Dear Dr. Scranton,

Thank you for taking the time to read and comment on our manuscript. We appreciate your observations and questions. Below is our response to your comments.

Comment: "The basic premise on which the authors base their study is that methane production depends on substrates produced during fermentation of organic matter and that methane production and sulfate reduction usually do not occur in the same sediment, presumably because of competition for these substrates." **Response:** In this comment, you seem to imply that methane production and sulfate reduction do not occur in the same sediment. Methanogens and sulfate-reducing bacteria do directly compete for organic substrates such that when sulfate or other more energetically profitable alternative electron acceptors are available, some methane is still produced, but it is at lower rates (please see manuscript citations: Achtnich et al. 1995; Lovley and Klug 1983, 1986).

Comment: "The authors mimic sea level rise by adding brackish water to their samples (which ignores the possible importance of increased water logging and decreased oxygen in the sediments)." **Response:** These sediments came from freshwater wetland ponds (0.4 – 1.2 m total water depth) that were already waterlogged and therefore had decreased oxygen in the sediments. We have no reason to suspect that the physical act of sea-level rise would change the saturation status of the sediments.

Comment: "All incubations are done after vials are flushed with nitrogen although the natural sediments apparently all had some level of oxygen present in situ." **Response:** In this comment and throughout your review, you seem to imply that purging vials with nitrogen gas is problematic. It is widespread practice in methane production bottle assays to purge with nitrogen gas in order to remove the oxygen that was introduced by removing sediment from below the water's surface and exposing it to the air in the field and in the lab (Lofton et al. 2014; Sinke et al. 1992; West et al. 2012, 2015). Of course, it is possible that there are microsites where oxygen is present in the freshwater wetland sediments, but we believe that the same may be true in the bottle experiments.

Comment: "Values for nitrate (another potential substrate for carbon remineralizing organisms) and for acetate (the putative important substrate for methanogens) were measured but data were apparently quite variable and are not reported." **Response:** Summary porewater acetate levels are reported in Results section 3.1.1. You are correct in that we did not report actual values for nitrate, and we will add a short line in the results to indicate that nitrate was usually 2+ orders of magnitude lower than total sulfate levels. We measured acetate, nitrate, and sulfate fairly extensively in both the water column and the porewater as well as other parameters. We did not present all the data we collected because we did not want to overwhelm the reader, but these data are all available in the archived data. If reviewers and editors feel that these data would enhance the manuscript, we will add them in tabular or figure form to the manuscript or as supplemental information.

Comment: "No measurements were made of H2, another potential substrate for methanogens and methanogens were not examined to see whether they were actually acetoclastic or hydrogenoclastic." **Response:** We did not measure H_2 production as we do not currently have an instrument available to measure this nor did we classify the type of methanogens. However, acetate was the predominant factor explaining methane production patterns in almost all experiments, which suggests that a large proportion of the methanogenesis is acetoclastic (e.g., pg. 8 lines 29-30). However, we do not deny the

potential of hydrogenotrophic methanogenesis occurring in our ecosystems (please see last paragraph on page 10).

Comment: "The data for the dsrA and mcrA genes are again not presented and in the results section appear to contract the conclusions made about their abundance. The authors also ignore the fact that numbers of genes do not directly relate either to number of cells (a cell can have more than one copy of a gene) or to gene activity." **Response:** The *dsrA* and *mcrA* summary data are presented in the Results section 3.1.3, and individual gene data per sample are also archived in the data set (see microbial tab). Although we acknowledge that the number of genes does not equate with number of cells or gene activity, qPCR of functional genes for particular guilds is a commonly used approach to estimate the abundance of a functional group. We also acknowledge in the manuscript that it is possible that some of the genes we detect are from dormant microbial communities (please see pg. 11, lines 8-12). The number of genes corresponding to abundance is always going to be a bit of an assumption, but in a broader survey of our wetlands, we have unpublished data from our system that *mcrA* abundance relative to *dsrA* abundance is correlated with methane production rates. Also, in *Letters in Applied Microbiology* (Volume 62, Issue 2, Pages 111-118), Morris et al. 2016 found that hydrogenotrophic methane production rates corresponded to *mcrA* abundance. Although observation of functional gene transcripts or even the actual enzymes catalyzing the reactions would clearly be preferred, logistics associated with our remote field sites precluded work with these more labile macromolecules. Despite challenges associated with interpreting DNA copy number of functional genes, we believe that our use of qPCR substantially adds to the literature as there are few studies that simultaneously measure methane production and microbial functional group abundances. For example, the studies listed on pg. 9, lines 22-26, have hypothesized that microbial community processes are behind methane patterns along a salinity gradient, but none of these studies actually tested this hypothesis by measuring microbial communities.

Comment: "Finally the authors use natural organic matter to enrich their incubations but do not indicate the amount of carbon added or the relative lability of that carbon. (Clearly a gram of sucrose and a gram of twigs would not be expected to stimulate microbial activity to the same extent)." **Response:** We added 3 g of live tissue per macrophyte species, which translated to 0.10-0.12 mol of C added to each incubation, and we will add these values to the methods of the manuscript. Dr. Scott Tiegs of Oakland University has measured the % C, % N, and % P of the four litter species we added: Marestail (45% C, 1.7% N, 0.17% P), Buckbean (44% C, 0.94% N, 0.15% P), Lily (45% C, 1.7% N, 0.17% P), and Horsetail (47% C, 2.5% N, 0.24% P), but the patterns we observed did not follow %C, %N, %P, C:N or C:P, which are considered standard indices of litter quality. We discuss the phosphorus in detail because Tiegs et al. 2013 found that litter decomposed more rapidly when P was higher (see pg. 11, lines 14-24). You are correct in that we did not specifically measure the lability of the carbon from each macrophyte species, but we are leveraging information from Tiegs et al. 2013, whose study had already assessed this by measuring decomposition rates of these litter types in Copper River Delta ponds. We will clarify this in the discussion.

Comment: "P2 Line 6: I do not like the term green house compensation point. It is not widely used and does not directly relate to what you are measuring (as you say nothing about carbon sequestration in this system)." **Response:** We believe that the term "greenhouse compensation point" accurately describes wetlands' dual roles in carbon sequestration and emissions, and we believe that because northern wetlands are on the edge of this compensation point that it is so important to study how

global change may alter the methane cycle. Greenhouse compensation point is a good way to set up the broader context of this study.

Comment: "You do not mention carbon fertilization again, and there is no indication in the text of whether any increases in carbon production in the CRD system would be due to warming or to carbon fertilization. I think you hurt yourself by trying to draw connections to too many issues. I would recommend drastic simplification of this section and sticking to the facts." **Response:** Increased CO_2 levels lead to warming as well as CO_2 fertilization, both of which could affect the amount of substrate available. We are not precisely sure whether increased organic matter will result from warming, or longer growing seasons, directly or from CO_2 fertilization and therefore it seems prudent to mention both potential mechanisms. Regardless, we will reevaluate our discussion of these points based on your recommendation and consider ways to focus our comments.

Comment: "P2 Line 16: redox conditions are only indirectly related to climate change. You seem to imply the mere presence of nitrate or sulfate changes redox conditions but this is not true, especially if oxygen is present. In fact there can be a lot of nitrate and sulfate in oxic surface sediments. Did you ever have sulfide in your samples? Does oxygen penetration vary with ecosystem? How far below the surface were the samples collected?" **Response:** True, redox conditions are a potential indirect effect of climate change. Sea-level rise could increase sulfate levels just as higher decomposition rates due to warming could deplete oxygen levels, both of which affect redox conditions. The sediment came from the top 20 cm, and although some oxygen is likely present in small amounts, it is also likely depleted quickly. For example, the layer of water directly above the sediments often has low DO levels of around 1 mg/L, particularly in the evening, and it is not uncommon to have anoxic groundwater upwelling in these systems. We also observe that water column DO levels drop throughout the season as vegetation begins to senesce. As for hydrogen sulfide, we did not directly measure that, but sediment characteristics (black coloration and pungent odor) suggest the presence of sulfide. Usually sediments become anaerobic within the first few cm of freshwater ponds, which is another reason that we make the incubations anaerobic by purging with nitrogen gas. Of course oxygen matters, but the amount of nitrate in particular is orders of magnitude lower than agricultural and other human impacted systems. Lastly, if significant levels of oxygen were present in these sediments, it would kill the methanogens because they are extremely sensitive to O_2 (please see manuscript citation: Whalen 2005) and we would therefore see very little methane production in our experiments.

Comment: "P2 Line 23: I don't think an "abundant supply of organic matter can reduce competition for methanogens by increasing substrate availability". Try instead "abundant supply of organic matter can increase substrate availability"" **Response:** We appreciate the wording suggestion and will make the change.

Comment: "P2 Line 28: replace "are likely results of" with "may be influenced by" " **Response:** We appreciate the wording suggestion and will make the change.

Comment: "P3 lines 7 and 10. Use same unit for sealevel rise (100 and 170 cm)" **Response:** We appreciate the wording suggestion and will make the change.

Comment: "P3 line 15: Numbers of methanogens not as important as whether or not the methanogens are active." **Response:** Please see our response to your major comment above about qPCR and the feasibility of RNA work.

Comment; "P4 line4: The range of physicochemical parameters in table 1 are actually pretty small for most measurements. Perhaps more important is whether the intertidal sediments are exposed to the atmosphere at low tide (tidal range?). How long are they submerged? What is the water content? Again it matters how far below the surface these sediments were collected. From the table it must be shallow since there was more O2 in these sediments than the freshwater wetlands." **Response:** The intertidal sediments were collected from the top 20 cm just as in the freshwater wetlands, and they are covered with freshwater during low tide and increasingly brackish water during high tide. So again these sediments are waterlogged with depleted oxygen levels. In Table 1, the DO data are actually from the water column, as we do not have oxygen data for the sediment. In the freshwater ponds where limnological profiles could be conducted, we reported DO levels from the bottom of the water column, but in intertidal marsh, DO levels came from the surface layer. This is detailed in the legend for Table 1. We will remove the water column DO data altogether since they seem more misleading than helpful.

Comment: "P4 line 26: Were no replicates run for sediments from a single site? How can you tell if observed variability is just typical of replicate samples? I would also think you MUST indicate how much macrophyte tissue you added (probably in terms of gC/g sediment or something like that) to even know if these treatments were similar since the lability of the carbon is likely not the same." **Response**: We ran five control replicates without added substrate at each freshwater wetland (n = 5), which we then averaged to form the basis of the delta CH4 production metric. Each macrophyte treatment was replicated in 5 different ponds. In the increased organic matter simulation, we used pond as the replicate because we were more interested in capturing how wetlands differing in biogeochemistry along a glacial to oceanic gradient would respond to organic matter addition rather than how much variability there exists within a single wetland's response.

Comment: "P5 Line 6: You mean incubation temperature not ambient temperature?" **Response"** Yes, we do. We appreciate this good suggestion for a wording change.

Comment: "P5 line 8: Purging with nitrogen will likely have a bigger effect than incubating at a few degrees cooler than the actual sediment. I would expect stimulation by this as you allow more anaerobic processes to occur. Were any blanks or controls run?" **Response:** Please see our response to your major comment above.

Comment: "P5 line 12: "flame ionization" not "flame ionizing" detector" **Response:** We will make this change.

Comment: "P5 line 20: You report that only sulfate was detectable in the water column. Did you present any water column data? Do they mean anything? What are detection limits for pore waters? Present nitrate and acetate data for sediments as well as sulfate." **Response:** We do have extensive water column data for these wetlands, some of which are in the archived data. If reviewers feel these data add to the manuscript, we can add them in tabular form or in supplemental information. Some (i.e., DOC and sulfate) are also already presented in Table 1, but acetate and nitrate were not detectible in the water column (we will add concentration detection limits to the manuscript which were about 10 and 5 uM, respectively).

Comment: "P5 line 27: the method you describe is typically called "loss on ignition" and represents a loss of organic matter. Did you convert to organic C? Is this data reported anywhere in the paper? Was DOC measured the same way? Blanks? Detection limit?" **Response:** Yes, we took the sediment organic

matter data (in % of OM) from loss on ignition and converted it to organic C (pg. 5, lines 25-29). We will add in the following citation about conversion to organic C: Thomas et al. 2005; Aquat. Sci. 67:424–433. The data were not reported in the paper as to not overwhelm the reader, especially since the amount of carbon did not affect the results, but the data are archived and available online. DOC was measured using a Shimadzu TOC-V (pg. 5, lines 18-19), and we did run blanks. All samples had detectible DOC, but our lowest standard was 1 mg/L and all samples registered above this level with the exception of five of the intertidal samples, which fell in between the blank and the lowest standard. If reviewers would like more detail on these analyses in the methods, we will add them to the manuscript.

Comment: "P6 line 2: I would use the word "converted" rather than "scaled" " **Response:** Thank you for the wording suggestion.

Comment: "P6 line 5: I couldn't tell what this composite sediment was used for. Were incubations for each site or was all sediment made into a composite? Why do this? It seems that then the averages for the genes refer to different samples than the physiochemical properties. Can you properly do statistics comparing the combined samples in one parameter to the individual samples in one parameter?" **Response:** The composite sediment was a combination of the five sediment samples from different sites for each freshwater wetland, but this was only done for the microbial analyses. Making a composite is highly practiced in soil microbial ecology for the purpose of controlling analytical costs while still capturing significant spatial heterogeneity. We made a composite for each freshwater wetland which were then compared against ten sediment samples from intertidal marsh. So yes, it is true that we could not directly link methane production of a single sample or its physicochemical properties to the functional group abundances, but overall we were able to characterize methanogen and sulfate-reducer abundances across ecosystems and relate these to average methane production rates.

Comment: "P7 line 2: move "log transformed" to immediately follow "four factors"" **Response"** Methane production rates were log-transformed, not the factors. We cannot make this change because it would misrepresent our statistics.

Comment: "P7 line 26ff: You use a lot of significant figures for something that is so variable. Maybe you are justified in two significant figures but not 3 or more. Again blanks and detection limits need to be mentioned as your averages are pretty close to zero (or at least almost include zero). I would like to see these data associated with specific samples" **Response:** We will reduce the number of significant figures and make the needed detection statements in the methods. We only reported parameters if they were detectible. The reason that these averages are close to zero in this particular line is that we converted actual concentrations to the total amount in the incubation bottle (i.e., µM * PW volume in incubation * (PW volume per ml of sediment) * sediment volume in bottle).

Comment: "P8 line 14: This paragraph seemed very odd. You seem to contradict yourself a lot. For intertidal samples, three had no dsrA but this was found in all freshwater AND the dsrA was independent of ecosystem type. This seems to contradict your hypothesis which you nevertheless cling to. Remember the presence of a gene doesn't mean the organisms are doing anything at the moment and the number of genes does not necessarily indicate the number of cells of a particular organisms." **Response:** Indeed, we did not detect *dsrA* in 3 of the 10 intertidal samples. Nevertheless, *dsrA* abundance tended to be higher in intertidal ecosystems, although not significantly so. In contrast, the *mcrA* presence and abundance did vary significantly by ecosystem, and methanogens are the group that

we tend to use to explain our methane production results directly, while we use sulfate and sulfate-reducer abundance as supporting evidence (please see pg. 9, lines 14-16).

Comment: "P8 line 27ff: I really don't like your equating sealevel rise purely with sulfate concentration. It is ok to say that sea level rise will flood current intertidal areas, but the vegetation will change and the water will be permanently water logged rather than periodically exposed to air. You have no data on estuarine wetlands to compare to the freshwater wetlands (although others have done this comparison in other systems). I would guess the reason you see no effect here is that you are mimicking the process the wrong way. You are really looking at increased sulfate levels, not sealevel rise." **Response:** In these systems, which are waterlogged throughout the year, we expect sulfate concentration to be a major biogeochemical change. Our simulation not only adds sulfate, but it also adds other nutrients and the microorganisms brought in with the saltwater. You are correct that over the long term vegetation and other characteristics of these ecosystems will also change. We will add comments at pg. 3., line 12, and pg. 4 line 20 to reflect our biogeochemical focus on the effects of sea-level rise. Our data on intertidal ecosystems for comparison with freshwater ecosystems can be found in Results section 3.1.2 and Figure 3.

Comment: "P10,line2: The sentence beginning "our study also demonstrates..."seems to directly contradict your own data as you said before that dsrA did not correlate with ecosystem type" **Response:** Just because *dsrA* presence and abundance did not significantly correlate with ecosystem type does not mean the presence or abundance of sulfate-reducing bacteria does not affect methane production in intertidal ecosystems. Even though these bacteria are also present in freshwater ecosystems, they could be limited by sulfate availability. Also *dsrA* tended to be higher on average relative to freshwater ecosystems, just not significantly so. We will change the wording to "tended to have generally higher sulfate-reducer abundances when present" to address this concern.

Comment: "P10 line 13ff and next paragraph: These two sections are wild speculation with absolutely no data behind them. There are a lot of other factors that might be important that you haven't included." **Response:** We provide potential hypotheses and explanations for our data that were grounded in the literature, which is not an uncommon practice in the discussion. Specifically, we discuss the acetoclastic pathway of methane production and how we might expect that to change in the future as well as possible reasons for why methane production rates varied by an order of magnitude throughout the season. We could omit these sections, but felt that it was important to say something about the results we observed. Regardless, we will reevaluate the nature and extent of our explanations.

Comment: "P11line14: This whole section is hurt by the fact that you don't characterize the macrophytes at all in terms of potential lability. Did you add a constant amount of C or organic N? Or just dry mass? Or just a "chunk" of leaves? Again much of this section is speculation without more facts" **Response:** Please see our response above to your major comment about lability of macrophytes. We evaluated many different possibilities for explaining the results of the different macrophyte treatments including % C, % N and % P, but the quality measurement that seemed to matter the most was antimicrobial properties. None of these potential indicators of lability appeared to have an effect on methane production, which we will clarify in the Discussion pg 11, lines 14-30, and pg 12, lines 1-9.

Comment: "P12 line19: You have no idea whether there are hydrogen utilizing methanogens around or how much hydrogen there might be. You can reiterate what other people have said in other papers but I

don't see you connecting these to your system with any facts." **Response:** We will try to improve the clarity of the manuscript, by acknowledging that methanogens in general utilize acetate, directly (acetoclastic) or indirectly (hydrogenotrophic). We will also add that CO_2 and H_2 are compounds that also come from fermentation of organic matter such as acetate. Hence, this is why acetate is important and was measured in this study.

Comment: "Modeling: I didn't really understand the models. If you are making linear equations including various parameters and then constant factors for each, I would like to see more detail on how this worked. If you have a lot of variables, you can fit most data but figure 6 is completely mysterious and doesn't convince me of the value of your model. I also didn't understand the columns in tables 2 and 3. I have never heard of an AIC before, for example. Explain the statistics a bit to an audience that may include non-biologists." **Response:** Akaike Information Criterion (AIC) is an increasingly utilized form of model selection that generates estimates of the model being the best representation given the data and the set of models being explored. It also penalizes for the number of parameters in the model. We will add more description of this in the methods. Basically, general linearized models are created, each of which is associated with an AIC value, which is then used to rank the models. We also corrected for small sample sizes (AIC_c). The model with the lowest AIC_c value is considered the best, and all remaining models are compared relative to the best approximating model using delta AIC_c (Δ_i). Models with a Δ_i less than or equal to 2 are considered to have substantial support, while models having a Δ_i greater than 7 have little support (Burnham and Anderson 2002). The relative strength of the models was then evaluated with Akaike weights (ω_i), which indicate the probability of a model being the best model, given the data and the set of candidate models (Burnham and Anderson 2002).

Respectfully,

Carmella Vizza, Will West, Stuart Jones, Julia Hart, and Gary Lamberti