such, I wonder if the same number of significant different taxa/species would be observed for other similar comparisons of the same data set (for example other time period before/after the heat wave). Did the analysis take this into account?

**Yes, our analysis accounted for the large variation across stations and years by incorporating all model-derived abundances for each species at each site in each year. We calculated 95% confidence intervals based on the observed variation among species-site-year abundance estimates. Line 202-210: “To evaluate the effect of the marine heatwave on CCLME fishes, we compared estimated species abundances before the marine heatwave (1996–2013) to that estimated both during and after the marine heatwave for each station (2014–2019). We first calculated the mean abundance for each species at each station for each model run. We then subtracted the pre-marine heatwave species-site abundance means from the post-marine heatwave species-site abundance means for each model run to evaluate changes in marine heatwave abundance per species per station per model run. We then calculated a 95% CI of change in marine heatwave abundance per species to identify which species were significantly different before vs. during and after the marine heatwave at each station.”**

**We recognize that this is a very simplistic, but statistically sound analysis of species community change. Given our limited spatial replication (4 sites at most each year) we did not have the statistical power to conduct a more robust multivariate analysis exploring the multi-stressor changes in the environment across the 23 year time period. Thus we focused on documenting large scale community assemblage shifts associated with the onset of the marine heat wave. We note that our findings are strongly corroborated by recent work based on the morphological identification of thousands of CalCOFI ichthyoplankton samples for which our analysis is a very small subset (Thompson et al. 2022a). Furthermore, our choice of time periods for analysis is not arbitrary, but based on the dramatic ecological changes that occurred in the Spring of 2014 through 2019 as detailed in a suite of cited literature within this manuscript ( Gentemann et al., 2017; Morgan et al., 2019; Ren & Rudnick, 2021; Schroeder et al., 2019).**

Line 293-295. I think this part should be moved to the discussion.

**We included this description of the 2014-2019 marine heatwave as justification for our analysis comparing fish assemblages before and after the onset of the event. We feel that this paragraph provides important context for readers who are not familiar with this large scale marine heat wave.**

Line 303. Is this analysis based on the joint model output or the morphological count data? I think it would be great to see the results of both methods to understand how much they two approaches differ.

**Yes, you are right these results are based on the model output. We have clarified the results section to make this clearer. We did not include a comparison of the ecological results of each survey method for a few reasons. First, as detailed in our model, we can not obtain reliable estimates of abundance from metabarcoding data alone because it is both compositional in nature and biased by amplification efficiencies. Thus any analysis of metabarcoding data alone would be deeply flawed at best. Second, the morphological data here are a very small subset of all morphologicall identified CalCOFI ichthyoplankton samples. And thus we refer the reviewer to Thompson et al. 2022a which conducts a more robust analysis of changes in ichthyoplankton assemblages on a substantially larger set of all available CalCOFI ichthyoplankton samples. Importantly, we corroborate the results from Thompson et al. 2022a, strongly suggesting that our joint model is capturing real changes in fish larvae assemblages in response to the marine heat wave.**

Line 333-337. I am getting slightly confused here. Please define technical replicates. Do you mean samples from the sample area here? Further, are the 56 species found using both methods and do you have species that are only found via the morphological analysis?

**Here we are referring to technical PCR replicates whereby identical PCRs are generated using template from the same exact DNA extraction. We have clarified this on Line 111: “Each DNA extraction was amplified in triplicate with each technical PCR replicate sequenced separately to capture stochastic variation within the amplififcation process.”**

**We included a greater comparison of the detection of species across both taxa (See Results). We note that none of the 59 species used in our modeling efforts were observed only by morphological analysis. In contrast, 11 species were identified by metabarcoding data alone.**

Line 353. The authors need to include some more details about plot b and c. It is very difficult to understand what they represent otherwise.

**Please see the above response for our clarification. In short, these plots represent a goodness of model fit plotting the predicted model results (given all estimated parameters) to the observed data.**

Line 373. I Don’t think that the references to the support material is correct here. Please check.

**We have corrected this mistake.**

Line 388. I think the title should be changed to abundance instead of biomass since it is not known how well number correlates with biomass

**We agree and have changed the title to abundance.**

Also, please state clearly what data set was used for this analysis. To me it looks like it is only the morphological data that are used here. Why not also the joint data? No matter what then I think it would be very interesting to include both the modelled data and the actual count data in the same figure for transparency.

**These are results from the model. We have clarified in the Results section that figure 3 and 4 are reporting the joint model data. The results of the joint model effort are abundances of larvae per site per year and thus have the same units as the morphological data. We have included supplemental figures derived from the morphological data alone which highlight nearly identical patterns, albeit with 11 fewer identified species, to address your concerns (Figure S6 and S7).**

I also wonder if there are any fisheries data for the studied areas that supports the findings here. Given that the data set still is fairly small it would be nice to understand if the observed patterns are supported by other data.

**We apologize for not making the context of our results clear. We included dozens of citations of recent manuscripts that find similar shifts in abundance changes in response to the marine heatwave. In particular, Thompson et al. 2022a leverages the full suite of thousands of CalCOFI ichthyoplankton samples across the 70 year time series (as opposed to the 84 samples used here) to explore changes in ichthyoplankton assemblages in the California Current in response to the marine heatwave. We have better clarified that our results corroborate those found by Thompson et al. 2022a throughout.**

Line 431-433. Is it possible that the observation is biased in any way? Either by the sampling or uncertainties with the estimates. Given that the observed pattern is unique this might indicate a problem. Are there any other studies that support this finding, or are there studies that contradict this finding? Also, is the finding the same in the morphological count data?

**Again we point the reviewer to Thompson et al. 2022a and the dozens of other citations in this manuscript that have characterized the ecological changes of the fish assemblages to the California Current. Our results are strongly consistent with other recent publications from the region and strongly suggest that we are capturing real changes in fish assemblages in response to the marine heatwave.**

Line 462-469. What about changes in sea current? Is it possible that the current changes during extreme heat waves and therefore advect larvae to places where they do not normally occur? I mean, how likely is it the patterns is as direct ecological response to temperature change, compared to transportation?

**As highlighted in the methods section, a suite of environmental changes including changes in current strength and speed and upwelling occurred during the marine heatwave (Gentemann et al., 2017; Morgan et al., 2019; A. S. Ren & Rudnick, 2021; Schroeder et al., 2019). Disentangling the effects of temperature and prevailing currents is the focus of Ren & Rudnick 2021 and we defer the reviewer to that in-depth manuscript. Here we do not have the spatial resolution to disentangle co-occurring changes in sea surface temperature and source water throughout the long persistence of the event (2014-16 and arguably through 2019 and the present). However, we note that Thompson et al. 2022a discusses this issue at length and highlights the role of advection and persistence of southern mesopelagics into the California Current ecosystem.**

Line 480-501. I think this paragraph can be shortened down a bit as it is slightly of topic and somewhat speculative.

**We have shortened this paragraph accordingly.**  
  
  
Reviewer: 2  
  
Comments to the Author  
The study presents novelties on the effect of the heatwaves on large scale marine ecosystems studying the species assemblages through application of eDNA quantification of fish larval abundances fo different assemblages.  
This study is of great interest in a management and conservations point of view taking into account an innovative and frontiers application of eDNA techniques: the quantification of abundances using a novel joint Bayesian hierarchical model described in Shelton et al. (2022).  
The scope of the work is very clear and well explained, the study design is appropriate for the research approach and the manuscript is well constructed, with clear introduction and justified objectives.  
Materials and sampling methods are clearly described. Results are easy to follow and Tables and Figures are justified.  The discussion is well documented and is really based on the result obtained in the study and the counts previously done for the ichthyoplankton counts, particularly for abundant species.  
My opinion is that the manuscript deserves to be accepted for publication in the journal Environmental DNA.

**We thank the reviewer for their kind response and hope to have the opportunity to publish in the environmental DNA journal.**

---  
HOW TO SUBMIT YOUR REVISION:  
  
1. Please highlight the changes made to your manuscript using 'track changes' or similar, or using bold or colored text.  
  
2. To submit your manuscript, log into <http://mc.manuscriptcentral.com/edn3>  and go to your Author Center.  You will find your manuscript title listed under "Manuscripts with Decisions."  Under "Actions," click on "Create a Revision."  Your manuscript number has been appended to denote a revision.  
  
3. When submitting your revised manuscript, you will be able to respond to the comments made by the reviewer(s) in the box provided.  You can also type in "see cover letter", and include your responses in your cover letter instead. Please also upload a file containing a copy of your responses (with author and co-author names removed); you should upload this file at the File Upload screen and choose the file designation “Additional File for Review but not Publication”. In order to expedite the processing of the revised manuscript, please be as specific as possible in your response to the reviewer(s).  
  
4. IMPORTANT:  Your original files are available to you when you upload your revised manuscript.  Please delete any redundant files before completing the submission.