Reviewer #1 (Comments for the Author):

In this manuscript the authors investigate the generally ignored contribution of the source of animals in microbiota analyses, by determining the microbiota profiles of mice from 6 sources during antibiotic treatment (clindamycin) and subsequent challenge with a toxigenic C. difficile strain. This allowed them to determine genera that are consistently impacted, irrespective of source, and developed models that were predictive of C. difficile clearance based on key species.

The study has implications for the way microbiota studies are reported and raises a number of interesting points regarding variability between subjects in microbiome analysis; the transparency with respect to code and data is commendable. I also appreciate the fact that the authors took care to formulate a title that does justice to the limitations of the study ("in mice"). Nevertheless, I would suggest that the use of "C. difficile clearance" more accurately reflects the content of the manuscript (as authors also state in L232-233 that susceptibility, possibly associated with colonization, is not affected).

> We thank the reviewer for this suggestion and have incorporated clearance into the manuscript title: "The initial gut microbiota and response to antibiotic perturbation influence *Clostridioides difficile* clearance in mice"

Generally, experiments are performed well, conclusions justified, and the manuscript is well written.

There are some issues that I would like to suggest to improve the manuscript further and to facilitate this I provide detailed comments below.

The authors do not define colonization. E.g. in the Abstract it is written that: "all mice were colonized 1-day post-infection" (L8-10). Are the authors sure this is colonization, not passage? The same can be said at multiple points throughout the manuscript (e.g. L101 "the colonization dynamics were siumilar"). I think the authors should either define their definition of colonization specifically, or rephrase (in other places this is done more accurately, e.g. "C. difficile clearance", "C. difficile was detectable"). In the reviewer's opinion, initial CFU likely represent transient presence of C. difficile upon outgrowth, but do not reflect (stable) colonization.

> Our mouse model challenges mice with a low dose of *C. difficile* spores instead of vegetative cells

to ensure the *C. difficile* we detect post-challenge germinated and expanded in the intestine. We have modified the following sentence to indicate colonization is transient in the clindamycin mouse model: "We have previously demonstrated mice are rendered susceptible to *C. difficile*, but clear the pathogen within 9 days, thus colonization is transient when treated with clindamycin alone (21, 43)."

There is limited discussion on the fact that the mouse model used allows recovery of nearly all mice (630 infections in mice are rarely lethal). It would be appropriate to present in the Discussion a paragraph on the possible implications of their findings for strains that could cause more severe disease.

We acknowlege that *C. difficile* 630 does produce mild infections in mice compared to other strains. It is interesting that we did see 2 deaths after *C. difficile* challenge with *C. difficile* 630. We expect that differences across sources of mice could yield to variation in disease severity and have modified the following sentence in the discussion to address this point: "Evidence for immunological toning differences in IgA and Th17 cells across mice from different vendors have also been documented (65, 66) and could influence the host response to CDI (67, 68), particularly relevant for strains that induce more severe disease than *C. difficile* 630."

There appears to be an issue with Supplemental Figure 2, as text refers to panels up to H, but Figure only contains panels up to E. Possibly, some panels have been lost during conversion. Please check. Also, the figure has poorly visible panels C-E. I suggest placing A and B next to each other, and increase the size of panels C-E and place them underneath panels A and B.

Thank you for calling attention to this discrepancy, we have corrected the text to refer to Fig. S1, which does have the correct number of panels: "There were some differences between the 2 experiments we conducted, as the experiment and cage effects significantly explained the observed community variation for the Schloss and Young lab mouse colonies (Fig. S1A-B and Table S4). However, most of the vendors also clustered by experiment (Fig. S1C-D, F), suggesting there was some community variation between the 2 experiments within each source, particularly for Schloss, Young, and Envigo mice (Fig. S1G-H)."

We have also rearranged Fig. S2 as suggested by the reviewer.

Authors have longitudinal data for all experiments (daily fecal samples), but appear not to leverage this potential as only pairwise comparisons appear to have been made. Methods designed for repeated measurements instead of cross-sectional based methods like Metalonda

2

or MetaSplines could be used. Statistics on the longitudinal dataset could give insight into the consistency of the significantly different abundance of OTUs between colonized/cleared mice at day 7, and could strengthen the findings (also relevant for Figure 6).

> Thank you for the suggestion, we opted for comparisons at specific timepoints because of the variation in the mice we had samples and sequence data for each day. We did test whether any taxa significantly varied between the colonized/cleared mice on day 7, but none were significant after correcting for multiple hypothesis testing. We suspect that the majority of the fluctuations occured during the first 3 days of the experiment.

Minor comments 1. L6: please include the region used (V4)

> Due to abstract word count constraints we specify that we sequenced the V4 region when we first mention 16S rRNA sequencing in the results section (L68, see point 3 below).

2. L64: were the Michigan mice originally obtained from a commercial vendor? If so, which one?

> Yes, the Michigan mice were originally obtained from Jackson in 2002, which is stated in the Material and Methods section.

3. L68: "sequenced the 16S rRNA genes" or "performed 16S rRNA sequencing"

> We have modified to specify the region sequenced, per the reviewer's suggestion with point 1 above: "sequenced the V4 region of the 16S rRNA gene""

4. L70 and on: the section states that "after they acclimatized". It is unclear how this relates to the study setup, in Figure 2 for instance. I would welcome a more detailed description on this.

> In the Materials and Methods section we define the acclimatization period as the 13 days between when the mice ordered from commercial vendors arrived at the University of Michigan and the start of the experiment. We have added the following sentection to the Figure 2A legend to define the acclimatization period: "Mice that were ordered from commercial vendors acclimated to the University of Michigan mouse facility for 13 days prior to antibiotic administration."

5. L83-87: in the reviewer's view, it is not entirely clear what comparisons are made (which groups, timepoints). Can they authors clarify?

> We are comparing OTUs that vary between the 6 sources and OTUs within a single source that

vary between experiments. Both comparisons are referring to the baseline (day -1 timepoint). We have modified the sentences to clarify: "After finding differences at the community level, we next identified the bacteria that varied between the 6 sources of mice. There were 268 OTUs with relative abundances that were significantly different between the sources at baseline (Fig. 1D and Table S5). Though we saw differences between experiments at the community level, there were no OTUs that were significantly different between experiments within Schloss, Young, and Envigo mice at baseline (all $P > 0.05$)."

6. L84: please place the number in perspective to the total number of OTUs detected (is 268 a minor or a major fraction of what you detect?).

After samples were rarefied to 5,437 sequences, we detected 1,935 OTUs in our dataset. We display the number of OTUs (richness) detected in each mouse in Fig. 1A, 3A, and 4A to place the number of OTUs that vary between sources of mice in context.

7. L101-102: to clear - better "to clear C. difficile"

We have added *C. difficile* to the sentence as suggested by the reviewer: "While the colonization dynamics were similar between the two experiments, the Schloss mice took longer to clear *C. difficile* in the first experiment compared to the second and the Envigo mice took longer to clear *C. difficile* in the second experiment compared to the first (Fig. S2A-B)."

8. L125-126: as earlier (see point 5), I would appreciate a clarification with regard to for instance timepoint.

We have modified the sentence to indicate the timepoint: "However, there were only 18 OTUs with relative abundances that significantly varied between sources after clindamycin treatment (Fig. 3D and Table S8)."

9. L135: "was primarily found only", suggest removing only or primarily

We have removed only as suggested: "*Enterococcus* was primarily found in mice purchased from commercial vendors and also increased in relative abundance after clindamycin treatment (Fig. 3D)."

10. L165-167: samples were grouped per timepoint (different sources of mice were put in 3 groups depending on timepoint), correct? Please clarify.

That's correct, we took the OTU relative abundances from all sources of mice that had sequence data at the 3 timepoints: baseline (day -1), post-clindamycin (day 0), and post-infection (day 1).

We have modified the sentence to clarify: "We trained three L2-regularized logistic regression models with either input bacterial community data from the 6 sources of mice at the baseline (day = -1), post-clindamycin (day = 0), or post-infection (day = 1) timepoints of the experiment to predict *C. difficile* colonization status on day 7 (Fig. S3A-B)."

11. L208: Perhaps more accurate to state 'ridge regression model' than logistic regression. A standard logistic regression model does normally not include any penalization.

For consistency with how we refer to the logistic regression models throughout the rest of the manuscript, we have added L2-regularized to specify the type of model: "We trained L2-regularized logistic regression models with baseline (day -1), post-clindamycin treatment (day 0), and 1-day post-infection fecal community data that could predict whether mice cleared *C. difficile* by 7 days post-infection better than random chance."

12. L221: facilitating is overstating. The data shows that it is associated with colonization, but facilitating would require experimental evidence to this effect.

We have modified facilitating to positively correlating: "We found *Enterobacteriaceae* increased in all sources of mice after clindamycin treatment, positively correlating with *C. difficile* colonization"

13. L257; certain mouse models use antibiotics other than clindamycin. I would appreciate a statement to this effect (how should the authors' findings be seen in this respect), considering that clindamycin treatment appears to be the dominant factor in determining susceptibility.

We agree that different antibiotic treatments will likely result in different *C. difficle* infection outcomes in mice. We refer the reviewer to L248-250 in the discussion section of the manuscript: "While we have demonstrated that susceptibility is uniform across sources of mice after clindamycin treatment, there could be different outcomes for either susceptibility or clearance in the case of other antibiotic treatments."

14. L262: comma missing after Additionally

We have added the comma after Additionally.

15. L263: types should be type

We have corrected to type.

16. L293 and 294: I think the authors mean to state that all these factors influence the microbiota, not the other way around? (at least for diet and age)

That is what we meant and we have corrected to the following: "The outcome after *C. difficile* exposure depends on a multitude of factors, including age, diet, and immunity; all of which also influence the microbiota."

17. L346-348: please include a single sentence on the results of the controls, to verify quality of processing.

We thank the reviewer for suggesting we add in the control results. We have added the following sentence to the Materials and Methods section: "Based on the mock communities, our overall sequencing error rate was 0.0112% and all water controls had less than 1000 sequences."

18. Line 360-362: Could the authors perhaps give some explanation on the used distance metrics? I am not familiar with this specific metric. More importantly, it is unclear to this reviewer whether the assumption of homogeneity of dispersion between the comparison groups is tested (can be done using the betadisper function). It is known that even if centroids of comparison groups are very close together in e.g. a PCoA plot, adonis can still yield a very significant value due to different dispersions between the groups. You would then not be looking at actual different compositions between groups, but at (large) differences in composition within a group. If this is the case for the current analyses, it would be useful to report this, or an alternative (recent) test could be used which is robust to this issue (https://microbiomejournal.biomedcentral.com/articles/10.1186/s40168-019-0659-9).

θ_{YC} distances are based on the Yue & Clayton measure of dissimilarity. We have modified the Materials and Methods Section to clarify: "PCoAs were generated based on the Yue and Clayton measure of dissimilarity (θ_{YC}) distances (75)."

Regarding the PERMANOVA tests used to analyze overall differences between the community structures of the 6 sources of mice, we agree that there are numerous approaches to differentiate microbial community structures. In our previous work we found that heteroscedacticity did not affect the Type I error of the PERMANOVA test (see AMOVA data in first column of Table 1 from DOI: 10.1038/ismej.2008.5). Regardless, the data for which PERMANOVA was used in the current study are pretty obviously different when looking at the ordinations (e.g. see Fig 1C). Furthermore, when we investigated the suggested test, there was not much documentation in the package to show us how to use the tool and it was not clear that it was possible to employ more sophisticated experimental designs than a single classification ANOVA. In contrast, we were able to use vegan to build a statistical model to account for the vendor and experiment. Based

on these points, we are confident in the results of the statistical tests using PERMANOVA as implemented in the vegan R package.

19. L371: please provide some additional detail about the cross-validation (e.g. k-fold cross validation, repetition of cross-validations). Why did the authors opt for a ridge regression and not a lasso or elastic net regression?

We have modified the sentence to include additional detail on the cross-validation: "The L2-regularized logistic regression models were trained and tested using the caret package (78) in R as previously described (79) with the exception that we used 60% training and 40% testing data splits for testing of the held out test data to measure model performance and repeated 2.5-fold cross-validation of the training data to select the best cost hyperparameter."

We chose ridge (L2-regularized) logistic regression because our group has previously implemented this model with microbiome data (Topcuoglu et al mBio 2020) and found the L2-regularized logistic regression model performed almost as well as a random forest model when classifying colorectal cancer cases using fecal microbiota data and did not require as many samples as random forest to avoid over fitting the data.

20. L377-378: have the authors considerd using a linear mixed model to investigate the dynamics of C. difficile clearance over time? (as I understand it now per-timepoint comparisons were performed).

The reviewer is correct in that we chose per timepoint comparisons. We chose to analyze at specific timepoints because we did not have datapoints for *C. difficile* CFU and sequencing data for each mouse across every timepoint.

21. In general, is anything known about the genetic make-up of the mice from the different sources? Are there distinct genetic differences that could underlie the differences in microbiota?

Yes, there is evidence of genetic differences across different C57BL/6 substrains that could influence the microbiota. Teasing out the degree to which environmental factors and genetic differences contribute to the microbiota differences we observed between the sources mice was outside the scope of our current study. We have modified the following sentence to acknowledge genetics as an additional factor influencing the microbiota. "The outcome after *C. difficile* exposure depends on a multitude of factors, including genetics, age, diet, and immunity; all of which also influence

the microbiota."

22. Fig. 1D: please indicate which test was used.

We have added that OTUs were identified by the Kruskal-Wallis test with Benjamini-Hochberg correction to the Fig. 1D legend.

23. Fig. 3: what is the reason for using high values for the y-axis, when no datapoints are shown in the higher regions? Would it be better to use a narrower scale so smaller differences between groups can be better appreciated?

The high values on the y-axis of Fig. 3A-B allow comparison with Fig. 1A-B and 4A-B plots, which have the same y-axis scale and highlight the overall impact of clindamycin on community diversity across all sources of mice.

24. Fig 5A: the horizontal light grey bars at the same position in each panel are not defined. What does this indicate?

These lines indicatee the limit of detection, we have added the following to the Fig. 5A legend to clarify: "The gray horizontal lines indicate the limit of detection."

Reviewer #2 (Comments for the Author):

The initial gut microbiota and response to antibiotic perturbation influence Clostridioides difficile colonization in mice

The study by Tomkovich et al. shows how the starting microbiota from mice purchased from six different vendors is different. They also define how the microbiota is altered with the addition of clindamycin followed by C. difficile (CD) challenge and monitor colonization over a period of 9 days. The main take a ways from this study are that mice from different vendors, including in house colonies, have different starting microbiotas. When they are treated with clindamycin to render them susceptible to CD colonization, they again are altered differently. From here the authors assessed CD load and saw that colonization patterns were affected differently. They went on to create a model to determine which OTUs were associated with CD clearance and or persistence over this time period. Again, adding to the list of bacteria that are potentially protective and or antagonistic to CD colonization.

This is a well written and well powered study. I also appreciate the transparency of the study by including many supplemental files that include all the data. I also think it will be received

very well from the field as it is showing how we could potentially get different results based on vendor and to consider this when working with CD mouse models. The study does fall short on mechanism and does not dig into how these bacteria are able to contribute to clearance, BUT that was not the point of the study. I do have a few minor comments that I think it would be nice for the authors to address below. Overall, very nice study!

Minor comments:

1.The authors use post infection vs. post challenge in the paper. Since they are not looking at disease I would suggest they go with post challenge throughout the paper. Also, did they look at disease at all throughout as this would have also been really interesting? I know they looked at weight loss, but this does not always correlate with disease. I also see that there were three mice that died during the experiments. Can the authors comment on why? Was this from CDI?

For consistency with our group's previously published studies with *C. difficile* 630, we have changed all instances of post-challenge to post-infection. Our hypothesis going into this study was the microbiota variation across different sources of mice would impact *C. difficile* colonization resistance. Previous studies have characterized *C. difficile* 630, and found it to be mild in terms of the amount of histological colitis and weight loss that was observed in contrast to mice infected with VPI 10463 (Theriot et al. Gut Microbes 2011). Thus, we opted to focus on clearance and not disease severity. We suspect that the 2 deaths we observed in our study that occured post-infection were from the CDI (the other death happened prior to infection and was unrelated). Reviewer #1 had a similar comment related to the strain we were using and we have modified a sentence in the discussion to address disease severity variation between *C. difficile* strains: "Evidence for immunological toning differences in IgA and Th17 cells across mice from different vendors have also been documented and could influence the host response to CDI, particularly relevant for strains that induce more severe disease than *C. difficile* 630."

2.I see that only female mice were used in this study. Can the author's note this as a limitation in the discussion and or justify why they did this?

Since sex has been shown to influence microbiota composition in mice (Wang et al. Frontiers in Microbiology 2019 "Core gut bacteria analysis of healthy mice"), we opted to eliminate this confounding factor by using only female mice. Additionally, we were interested in placing our findings in context to studies that have observed different *C. difficile* infection outcomes despite

using similar mouse models and several of these studies also used only female mice. We have added the following to the discussion to highlight this limitation: "One study limitation is that we only used female mice. Sex has been shown to influence microbiota variation in mice (45), so we used female mice to reduce this confounding variable and also match the sex used in previous CDI studies that administered clindamycin to mice (32, 33, 44, 46)."

3.I found Figure 5A to confusing. I would suggest trying to make this clearer for the reader.

Figure 5A displays the 10 highest ranked OTUs in our model with the best performance. We have added the following sentence to the text to clarify: "On day 0, the majority of these OTUs were impacted by clindamycin and had relative abundances that were close to the limit of detection (Fig. 5A)."