Response to Reviewers

The line numbers we refer to below are taken from the "clean" version of the manuscript we have submitted (manuscript.docx).

Reviewer #1:

All of my prior concerns have now been addressed

Reviewer #2:

In the revised manuscript, attempts were made to address the comments on the previous submission. However, I cannot follow the way one of my major previous points was addressed: the lack of temporal analyses.

The new figure S4 is said to show that dynamics are fairly linear --- I cannot see this. I also do not know what the explanation in the legend: "OTUs [...] that have a significantly different temporal trend by the histopathologic summary score by LEfSe analysis" means. This sentence contains insufficient information as to what has actually been done. The methods in line 383, "We tested for differences in temporal trends through fitting a linear model to each OTU and testing for differences between histopathological summary scores with LEfSe (85) in mothur (default parameters, LDA > 3)" do not clarify this.

We thank the reviewer for their feedback on this figure highlighting areas that needed to be clarified. We understand "fairly linear" was an incorrect word choice to describe the figure S4 in our Response to Reviewer document. We meant that relative abundance relationships at day 0 are maintained throughout the experiment for most of the bacteria identified with significantly different trends. For example with Alistipes OTU 160, severe mice had high relative abundance on day 0 and at their final time point. We acknowledge there is quite a bit of variation in these data. We have clarified how we determined the significant trends in our methods. "We tested for differences in temporal trends through fitting a linear regression model to each OTU and tested for differences in regression coefficients by histopathological summary scores with LEfSe (85) in mothur (default parameters, LDA > 3)." (Lines 386-388)

In the main text, these temporal results are discussed in the following, hard to follow way

"These bacteria groups increased in severe outcomes maintained their differences throughout the length of the experiment (Figure S4). These results agreed Aside from Helicobacter, these groups of bacteria that associated with more severe outcomes did not have a conserved association between their relative abundance and the disease severity across all mice." Besides the fact that this sentence seems to contain some mistakes, the taxa discussed in the preceding section of the manuscript have little overlap with the data shown in figure s4. Where are Klebsiella, Eggerthella, Bacteroides,...?

*We agree this statement was hard to follow. We have added text to the results to introduce Figure S4 and then more clearly connect Figure S4 to specifically Helicobacter. We added "Lastly, we tested for associations between temporal changes and disease severity (Figure S4). Most groups of bacteria maintained higher relative abundance, relative to the other other outcome groups, from day 0 through the end of the experiment." (Lines 169-171), "In our experiments, when *Helicobacter* was present, the infection was more likely to result in a high histopathologic score (Figure 4C, S4)." (Lines 283-285)*

I have made several comments on the model underlying figure 5. This has been removed and replaced by a simpler logistic regression. Again, different taxonomic aggregation levels are used in A, B and C. The legend mentions colors that are not visible in the figure, the confidence intervals (?) of coefficients are not labeled as such. The model in C) investigates predictors of a high histpathologic score (presumably, and we must guess) a coefficient great than 1 indicates an association of the corresponding taxon with a high histopathologic score while a coefficient less than one indicates an association with low histopathologic score. Lowest coefficient here is for Alistipes. Yet, this genus is highest in the "severe" category in figure s4, and more abundant over time in the high histopathologic group than the low group.

We thank the reviewer for bringing these issues to our attention.

We chose to aggregate taxonomic levels based on the specific model since higher classification levels performed as well as lower levels. Since the higher levels performed as well as lower levels, the effect may be attributed to group effects as opposed to OTU specific functions. A and B did not perform better with more resolved taxonomy, indicating that genera or order sufficiently classified the outcome. We have added this explanation to make our reasoning more explicit. "We used the highest taxonomic classification rank which performed similar to lower ranks, which suggested the effect is associated with general attributes of the bacterial group as opposed to specific functions of more refined grouping." (Lines 177-180)

We have corrected the descriptions of the data and colored the data to match the description. "Median (solid points) and interquartile range (lines) of the odds ratio are plotted.

*Bacterial groups are ordered by their odds ratio. * indicates that the bacterial group was unclassified at lower taxonomic classification ranks. (A) Bacterial members grouped by genus predicted which mice would have toxin activity detected at any point throughout the infection. Data with a decreased probability of toxin activity are colored light purple and those with an increased probability of toxin activity are colored dark purple. (B) Bacterial members grouped by order predicted which mice would become moribund. Data with a decreased probability of moribundity are colored light blue and those with an increased probability of moribundity are colored dark blue. (C) Bacterial members grouped by OTU predicted if the mice would have a high (greater than the median score of 5) or low (less than the median score of 5) histopathologic summary score. Data with a decreased probability of high histopathologic score are colored light green and those with an increased probability of high histopathologic score are colored dark green." (Lines 806-819)*

Alistipes in Figure 5 (OTU 166) is a different OTU than the Alistipes OTUs in Figure S4 (OTU 89, 92, and 160). OTU 166 in Figure 5 did not follow the same pattern as the other Alistipes OTUs and thus was not identified in the temporal analysis as significantly different over time. Therefore, OTU 166 may have a different interaction with its community than the other Alistipes OTUs identified.

In summary, I think the results are too messy to be clearly interpretable and I do not feel that the comments were addressed with sufficient care.