

We thank all three reviewers of revision round #2 for their time they have taken to evaluate our manuscript (revision #1). Referee #3 and #4 provided further detailed reviews with very helpful and constructive comments which we greatly appreciated. In our revision #2, we considered all of them very carefully and are convinced that they helped to further improve the quality of our manuscript.

The major changes we performed in revision #2 in response to consistent comments of both, referee #3 and #4, are that we turned the objectives that we had outlined for our work into clearly formulated working hypotheses that were subsequently tested with repeated measures ANOVA (rmANOVA, presented in the new Table 1). Furthermore, the samples taken in the Pacific were apparently misunderstood as (treatment) control, which they are not. However, we realized that we missed to make this apparent also for readers not so familiar with such type of mesocosm studies. To prevent further misperception, we included a sentence (last) into the introduction explaining that accompanying samples were taken in the Pacific to gain insight into the in situ plankton community development. Table 1 (rmANOVA results) aids further transparency to the fact, that the Pacific samples were not used as statistic controls. Please note, that the accompanying monitoring of the plankton community in the water directly adjacent to the mesocosm field is the accepted approach for these type of mesocosm experiments (Bach et al. 2020, and many more from previous studies). For further detail, please see below our point-by-point response (in blue) to each of the referees' comments.

## Author's point-by-point response to comments of referee #3 and #4 of revision round #2

### bg-2022-157-referee-report-2, referee #3

**Author, general comment:** Referee #3 made all comments, changes and corrections into the pdf version of our revised manuscript #1. In order to be able to reply in a point-by-point manner, we have collected the comments and added the line numbers where they were made in the pdf. Text corrections made by referee #3 were performed by the authors directly in the revised manuscript version #2 and can be followed in the marked-up manuscript version #2.

**L2:** I am unclear what this means. Sounds like what is meant is that the system will be negatively impacted by climate change. Is that a "hotspot"?

**Response, Revision #2:** The term "hotspot" has no positive or negative implications per se, but is used here to characterizes EBUs as regions of particular climate change impact. We used this term in accordance with Bach et al. 2020.

**L5:** The previous two sentences do not really provide context for the reader, unless they are familiar with these systems.

**Response, Revision #2:** We agree and do understand the referees' objection. However, due to the word limit for the abstract we have to confine the content to the most important information and have presume some familiarity with the subject.

### L37

**Response, Revision #2:** "uniquely" was replaced by "pronounced" in revision #2.

### L38

**Response, Revision #2:** We didn't include "- between 10 m in revision #2 as this is uncommon in oceanographic literature.

**L39:** As a reader, I am confused here. Is low N:P a characteristic of upwelled water here in HCS specifically, or in general? The "in consequence" leads me to think this is the condition of HCS, but

the rest of the sentence seems to suggest this is a general phenomenon. I suggest re-writing this sentence to simplify its meaning for the reader.

[Response, Revision #2](#): Indeed, the low N:P is characteristic for the HCS. In the revised version, we have rephrased this sentence to make clear that this is a specific characteristic of the HCS.

**L42**: I do not think you mean the upwelled water changes in N:P, but the upwelled water is a different N:P than the surface water it is replenishing?

[Response, Revision #2](#): Thanks for making us aware of this inaccurate formulation. We have rephrased this sentence accordingly.

**L46**: Is this different than HCS? or are you now narrowing your focus to a specific region within the HCS?

[Response, Revision #2](#): Yes, this is narrowing to the specific region because our experiment took place on the shelf.

**L50**: As a reader, I think I am missing a connection. What is the connection between small zooplankton feeding on blooms on shelf waters, and larger zooplankton communities offshore that fuel Peru anchovies. How are the two zooplankton communities connected to each other, resulting in high anchovy biomass? this might be well known to marine researchers, but maybe need to connect the dots for non-marine people.

[Response, Revision #2](#): We have rephrased this sentence to make a better connection between shelf and offshore zooplankton importance for anchovy.

**L73**:

[Response, Revision #2](#): “larval” relates to zooplankton as mentioned at the beginning of this sentence.

**L76**: This statement (top-down control of phytoplankton by zooplankton) confuses me a little because the previous sentence states zooplankton is regulated by food quality and composition (bottom up control of zooplankton by phytoplankton).

[Response, Revision #2](#): Please apologize, but we do not quite understand this objection. This is a general statement that doesn't state top-down control of phytoplankton by zooplankton but states that when primary productivity is low (for example because nutrients are low) euryphagous and carnivorous species take advantage.

**L81–89**: This seems a bit redundant with the end of the first sentence of the paragraph (to investigate the impacts of...). Perhaps this text can be combined with the end of the first sentence to make a single statement to reduce the redundancy?

[Response, Revision #2](#): We have removed this sentence in the context of the extensive revisions performed to the introduction, for example the inclusion of working hypotheses, etc.

The authors have set up their background nicely, and have included several hypotheses as to why this system is so productive. As a reader, I was expecting to see which hypotheses they were going to test with their experiments. However, the objectives are just to describe the response of plankton - not what I was expecting as a reader.

[Response, Revision #2](#): We appreciate the referees' overall assessment of the introduction of our manuscript. The decision to present our work in a descriptive nature resulted from the fact, that the difference of N:P signatures of the two collected deep-waters was much smaller than anticipated (high temporal variability), and thus, larger treatment differences of the mesocosm plankton communities could not be expected. This was unfortunate but for many mostly logistic reasons (time, effort, budget, availability of ship time and deep-water masses with larger N:P differences, etc.) it was impossible to sample two new deep-water batches. Therefore, in accordance with Bach et al. 2020, and like Schulz et al 2021 and Chen et al. 2022 (this SI) we had decided to present our data in a

descriptive way and provide important baseline information of plankton dynamics in the HCS under the impact of upwelling events over time that future studies can build on. Yet, we do understand the reviewers' request to see an hypotheses-driven approach.

To me, it seems like these experiments are/could address Hypothesis 1 (zooplankton are extremely efficient at using primary production) but need to identify and test the mechanisms that are at work under different OMZ conditions that would enhance or inhibit zooplankton conversion efficiency. To me, having objectives that are "to describe" is limiting the potential of this work. Can the objectives be turned into statements related to scientific understanding based on interpretation of data analysis to test hypotheses/mechanisms? Which of the experimental conditions do the authors expect will lead to the great trophic transfer efficiency and why? What are the response variables that can be compared to those conditions to lead the authors to infer one mechanism over a competing mechanism? I think the objectives could be improved by going beyond just "describing". [Response, Revision #2](#): As explained above, established differences in inorganic N:P signatures through the addition of OMZ water were only minor, which we had mentioned probably too late, only in the Discussion, L524–526. In revision #2, we have moved this important information to the introduction to provide the frame of the actual experiment conditions beforehand. As explained just above, formulation of hypotheses and subsequent hypotheses testing seemed to us a bit meaningless as no actual treatment differences could be established. That's why we decided to follow the approach of Bach et al.. However, as we do understand the referees' request, in revision #2, we have formulated two hypotheses in line with the objectives that we had formulated and, for the datasets that allowed for, performed ANOVA statistics for hypothesis testing.

My sense is that the authors probably have expectations/predictions. More detail here on what the response variables are, and how they expect them to respond to different treatments would help the reader. For example, "If mechanism X, we would expect to see increased GF in treatment 1 and decreased GF in treatment 2..." This will help the reader better understand the framing of the work and how we will learn something about pattern and process in the system.

[Response, Revision #2](#): As indicated just above, in revision #2 we have rephrased this whole paragraph, formulated two hypotheses and elaborated on our expectations and response variables (performed analyses).

**L91**: The authors have setup an experiment with two levels of OMZ treatment. However, they do not really describe why these two levels were chosen, what questions/hypotheses the experimental design is addressing, what comparisons of which response variables are to be done to evaluate their expectations (hypotheses). From the Introduction, it seems to me there are a number of mechanisms hypothesized to be at work to make this system so productive. I think the manuscript could be greatly strengthened using a hypothesis-driven approach rather than an approach that just describes many variables without any context.

[Response, Revision #2](#): As mentioned above, we have revised the introduction and formulated hypotheses to address the reviewers' request. With respect to the OMZ treatments, they cannot be (exactly) foreseen as the N:P stoichiometry of upwelled water masses as well as upwelling events as such are highly variable on temporal, local, vertical scales, i.e. we had to take what was there at time of collection of OMZ water. Moreover, OMZ water was collected at permanent time-series stations of IMARPE, i.e. stations that are regularly visited and are within reasonable distance (with respect to logistic, etc. considerations). We have described this in L99. More details are given in the overview paper by Bach et al. (2020) and are beyond the scope of our manuscript. We acknowledge, this might be rather specific knowledge for non-marine readers, but to stay within the scope of our work, some things have to be presupposed, thus, we kindly ask for your understanding.

**L116**: As a reader, going into this section on zooplankton sampling, I am unclear on why the experimental treatments were used in relation to this work. The authors must have some expectations based on hypotheses - what is going to be measured and for what purpose?

[Response, Revision #2](#): Hypotheses have been included, please see responses to previous comments.

**L130**: Maybe I missed it, but how were microzooplankton sampled from the water samples? what mesh size range for example? what is considered "microzooplankton"?

[Response, Revision #2](#): In this section zooplankton sampling in general is described (microzooplankton was collected with integrated water samplers), further processing of microzooplankton samples is described in detail in section 2.8. In revision #2, to avoid redundancy we included reference to section 2.8 to make the reader aware that more detailed information follows further below.

**L134**: after what or when?

[Response, Revision #2](#): This refers to Day 20 at the beginning of this sentence. To make it more obvious, in revision #2, we have replaced "afterwards" with "after Day 20."

**L172**: spell out to start sentence

[Response, Revision #2](#): We really appreciate the detailed and critical reading of our manuscript by referee #3, but in this case we think this is rather a matter of taste. Also using an abbreviation for one of the most important parameters at start of a sentence complies with the style in the overview paper by Bach et al.. Therefore, we refrained from changing.

**L187**: Long sentence. Suggest splitting into two.

[Response, Revision #2](#): As suggested, the sentence was split into two.

**L198**: Unclear.

[Response, Revision #2](#): „such“ was replaced with „GF“ to clarify the meaning.

**L203**: Do you mean three individuals were selected for each sample?

[Response, Revision #2](#): A triplicate sample is a commonly used term meaning three samples with in our case one female per mesocom. To clarify, we have added this information in brackets.

**L204**: This should be paired with the last sentence to keep the C:N work together, and the DM work together. As is, the DM work is split by this sentence about C:N sentence.

[Response, Revision #2](#): In revision #2, we have performed the changes as requested.

**L206**: specify whether one individual went into one tin cup, or whether individuals group and then total biomass was divided by # of individuals

[Response, Revision #2](#): Individual female copepods were picked and transferred (see L203). To clarify we have replaced "individual" with "single".

**L217+219**: Spell out to start sentence?

[Response, Revision #2](#): Kindly, we decided to use the abbreviation also at the start of the sentence (in line with Bach et al.)

**L227**: is this period supposed to be here?

[Response, Revision #2](#): Yes, it is an abbreviation.

**L258**: I am confused by the wording here. If the organisms are being retained on the 15-um mesh, how does once concentrate to a volume of 2-4 ml? What are the remaining water samples?

[Response, Revision #2](#): The 200 µm pre-filtered water samples were further concentrated via inverse filtration to a volume of 2–4 ml. These 2–4 ml "remaining samples" were then further processed as detailed in the text. We have rephrased this passage for better clarity.

**L279:** It is not clear to me what statistics were used in many parts of the Results, or if any were with some statements in the Results. I think this section could be improved by outlining which comparisons are made and why (hypotheses), and which statistics were used for each comparison.

Response, Revision #2: Following others comments of referee #3 and #4, we performed comprehensive additional statistics (repeated measures ANOVA) on our data set and revised section 2.9 accordingly with more detailed information on the response variables that were tested for possible treatment effects due to the addition of deep water with moderate or extreme OMZ signature. Moreover, we explain better why for the fatty acids and gut fluorescence “only” means with corresponding CIs/standard deviations are shown and why rmANOVA could not be applied.

**L282:** I am confused here. How can one have a "mean" of an individual? Clarify.

Response, Revision #2: We agree, our phrasing was not clear here. In revision #2 we have rephrased this sentence to increase clarity: *“Instead, we calculated mean percentages and their 95% confidence interval (CI) of the dominant fatty acids per mesocosm treatment and experimental phase to allow for basic statistical inference.”*

**L283:** I may have missed it, but significance of what? What is being inferred, what is being compared here?

Response, Revision #2: Because the fatty acid dataset is restricted due to unequal sample size and low number of replicates (because about half of the samples had to be omitted from the data set due to impurities, described in section 2.9), comprehensive statistical analysis (rmANOVA) was not applied. Instead, we used means of fatty acids (per phase and mesocosm treatment) and confidence intervals to provide insight on potential significance between any two means. Confidence intervals provide information on the significance between any two means based on their overlapping or non-overlapping CIs (Field et al. 2012), so that from comparing any two means and their corresponding CIs in Table 3, inference on potential significance can be drawn. However, we realize that the respective text passage was not clear and comprehensive enough, and thus we have revised this passage in section 2.9.

**L283:** What is the purpose? What hypothesis is being tested? What is the expectation? Which treatment is expected to have higher abundance, and will it be higher than ambient conditions (Pacific)?

Response, Revision #2: As suggested, in revision #2 we have formulated two hypotheses (introduction) to frame our work better. According to these, we additionally performed rmANOVA to test for significant differences between the response variables (see new Table 1) and the two deep-water treatments. This should also help to raise awareness that the Pacific sample was not used as control but only to provide a background on the natural development of the plankton community in the adjacent waters (consistent with Bach et al., Schulz et al., Chen et al, etc.)

**L284:** Unclear to me why GF is included in this manuscript if no meaningful results could be taken from the data/analyses? I have not looked at results yet, but my first thought is to remove all of the GF work if the data are not sufficient to make any meaningful interpretation....After having read the Results, it seems like GF was very low suggesting low feeding. I would change this wording b/c low feeding is information....

Response, Revision #2: It is important here to realize the difference between gut fluorescence (the amount of fluorescence in the gut) and gut clearance rates (from which feeding rates can be estimated). GF was generally very low, indicating low feeding on autotrophic food, and thus, a valuable result of the study. At the same time, because GF was so low, no clearance rates could be inferred, and therefore clearance rate measurements are not shown. We have rephrased this passage for clarification, however the difference between GF and clearance rate needs to be realized by the reader.

**L288:** Did the authors conduct this analyses using time lags to account for delays between interactions and the realizations of those interactions? Zooplankton will not respond instantaneously to consuming prey, so comparing zooplankton at time  $t$  with phyto at time  $t-1$  would seem appropriate to explore.

[Response, Revision #2:](#) Unfortunately, it is not possible to do correlations with time lags of  $t-1$  between phyto- and zooplankton because these samples are taken at different frequency to limit density changes in the mesocosms (due to zooplankton being sampled with nets whereas phytoplankton is sampled with water bottles) to a minimum. This is common practice for mesocosm experiments. In the case of our study, zooplankton was sampled every 6 days whereas phytoplankton was sampled every 2 days.

**L291:** I think the manuscript could be strengthened if the authors presented some hypotheses to be tested. "gain insights into food web relations" seems very generic and somewhat meaningless, especially when it seems like a lot is known or hypothesized about how the HCS works.

[Response, Revision #2:](#) As requested, we have included some hypotheses in the introduction to better frame our work (see our replies to previous comments, please).

**L293:** suggest being more specific here. what is meant by "elemental composition"?

[Response, Revision #2:](#) In response to the reviewer comments, we performed rmANOVA on our dataset (where possible) and accordingly extensively revised Section 2.9. The single response variables are listed in the new Table 1 together with statistic results ( $p$ ,  $F$  values).

**L299:** is this redundant? "mesozooplankton" is a group of all the mesoplankton, implying "total"

[Response, Revision #2:](#) Yes, we agree, it seems redundant, however, this wording can be typically found in marine zooplankton literature.

**L299:** I believe this should be "density" throughout the manuscript (# per unit area or volume). Abundance is how many (total numbers in a system, population, community). Should be changed throughout the manuscript

[Response, Revision #2:](#) The term abundance is defined as "the number of individuals per unit volume of water" (Zooplankton Methodology Manual, Harris et al. 2000, p147) and is typically used in zooplankton literature.

**L303:**

[Response, Revision #2:](#) Kindly, we did not correct "numbers" as this is typical wording in marine zooplankton literature.

**L304:** not really needed (to me)

[Response, Revision #2:](#) We do explicitly acknowledge the different perception of necessary detail. Nevertheless, we prefer to keep this information, because it gives max. abundance values that cannot directly be read from Fig. 3, and from our own experience it can be very helpful if specific values can simply be read from a paper, for example for comparison reasons.

**L316:** This is a lot of detailed information in this paragraph - does it contribute to significant inferences?

[Response, Revision #2:](#) We do understand the referees' objection, however, both referees of revision round #1 agreed with the detail, respectively asked for additional information with respect to the chordata (please see our point-by-point-response to revision round #1 and the revision #1 of our manuscript). As we have to adequately consider the comments of all reviews we received, we would leave this paragraph as it is.

**L324:** This section comes across as a "play-by-play" of patterns of each taxon. I'm not really sure how important this level of detail will be for making inferences. Maybe it will be clear in the Discussion,

but my first impression is this seems just to be a list of information without comparing experimental treatments.

[Response, Revision #2](#): We do understand the reviewers' objection with respect to a too detailed list of information, also this comment is in line with a comment made by referee #4 of revision round #2. Yet, we would like to kindly point out, that the information listed always compared the experimental treatments. In revision #2, we have substantially shortened this section and restricted the description to the main results depicted in Fig. 5.

**L361**: Same comment as previous section. Seems like a lot of play-by-play of detailed patterns, but not really sure how this fits into any context of OMZ effects from the experimental treatment. Comes across as a "list" of facts with no context.

[Response, Revision #2](#): Like the previous section and comment, in revision #2 we have substantially reduced this part of the results comparing stage succession between mesocosm treatments, which is again well in line with a comment made by referee #4.

**L385**: The design has a treatment with two levels (moderate and extreme OMZ) and a control. Why not report the means and test for differences? As a reader, I find it very difficult to discern any patterns with reporting of these specific details - given the context of the experimental design, I think providing results that compare the treatment levels would be much more informative. Perhaps a supplemental table showing all the raw values for each mesocosm would be a better way to present this level of detail. As a reader, I want to see comparisons - were zooplankton more abundant in moderate or extreme OMZ, and how does this relate to the hypothesized mechanisms of how the system works...

[Response, Revision #2](#): As suggested by referee #3, we have performed rmANOVA on our data and present the results in Table 1. No OMZ treatment effects were determined. We have added this information in the respective paragraph with reference to the (new) Table 1. The raw values are published in the PANGAEA database, and therefore, were not additionally included in the supplement.

**L391**: A design that include 4 stations distributed around the mesocosms would have provided the opportunity for a balanced comparison of control and two treatment levels (each with N=4). This is hindsight of course...

[Response, Revision #2](#): As pointed out above already, this is a fundamental misperception. The samples taken in the Pacific were not introduced as control and do not serve as treatment controls. This is the general accepted approach for KOSMOS mesocosm experiments to give insight on the plankton development in the adjacent waters. As control such samples even with larger n would be unsuitable because the mesocosms are closed systems. Please compare also with other publications in this SI (Bach et al, Schulz et al., Chin et al.)

**Fig.7**: I do not see these

[Response, Revision #2](#): Referee #2 of revision round #1 and the Editor requested inclusion of vertical lines in all figures to indicate times of deep-water additions. We followed this suggestion and removed vertical lines indicating phases because figures were overloaded with two types of vertical lines. Unfortunately, we forgot to remove mention of "phase vertical lines" in the caption of Fig. 7, but we have done it in revision #2.

**L415**: Suggest removing the entire GF part of the paper if it did not produce any useful results (significant or non-significant).

[Response, Revision #2](#): As mentioned above, it is of absolute importance here to notice the difference between gut fluorescence as such and gut clearance rates while reading this section (explained in M&M). The result "very low GF" is definitely meaningful and tells that not much autotrophic food was incorporated. This is further supported by the fact that clearance rates could not be measured because GF (over time of defecation) was even lower than initially. The HPLC

attempt supports the extremely low GF even further. Therefore, we will keep the GF part in the manuscript but have rephrased this text passage for better clarity.

**L420:** I'm unclear on the value of the C:N work when only data from a single night were available. Perhaps the usefulness will become clear in the Discussion, but at the moment, I'm just not sure what can be learned from one observation over a 50 day experiment....

Response, Revision #2: We agree, the C:N values depict "only" single time-point measurements, but they summarize the shorter term feeding history of the copepods and give background information. In our opinion, not every measured value presented in results must be taken up in the discussion, but it can still provide valuable information for future comparison.

**L420:** presumably no difference between these treatment level means?

Response, Revision #2: We have added confidence intervals for the treatment means in Table 2 to provide evidence on statistical significance. CIs between the two treatment means overlap, thus do not suggest for significant difference (please note, that the sample size for any further statistics is too small). Accordingly, we have rephrased and shortened the respective text passage.

**L437:** provide statistical test results between phases and between treatments. detail the methods in the Methods

Response, Revision #2: As mentioned in response to a previous comment, the fatty acid dataset is restricted due to unequal sample size and low number of replicates, therefore comprehensive statistical analysis (rmANOVA) could not be applied. Instead, we used fatty acid means (per phase and mesocosm treatment) and confidence intervals to provide insight on potential significance between any two means (see above).

**L438:** Cannot make a comparative statement of TFA levels for female *Hemicyclops* sp in the Pacific with an  $n = 1$ . The single observation may have been lower than any observation from the mesocosms, but that is all that can be said. the CI of 8.0 for stn 3 treatment suggests the 4.7% from the Pacific was likely in the range of observations from stn3 treatment (Table 2)

Response, Revision #2: We absolutely agree, and we didn't mean to say more than the single observation was lower. But we realized that our wording was awkward and misleading. Therefore, we have rephrased the last part of this paragraph and corrected this imprecision.

**Table 2:** what % CI?

Response, Revision #2: This is a misunderstanding of the information in the table heading: It is not % CI but mean contributions of fatty acids (in %) and the corresponding CIs of these means.

**L450:** Report stats. I am unclear on what stats tests were used. See my question on "elemental composition" in the Methods section for data analysis.

Response, Revision #2: As mentioned above, we have extensively revised section 2.9 and explained better that and why comparisons of the fatty acid compositions are based solely on phase-treatment means with their CIs. Accordingly, we have also rephrased this text passage and emphasized in the first paragraph of section 3.4.2 that CIs of any two means that are compared do not suggest for significance because they always overlapped (except in one case, that is mentioned specifically later in the text).

**L456:** Comparing a single value is tenuous at best.

Response, Revision #2: We absolutely agree. But as detailed above, the samples taken in the Pacific are no statistic controls but are to provide some background information on the situation in the adjacent waters. We realize to have missed to make this clear enough for the reader but have pointed it out in revision #2 already in the introduction (last sentence).

**Fig. 9:** adding in vertical lines for the phases would be helpful.



[Response, Revision #2](#): In response to reviewer #2 from revision round #1 we included the green vertical lines indicating deep water additions instead of vertical lines for the three phases. Adding both line types overloaded the figure. As a compromise, we added the phase durations in the figure caption (compare with our response to a similar comment above).

**L519**: I'm not sure how much this first paragraph contributes to the Discussion. Much of this information is background, and could be incorporated into Introduction to help set up expectations for how zooplankton communities would respond to the treatments, given what was found for the phytoplankton and biogeochemistry reported in Bach et al. (2020). As presented, this paragraph is confusing to me.

[Response, Revision #2](#): We agree with the referees' objection and have incorporated this paragraph into the introduction and used it to build hypothesis on in revision #2.

**L535**: This seems to be Results

[Response, Revision #2](#): We have deleted the respective text passage from the discussion and moved the first two sentences and also the following sentence ("Thus, any potential small-scale ...counting bias.") to the start of the results section 3.1.

**L540**: This is Results

[Response, Revision #2](#): We have deleted the respective text passage and included it into the results sections 3.2 (first two sentences of the yellow-marked part) and 3.1 (last sentence).

**L550**: Seems like Results and background.

[Response, Revision #2](#): We have deleted the first sentence of the yellow-marked text ("All copepod species were pooled... and adults.") which is really more a repetition of results. The remaining part of this text passage provides the necessary context/background for the comparison with what is known for Peruvian coastal and neritic waters, which we think should not be moved elsewhere.

**L556**: This paragraph seems to be the real start of the Discussion, where the authors begin to synthesize their results.

[Response, Revision #2](#): We mostly agree and as explained above, in revision #2 we have deleted/moved most text of the preceding two paragraphs of revision #1 so that in revision #2 only a short "introductory" paragraph precedes this respective paragraph.

**L610**: This is a different message than in the Results, where it was stated the data were unreliable. Please clarify. Were the data indicating low feeding, or the data were unreliable. If the former, then keeping the GF data in the manuscript is good. If the latter, then should be removed (if authors cannot make any inference either way).

[Response, Revision #2](#): We agree, the conveyed message was contradictory. Our data are reliable and indicate very low feeding on autotrophic food particles. Accordingly, we have clarified this, both in the discussion and the results.

**L646**: I'm not quite clear what this means?

[Response, Revision #2](#): This sentence (and also the sentence in L567/568) was inserted in response to a comment of reviewer #2 from revision round #1 who pointed out that reduced feeding of zooplankton might have been partly due to captivity (wall effects). We have deleted this sentence again, as it rather seems to confuse. Furthermore, "wall effects" in the relatively large mesocosm bags are probably negligible.

**L654**: what is turnover time of delta C in these copepods? this assumes that the overlap was of sufficient duration prior to day 36 that the C signal in the copepods could change accordingly.

[Response, Revision #2](#): The turnover time of delta 13C varies from days to weeks. Hence, an increase of Chlorophyceae should be seen in the C signal of the copepods after several days. We have included this information in the respective text passage.

**L670**: (include some references. Hobson et al. 1993; Gannes et al. 1997)

[Response, Revision #2](#): We have included the requested references.

**L672**: This seems like excellent material on which to build hypotheses for the experiment, and use a hypothesis-driven approach to this work rather than purely descriptive approach. This information should be in the Introduction.

[Response, Revision #2](#): This information was already given in the introduction (both in the initial version and revision #1, L39–45). In revision #2, we have removed it from this text passage.

**L682**: My take here is that the conclusions basically highlight everything the study did not answer. This detracts from the efforts and what was learned - I think the authors are shooting themselves in the foot as the reader will walk away thinking about everything they did not answer, rather than what they did answer. I think the Conclusions could be strengthened by focusing on what was accomplished, not what wasn't.

[Response, Revision #2](#): We agree and appreciate the referee's evaluation very much. In revision #2, we have revised our conclusion section and strengthened it by emphasizing what was accomplished in our study.

**L749**: this link did not work for me.

[Response, Revision #2](#): It does work for me without any difficulty. So, it's probably a user specific problem rather than generic.

### bg-2022-157-referee-report-3, Anonymous referee #4

**Overall thoughts:** Overall, I found the manuscript to be clear and well-thought out in terms of their methodologies. One concern was the amount of information provided in the Results section which was not adequately elaborated on in the Discussion. There were a lot of figures and tables which I don't think are all necessary in contributing to the story the researchers are trying to tell. I also thought that the Introduction could be improved through talking more about expected results and specific hypotheses. Finally, I have a concern about the Pacific "control" used in this experiment, in which replicate samples were not taken, and the authors didn't discuss potential behavioral differences of the zooplankton in the mesocosm enclosures vs. at the Pacific "control" site.

Response, Revision #2: We thank referee #4 of revision round #2 very much for the very constructive and helpful comments on our manuscript which we have considered carefully. As mentioned above already, a general misperception is that we used the samples taken in the nearby Pacific as a control treatment, which is not the case but the general and accepted approach in our mesocosm studies. However, we realized that we missed to make this fact clear from the beginning, so that also readers less familiar with these type of studies are made aware early. Therefore, in our revision #2 we bring this point to the readers' attention already in the introduction. Furthermore, we formulated the objectives of our work into specific hypotheses to better frame the context of our study and reduced the detail of information in the results section. Please see below, where we reply to each of the referees' comments point-by-point.

#### **Abstract and Introduction, general comments:**

- In the Abstract, they presented a lot of results and not much interpretation of those results. I would have liked to see more interpretation / key takeaways from their results. Their concluding sentence in the Abstract is about how further studies are needed to answer their research questions, but what are their main conclusion(s) that they are able to draw from their results?

Response, Revision #2: We agree with the referee's criticism and accordingly have restructured the abstract. We deleted the last sentence and instead inserted our two main conclusions that we draw from our study at the end of the abstract: "*Our findings suggest omnivorous/opportunistic feeding of copepods may have played an important role in the pelagic food web. Overall, projected changes in frequency and intensity of upwelling hypoxic waters may make a huge difference for copepod reproduction and may be further enhanced by varying N:P ratios of upwelled OMZ water masses.*" Moreover, we deleted some unnecessary detail: "*(pooled and species-specific for Paracalanus sp., Hemicyclops sp.)*" and "*Stable isotope signatures of copepods varied over time but not between treatments.*"

- Overall, I thought the Introduction made it clear what the researchers were seeking to study.

Response, Revision #2: We appreciate that the general layout of our introduction was acceptable for referee #4.

- However, I would have like to see some hypotheses laid out about what they expected to happen over the course of the experiment. A lot of their results were presented in terms of changes from Day 0 to Day 48 of the experiment, but I don't think the Introduction did enough to set up hypothesis on what they thought would happen across this time frame.

[Response, Revision #2](#): Referee #3 made very similar comments with respect to the lack of hypotheses in the introduction. Accordingly, in revision #2, we have comprehensively revised the last paragraph of the introduction with formulated hypotheses and expectations.

- Similarly, I would have liked to see more concrete predictions about anticipated differences in results for the moderate OMZ vs. extreme OMZ vs. Pacific “control.”

[Response, Revision #2](#): Both, referee #3 and #4 criticize the lack of adequate working hypotheses of anticipated differences of zooplankton response between the deep-water treatments. Accordingly, we have formulated two main working hypotheses that we have included in the last paragraph of the introduction. We believe, they give now an adequate study-frame that will help the reader to easier grasp the main intention of our study.

#### **Abstract and Introduction, specific comments:**

- Line 5: Remove the word “how” or “on”

[Response, Revision #2](#): “On” was removed in accordance with reviewer #3 of revision round #2.

- Line 17: The wording of the sentence starting with “Possibly” makes it seem like this is their guess about what the major diet of copepods was - did results show this, or is this just a guess?

[Response, Revision #2](#): This is a guess, and to indicate this, we started the sentence with “possibly”.

- Lines 61-62: Move the citation for Garcia-Reyes et al. 2015 to the end of the sentence

[Response, Revision #2](#): In accordance with reviewer #3 of revision round 2, the citation was moved to the end of the sentence.

- Line 80: I think they should set up in the Introduction what the “two different types of OMZ waters” are. They just mention that they do two types with “different OMZ signatures,” but I think they should elaborate on this and say explicitly that they are simulating upwelling with differing inorganic N:P ratios (extreme and moderate OMZ signature).

[Response, Revision #2](#): In accordance with comments made by referee #3, we have included information on the N difference of the two collected deep waters (some of the information was formerly mentioned in the first paragraph of the discussion) in the last part of the introduction.

#### **Methods, general comments:**

- I would have liked to see more information provided about the Pacific “control” site where they also collected samples from - they should give more details on the location of this site and possibly other characteristics, such as its depth, distance from the mesocosms, etc. Its location should be added to the map in Figure 1. It seems like at this Pacific “control” site, no replicate samples were taken (indicated by no error bars for the data from Pacific), which I am hesitant about. Can this really be considered a “control,” since it wasn’t under the same conditions as the mesocosms (enclosing the zooplankton) - would they expect to see behavioral differences in zooplankton that are enclosed vs. not enclosed?

Response, Revision #2: Obviously, this is a fundamental misperception of both, reviewer #4 and #3 from revision round #2 as these samples were not meant to be controls nor were they anywhere in the manuscript introduced as controls. As reviewer # 4 (round 2) correctly states, such samples are unsuitable as controls because they are open ocean samples, whereas the mesocosm samples are from a closed system. They are to give an idea on the natural plankton dynamics in the adjacent open ocean and are regularly taken during all such mesocosm studies (please compare also with Bach et al, Schulz et al., Chin et al., etc. in this SI where this approach is accepted like in all other preceding KOSMOMS publications). To emphasize, we have included a sentence at the end of the introduction: *“Concomitantly, we monitored the MeZP in the adjacent Pacific shelf waters at one sampling point nearby the mesocosm field to gain insight into the in situ zooplankton community development.”*

- Were zooplankton communities evaluated in the water that was added and removed from the mesocosms? I think this would be helpful to know when considering community changes over time after water addition and removal.

Response, Revision #2: Please compare with our point-by-point-response to a very similar question by reviewer #2 of revision round #1. Unfortunately, no net samples were taken from the collected deep water and the zooplankton removed/added was not quantified (mostly for logistical/manpower reasons during deep-water exchange that is a physical highly demanding, exhausting and time-consuming operation).

#### **Methods, specific comments:**

- Lines 111-115: Were these trends that determined the 3 “phases” consistent across all eight mesocosms and the Pacific “control” site? I would have liked elaboration on whether all of these showed the same trends which determined these “phase” definitions.

Response: The phase definitions are described in detail in the mesocosm experiments’ overview paper in Bach et al. (2020) and are based on phytoplankton development in the mesocosms (of course, they were not and could not be applied to the samples from the Pacific). Thus, they are not the focus of our study. In our manuscript, we give this information to provide some general background of the mesocosm experiment to our data to help interpretation. With respect to the slower mesozooplankton development (than that of phyto-/protozooplankton), the phases cannot be transferred. To clarify, we have inserted *“based on the phytoplankton development”* in the respective sentence.

- Lines 123-124: They sampled zooplankton in the afternoon only, so I would have liked to see them discuss somewhere whether DVM of the zooplankton could have influenced their results - perhaps a lot of the zooplankton were near the bottom of the mesocosms during that time that they sampled.

Response, Revision #2: Zooplankton net samples were taken at maximum sampling depth (17 m) in the mesocosms right above the entrance of the mesocosm sediment trap (total length of the mesocosm bag 19 m, length of the sediment trap is 2 m, see Fig. 1).

- Lines 133-135: I would have liked to see an explanation for why they changed the depths they sampled microzooplankton from Day 20 onwards. Why did they stop sampling the 17-10 m depth interval?

[Response, Revision #2](#): Preparation and analysis of microzooplankton elemental composition is time- and cost-intensive. In the 10–17 m depth range, concentrations of microzooplankton were often too low to reach detection limits, thus, after Day 20 only the upper depth interval, where concentrations were sufficient was continued to be sampled. In revision #2, we have included this information for clarification.

- Figure 2: Legend should say what % CIs are shown. (I assume 95%, since that's what the rest of the figures use, but should be stated explicitly.)

[Response, Revision #2](#): We agree and have added the explanation to the caption of Figure 2.

- Line 168 and Figure 2: Why were only female *Paracalanus* sp. used for this analysis? Why not males and females? I didn't think this was explained well enough. Same comment for Line 210 about fatty acid analysis.

[Response, Revision #2](#): Adult females dominated (L168), and male copepods often do not feed at all. As picking enough organisms to reach detection limit for GF analysis is very time consuming, it was not possible to also pick other stages. Thus, we (had to) chose the numerically dominant stage to assure enough organisms per GF sample in a reasonable amount of time. The same explanation applies with respect to the fatty acid analysis. In our revised manuscript, we added "*the dominant developmental stage*" to both sentences slightly to enhance clarity.

- Line 187: This is a pretty high variance in sample size (between 8 and 52 individuals) - did they find any correlation between sample size used and their results on gut clearance rate?

[Response, Revision #2](#): True, but the high variance in number of organisms per (pooled) sample is due to high variability in the availability of female copepods. According to the standard protocol, both, gut clearance rate and gut fluorescence were normalized to dry mass (L202, Table 1) for comparability.

- Line 198: Remove period after *Paracalanus*

[Response, Revision #2](#): Period was removed.

- Line 214: What was the minimum number of individuals? They state 80 as the maximum, but what minimum did they use?

[Response, Revision #2](#): Minimal availability was 16. We have included this information in the respective sentence.

- Line 295: Add a citation for the R Core Team

[Response, Revision #2](#): Thanks for making us aware of this missing citation. We have included the reference in the revision #2.

### **Results, general comments:**

- This study has a lot of results and a lot of figures. I feel like overall, the text could be reduced. I noted a few chunks of text that I felt contained too much listing of data without any elaboration: Lines 330-337, Lines 369-383, among others. If these results aren't explicitly discussed further in the Discussion, I don't think it's necessary to list them if they are already contained in a figure.

[Response, Revision #2](#): This comment agrees with comments made by referee #3 of revision round #2. In revision #2 we substantially shortened these text passages comparing results of the copepod community between the mesocosm treatments.

- Another general comment on the Results is that I didn't see many statistical test results (p-values, F-values, etc.) presented in the Results section to back up their claims. For example, statements like those on Line 309 could be backed up with a relevant statistical test result.

[Response, Revision #2](#): We have taken the reviewer comment to heart (that agrees with comments made by referee #3) and performed repeated measures ANOVA on most of our data sets (where it was possible, summarized in the new Table 1). The outcome of this is, that the short-term higher abundance in the moderate treatment mesocosms after deep-water addition was not a significant difference. In revision #2, we have added this information in the text of section 3.1, first paragraph.

- Significant issue with Figure 9 that needs to be fixed: Either the color scheme or the legend needs to be fixed on this figure - on all other figures, red signifies extreme OMZ mesocosms (M1, M4, M5, M8) and blue signifies moderate OMZ mesocosms (M2, M3, M6, M7). On Figure 9, the legend is different and either the points are colored wrong or the legend is wrong, since it's implying that M1-M4 are one treatment and M5-M8 are another treatment which is not the case.

[Response, Revision #2](#): Thank you for spotting this error in the legend, which occurred during revision round #1. However, in response to other comments made by this referee, we now present mean and CI values and doublechecked our script and data carefully and the correct treatment IDs were applied to the respective mesocosms.

#### **Results, specific comments:**

- Figure 4b: In my opinion, the color used for Echinodermata and the color used for Others were difficult to tell apart on the figure, which made it difficult to interpret.

[Response, Revision #2](#): We revised Fig. 4b and chose a slightly brighter color for Echinodermata.

- Figure 4b: I think they could add a label for the "moderate" column and the "extreme column" above each column, and then just have "M2", "M1", etc. in each panel, to make this easier to interpret.

[Response, Revision #2](#): This is probably more a matter of taste, but we have performed the requested changes in revision #2.

- Line 322: Why were fish eggs added to the mesocosms on Day 31? This should be explained further and discussed in terms of whether it could impact any of their results.

[Response, Revision #2](#): We mentioned in L323 that fish eggs did not stay long in the mesocosm and further details are given in Bach et al. (2020). The fish introduction was part of another research question within this mesocosm experiment that was not further elaborated because the fish didn't survive, thus, their impact on the zooplankton community was neglectable.

- Lines 342 and 345: In my opinion, the phrases “of some importance” and “of minor importance” indicate some level of influence of those communities. If the authors meant to indicate just abundance or biomass contributions, they should say that instead of stating things in terms of “importance.”

[Response, Revision #2](#): These phrases have been deleted as the section was substantially shortened in response to the previous comment and comments made by referee #3.

- Figure 5 caption: “genera” instead of “genus” in the second to last sentence

[Response, Revision #2](#): Thanks for spotting, we have corrected it in revision #2.

- Figure 6: The different colors here aren’t really necessary since the panels are broken up by genus. If they wanted to, they could just have a black (adults) and grey (copepodids) line for each panel.

[Response, Revision #2](#): In principle, the reviewer is correct, colors are not really necessary. However, we think, colors increase clarity here. It’s probably again a bit a matter of taste. In this case, we would like to keep Fig. 6 as it is.

- Line 390: Confused by what they mean by “M2 and M3 reached biomass maxima for both copepods and polychaetes” - what maxima?

[Response, Revision #2](#): We agree, this sentence was a bit unclear. In revision #2 we have rephrased this statement to read: *“Highest biomass values were observed for both total copepods and polychaetes in mesocosms M2 and M3 (both in the “moderate OMZ signature” treatment). The main contributors to copepod biomass were Paracalanus sp. and Hemicyclops sp.. Due to the high variability between mesocosms, however, this treatment effect was insignificant (Table 1).”*

- Figure 7: The caption says there are vertical lines on the figure indicating the three phases, but these lines are not on the figure.

[Response, Revision #2](#): Apologize. In response to an earlier review of revision round #1, the vertical lines were removed, but we forgot to delete the respective description from the figure caption, which we have done in revision #2.

- Figure 7: Why not do averages and confidence intervals like previous figures? Why have a different line for each mesocosm? Do we really need to see variation between single mesocosms, or would it convey the same information if the mesocosms in each treatment were averaged like in Figures 3, 4, and 5, with error bars?

[Response, Revision #2](#): For consistency with the manual count data, we now present means and CIs in all figures.

- Figure 8: Same comment as Figure 7 - can the replicates in each treatment be averaged?

[Response, Revision #2](#): For consistency with the manual count data, we now present means and CIs in all figures.

Line 423-424: Add citation(s) for the sentence “This is a common problem in small copepods with rather low total lipid mass close to the detection limit” to back this up.



[Response, Revision #2](#): We included Lischka and Hagen (2007) to back this statement up.

- Line 425-426: What proportion of samples hit this purity cutoff?

[Response, Revision #2](#): About half of the samples hit the purity cutoff (44 out of 86). However, at this point we would like to point out that our fatty acids samples are unique for the considered copepod species not only in this region but also elsewhere, thus, they give valuable first insights on the fatty acid composition and short-term dynamics of *Paracalanus* and *Hemicyclops* in the important Humboldt Current upwelling system.

- Table 2: Maybe this could go in the supplemental information

[Response, Revision #2](#): Table 2 presents some of our main results, namely mean fatty acid compositions with their CIs, that we prefer to be included in the main manuscript. Please note that this table also includes our statistical results as the overlapping/non-overlapping nature of the CIs is indicative of significance or not, respectively. Alternatively, we could have showed these data graphically, but that would have resulted in too many single figures.

- Section 3.4.2. Fatty acid and fatty alcohol composition: Lots of information in this section that isn't set up adequately in the Introduction or elaborated on adequately in the Discussion

[Response, Revision #2](#): In this case, unfortunately, we cannot quite share the reviewers' opinion on the respective manuscript passages. As outlined in the results section, fatty acids/alcohols did not play a major role for the respective copepods, showed not treatment differences, but supported our findings on GF. Our discussion is focused on the important biomarker fatty acids (basically lacking), the lack of storage lipids, etc. From our perspective, more detailed elaboration is not necessary but would inflate the results. With respect to the introduction, and in response to different comments of reviewers #4 and #3 of revision round #2, in revision #2 we formulated working hypothesis in the introduction to frame our work better, including mention of the use of fatty acid compositions (well known as biomarkers) to conclude on utilization of possible phytoplankton blooms.

- Lines 483-494: Were these Pearson correlations calculated with abundance or biomass values?

[Response, Revision #2](#): Thanks for making us aware of this missing information. We used copepod abundance and have inserted this information in the respective text passage in section 2.9.

- Figure 9: Since these results are discussed in terms of the 3 phases, I think the 3 phases should be indicated on this figure with vertical lines.

[Response, Revision #2](#): As a trial, we generated the figure with both line types (green for deep-water addition, gray for the three phases), however that overloaded the figure. The green lines had been added in response to reviewer comments of revision round #1 and are in accordance with figures in Bach et al. Therefore, we decided to restrict Fig. 9 to the green lines. However, as a compromise, we have inserted the information of the three phases in the figure caption.

#### **Discussion, general comments:**

- To me, the Discussion felt slightly disjointed. I think it could be improved by strengthening the first and last sentences of each paragraph, to connect the flow of ideas and provide the reader an indication of what each paragraph will be about from just reading the first sentence of each paragraph.

[Response, Revision #2](#): In revision #2, we have rephrased the discussion paragraphs to strengthen the central theme and indicate already through the first sentence of each paragraph our main thought.

#### **Discussion, specific comments:**

- Line 524: Elaborate on “lower than expected” - what values for N concentrations were initially expected, and why?

[Response, Revision #2](#): Information in this paragraph was moved to the introduction in response to comments of referee #3 and #4 of revision round 2. Further, in revision #2, we included details on the established difference between OMZ treatments with respect to dissolved nitrogen concentrations.

- Table 3: Pretty low sample sizes - I would have liked to see more discussion of this and justification for these results.

[Response, Revision #2](#): Organisms for the samples in Table 3 were picked from the second net that was collected for picking live organisms for fatty acid, DM and elemental analyses (see section 2.2). First, a sufficient number (if available) of organisms for fatty acid analyses was picked, afterwards of the remaining samples organisms for DM and elemental analyses were picked. Hence, the number of replicates resulted from the biomass that was still available after organisms for lipid analyses were removed. Because of the restricted number of net samples that could be taken from the mesocosms, a larger sample size was not possible. Yet, we believe these data add valuable “side-information” to justify inclusion in Table 3, but we wouldn’t like to extend the discussion much further not to “overrate” them.

- Lines 604-632: Very long paragraph - try to cut down or break up into multiple paragraphs

[Response, Revision #2](#): We agree and have split this paragraph into three shorter paragraphs.

- Line 686: The information in the parentheses here seems disjointed and confusing

[Response, Revision #2](#): In response to a comment made by referee #3, we have performed extensive revision to the conclusion and in this context we have deleted the information in parenthesis.

- Line 690-691: I would take out the “but implementation....more time and manpower” part of this sentence. In my opinion, ending with this statement diminishes the results that they found in this study. I would rather have them end the Conclusions section with a final take-home message illustrated by their results. But this is just my opinion - up to the authors.

[Response, Revision #2](#): In response to comments of referee #3, we have revised the conclusion for revision #2. In this context, the statement “but implementation...” was deleted.