Review of Chambers et al. "Ocean acidification effects in the early life-stages of summer flounder, *Paralichthys dentatus*"

This laboratory study represents an important new piece of empirical data on the sensitivity of fish early life stages to elevated CO_2 and reduced pH conditions – a topic that is of increasing relevance due to predicted anthropogenic changes in marine chemistry, also known as ocean acidification. The language is clear, and I agree with the premise of the authors that the field is still in its infancy and in dire need of additional observations under controlled laboratory conditions, which are particularly needed for economically important species such as summer flounder. I also agree that the few existing studies cover a wide range of different fish species; however, there is a good empirical basis for assuming that the most sensitive ontogenetic stages are the early life stages. Hence the focus of this study on embryos and pre-metamorphosis larvae appears well justified.

That said, I'd like to offer a few important criticisms and suggestions, which will strengthen the paper, if incorporated. As detailed below, the introduction is a bit misleading in its breadth, because the actual presented data are on a much smaller subset of the whole experiment that has been presumably conducted (e.g., no temperature-dependence). Second, I point out a number of places where critical information needs to be added in order to allow subsequent researchers to properly place these findings into context. Lastly, I'm urging the authors to reconsider some of their stronger statements in the discussion, and better evaluate the many little short-comings of their approach.

Major comments:

Justification of experimental CO_2 levels:

The authors use very high CO_2 levels. While the intermediate level of almost 1,900 ppm represents the predicted CO2 concentrations in the open ocean about 300 years from now (Caldeira & Wickett 2003, Riebesell et al. 2010), the high CO_2 level of 4,700 ppm is not a plausible scenario under any current open ocean acidification scenario. For the latter, the authors admit as much, stating explicitly that their goal was to evaluate the broad spectrum first, before focusing on other more realistic parts of the response curve. I have no problem with that. However, with the intermediate level, the argument is getting logically flawed. Because, on one hand, the reader is told that coastal areas are generally more acidic than the open ocean, thus implying that early life flounder would experience higher CO_2 levels, but then we are told that summer flounder larvae develop in the relatively stable pelagial, where there are no high-frequency temporal changes in CO_2 levels. If that's the case, then the 'open ocean acidification' scenario should apply, which means that the intermediate level is something that is not expected to occur for at least 300 years. This is in conflict with statements from the authors such as (P13918,L15) "… CO_2-induced variations, even those at intermediate, next-century CO_2 values used here, …" or

(P13901, L20) "the twice ambient CO_2 target reflects IPCC predictions … for later this century…" (are the authors perhaps referring to regional coastal levels here? If so, that's not clear)

In my opinion, the authors need to take at least a descriptive stab at the current range of pH conditions that contemporary fish realistically see during the early life stages. They should put less emphasis on responses that were seen under the unrealistically high CO_2 levels (for open ocean) of 4,700ppm.

Replication and genetic diversity

Another major limitation of the study is the small number of females and males used to produce offspring. I am fully aware of the logistical challenges of lab spawning fish with pelagic eggs, but it cannot go unmentioned that this is unlikely to be a representative sample of summer flounder populations, not even for those locally occurring in the New York Bight. Three females and 3-5 males are really way too few. In the first sub-experiment, these three females are used as replicates, when in fact, there are not. I don't understand, why there weren't any true replicates for each female and treatment, which would yield a better evaluation of the true variation in responses. However, the variation between the three females is already substantial, and while I agree that all three showed the same directional response (good), the potentially large variation between individuals needs to be discussed.

In addition, while the survival to hatch is certainly important, it's currently awarded prominent space in the abstract, whereas the survival during the much larger experiment is not reported at all. That should be corrected. At the very least, you know with how many individuals you started, you know how many you took out for various sub-samples, and you know how many remained in each replicated after 28 days. It would be very important to report the resulting survival numbers, and discuss then why the small genetic variability and the minimal level of replication might have limited the conclusions.

Histopathology

The descriptions of the histopathology are currently hard to understand. I believe the reader should be told in a bit more detail what 'minor liver sinusoid dilations' or 'focal hyperplasia' are, and why they are believed to be detrimental rather than incidental. In addition, the current microscopy imagery is not helpful, because the reader cannot compare between 'normal' and 'abnormal' features. Lastly, but most importantly, since all of this is so qualitative – did the person who evaluated those sections know what treatment the fish came from? If so, there is little to convince me that these qualitative evaluations are indeed unbiased.

Water chemistry

The authors chose to omit many details of their methods for sea water chemistry manipulations. For instance, we are simply told that CO_2 was stripped from the incoming water, but not how. There is no mention of which pH meters were used and how they were calibrated (NBS or total scale?). Most importantly, I think it's not appropriate to refer readers interested in all these details to an unpublished paper. The reader of this paper has a right to look at a table that summarizes the water chemistry measurements and their variation over the course of the experiment and between replicates. Parameters to report may include pH, DIC, Total alkalinity, pCO_2, CO_3^{2-}, and perhaps aragonite saturation state.

Minor comments & edits

P13898, L9: "… water temperature, and their interactions." This was perhaps tested but not included in this manuscript; hence, it should be deleted from the abstract. It's confusing.

P13898, L11: "… of this region." This is unclear, because there is nothing in this sentence stating what region.

P13898: Add survival estimates from the main larval trial to the abstract to balance the extreme responses seen in one subset of experiments

P13898, L18: "…levels, with …"

P13898, L24 " … and dilation of the liver sinusoids …" That seems redundant.

P13899, L3: "…will be challenged by ocean acidification in the near future." Given your limited parentage, unreplicated experiments, and very high CO_2 treatment levels, I strongly disagree with the boldness of this statement. Please conclude something more cautious.

P13899, L13: "… decrease in surface pH of ~0.4 units …" Please double-check this figure, it seems way too high to me. We know that a current pH of ~8.0 translates into ~400 ppm CO_2, a rise to 750 ppm will likely result in a pH of about 7.8. Your own ambient water had about 750 ppm has a pH of 7.8.

P13899, L15: "…especially for continental shelf …" Well, further down you say that the pelagic shelf water where summer flounder eggs develop are relatively stable, but I seriously doubt that they are at 750 ppm already, especially not in autumn and winter. The whole logic here is flawed.

P13899, L23: "… antherid…" I believe you want to say 'atherinopsid'. That's the correct term for the family of New World Silversides.

P13900, L8: replace “youngest fish life-stages which …” with “youngest life-stages, which …”

P13901: General approach: I understand the motivation of the authors to outline their greater ‘step one-two-three strategy’ here. However, I feel it’s more confusing that helpful, e.g., L9: “…series of conducted or planned experiments …”. To improve clarity and conciseness, I suggest sticking to the experiments that were actually conducted and which are reported here.

P13903, L5: “source water … is cleaned …” Does that mean filtered? If so, through what mesh size was used? Were there other ‘cleaning procedures’? Please provide more details here.

P13903, L6: “…CO_2 stripping from air and water …” Please provide details of that methodology.

P13903, L8: What volume had the specimen exposure containers?

P13903, L9: “pH probes” Please provide make and model, plus specify, what ‘regular’ means. At what intervals were DIC samples taken? Please provide DIC data et al. in a supplementary table, regardless of the status of the publication by Wieczorek et al.

P13903, L14: Please specify the approximate collection date of your broodstock adults. How long was the acclimation period? Weeks? Months?

P13903, L16: “2 to 3m diameter” Were the tanks of different size? That’s a huge difference. Please give volumes rather than diameters.

P13903, L24: “3 to 5 males” The way this is written it is not clear whether the 3-5 males fertilized all three females, or whether there were 3-5 different males for each female.

P13904, L17: “4 week post-hatching” → “four weeks post-hatching”

P13904, L26: “…protocols previously used at our laboratory.” You need a reference here, even if it’s a report or something. If none is available, please briefly describe these protocols.

P13905, L10: Reference needed for claim that response variables “have the potential to be correlated with the Fisherian fitness of the individual.” Why isn’t it sufficient to say ‘fitness’?

P13906, L4: “Image Tool software”. Is this a particular software name? If so, please give the maker. If this is just some generic term, maybe it’s better to call it image analysis software.

P13906, L6: “development, and cranio-facial …”

P13907, L6: “… were examined for differences in tissue and cellular morphology…” This qualitative evaluation is only unbiased if the observer has no knowledge of the treatment origin of the larvae. Was this done in a blind or double-blind way?

P13909, L2: “The relative survival to hatching was reduced by approximately half…” I think, here in the results, these descriptive statements should be followed by actual numbers in parentheses.

P13909, L10: “…each replicate.” Given the distinctly different parentage, these are not true replicates.

P13909, L22: “…expectation that sets of morphological …”

P13912, L18: “minor liver sinusoid dilation.” What does this actually mean? How many of the examined larvae showed this? If the authors believe this is in some ways detrimental to the fish, they need to discuss this somewhere.

P13912, L21: “… evident, which …”

P13914, L8: Baumann et al. (2012)

P13915, L10: better reference needed for trade-off claim, unpublished data seems inappropriate here

P13915, L14: “Greater larval lengths and growth rates in CO_2-enriched environments have been observed in inland silverside…” No, that’s wrong. The paper by Baumann et al. (2012) shows the opposite, smaller larval lengths at higher CO_2 concentrations.

P13916, L22: “…considerable ecological significance in summer flounder (Chambers, unpublished data). Please replace with a proper reference.

P13917, L1: “…would amplify the consequences of ingressing in the autumn versus spring.” I did not understand the argument made here. Can you elaborate in one sentence, how exactly you hypothesize this amplification to occur?

Figure 1: Unit of pCO_2 missing on the X-Axis label.

I’m confident that the authors can use the suggestions to strengthen their paper, which I’m looking forward to read in Biogeosciences.

Literature cited

Baumann H, Talmage SC, Gobler CJ (2012) Reduced early life growth and survival in a fish in direct response to increased carbon dioxide. Nature Clim Change 2:38-41

Caldeira K, Wickett ME (2003) Anthropogenic carbon and ocean pH. Nature 425:365-365

Riebesell U, Fabry VJ, Hansson L, Gattuso JP (2010) Guide to best practices for ocean acidification research and data reporting. Publications Office of the European Union:260