**Editor comments:**

**Unfortunately, the reviewer and editor believed that the revised manuscript did not meaningfully engage with the original review. As such, the paper in its present state cannot be considered for publication in mBio. I agree with the editor that the paper might be better suited in its current form to a specialized journal. A new manuscript could be considered if the authors are willing to revise the manuscript according to the reviewer's comments.**

**Invited Editor:**

**As you can see from the comments of the reviewer, they indicate that the concerns raised during the first round of reviews have not been adequately addressed. I also agree with this assessment, particularly with respect to the demonstration of a more general applicability of the findings beyond the specific case study presented here.**

**Thus, I recommend either submitting to a more specialized journal that would be more appropriate for publishing such a case study, or substantially revising the manuscript and resubmitting. In particular, it is important to place this work in the broader context - using clustered sequences as references is already common practice, yet not adequately acknowledged here. Also, the absence of a comparison to ASVs is notable given that the vast majority of studies now rely on ASVs. Finally, while I understand that MCC is a measure of clustering quality, the reality is that end users care more about the biological insights that can be derived from the data. I would argue, that it is easy to identify scenarios in which a good clustering is not biologically meaningful, or where biologically relevant findings can be derived from (mathematically) suboptimal clusterings. Thus, it is critical to include a thorough analysis of the biological implications of the methodology beyond what is currently discussed.**

We thank the editors and reviewers from previous rounds of review at *mBio*. We are anxious to have this work published. Unfortunately, it appears that we are at an impasse with them on the content of this study. They seem to want to readdress the results of other papers that we have published in *mBio* and *mSphere*. We hope that the manuscript is improved because of their feedback. We look forward to receiving an assessment of the revised manuscript from *mSphere*.

**Reviewer #1:**

**Secondary review of Machine learning classification by fitting amplicon sequences to existing OTUs**

**Qt 2-3 ... "Machine learning classification using the gut microbiome relies on assigning 16S rRNA gene sequences into operational taxonomic units (OTUs) to quantify microbial composition."**

**OTU classification of 16S sequencing is not the only to characterize the gut microbiome. Even if restricting to DNA sequencing methods, ASV analysis of 16S sequencing and taxonomic profiling (or phylotyping) from 16S or shotgun data are widely used. More widely used than the methods acknowledged here. But they are ignored in this manuscript.**

**The ASV and phylotyping methods being ignored do not have the "lack of stability" issue of de novo OTUs. ASV analysis is also de novo. Every commonly used method other than de novo OTUs already has the "ability" described in L45-47.**

**Inadequate recognition of standard methods for the same problem permeates the manuscript. This revision has not adequately addressed this commont from the original review "The authors need to situate their methods within the universe of methods commonly used in this area."**

We have thoroughly rewritten the introductory paragraphs to better frame the scope of this manuscript. With regards to the specific question of being able to diagnose colonoic screen relevant neoplasias (SRNs), we have previously published models using other taxonomic levels including ASVs, OTUs, genera, families, order, class, and phylum in *mBio* (DOI: 10.1128/mbio.03161-21). That analysis showed that de novo OTUs at a 3% distance cutoff did better than ASVs and other taxonomic levels. Thus, we do not feel it is necessary to represent this previous work. The updated text does a better job of highlighting this specific case where de novo OTUs outperform other approaches. Furthermore, it highlights that there are likely other situations where models based on de novo OTUs will be needed. Because of the challenges of de novo OTUs (e.g., lack of stability), we feel that this paper using OptiFit to fit new data to existing de novo OTUs makes a meaningful impact.

When we received reviews from the paper describing OptiFit (manuscript mSphere00916-21), which was published in mSphere (DOI: 10.1128/msphere.00916-21), a reviewer had this to say:

*From my perspective, the development of OptiFit makes a perfect partner for OptiClust in a Train/Test (Train/Classify) splitting situation, where the users would split data, train using OptiClust and then classify the remaining reads using OptiFit. Of course, pursuing this would be well beyond the scope of this manuscript, but the authors may consider discussing such a possibility for future work.*

This work directly addresses this insight from the reviewer.

**From response to reviewers: "The purpose of this paper is to provide an example of a scenario when the previously publish OptiFit algorithm could be useful. We are demonstrating here that it is possible to use your own data as a reference for consistent OTU classifications without a needing a reference database."**

**This is a narrow scope. De novo clustering to serve as a reference for subsequent closed-reference OTU assignment is as old as the microbiome field. All the 16S references commonly shared are de novo cluterings, that are then used by a closed-reference cluster method like Optifit.**

We disagree with the reviewer on their point that “De novo clustering to serve as a reference for subsequent closed-reference OTU assignment is as old as the microbiome field”. Prior to OptiFit, closed reference clusterings were done to de novo clusterings of the greengenes database using a tool like VSEARCH; not to a reference collection of sequences from the same dataset. As has previously been shown in previous papers from our lab, these approaches to closed reference clustering (1) defined reference OTUs using full length sequences, (2) only included one representative sequence per OTU, (3) was dependent on the order of the reference OTU sequences, (3) when a subregion (e.g., V4) is used, many reference sequences from different OTUs are found to be identical rather than 3% different, (4) the OTU definition is effectively 6% since the mapping threshold is a radius rather than a diameter as it is with de novo clustering (see DOIs: 10.1128/msphere.00916-21, 10.7717/peerj.1487, and 10.1128/mSystems.00027-16).

In regard to the claim that our study is “a narrow scope”, we would point back to the comment from the reviewer of the OptiFit manuscript. We hope that the revised text does a better job of conveying the value of the approach described in the current study.

**L62: Vsearch is a software package that implements a lot of methods? More clarity on the closed-reference method used.**

This statement has been revised.

**From thh response: "The MCC score is a measure of OTU cluster quality based on the similarity of sequences and whether they are appropriately clustered together or not. Since we added additional methods, we can see that some methods do have lower MCC but equal model performance. This likely indicates that the model depends on well clustered OTUs. We've added some discussion on this to the paper (lines: 118 - 122)."**

**I don't follow the logic here, and it doesn't seem to correspond to L118-122. That model performance is unrelated to MCC seems to indicate that the model does NOT depend on high-MCC OTUs.**

**I also don't think the question from the original reviewer about why MCC is being evaluated here has not been answered. Reading the Optifit paper, it appears that Optifit specifically optimized the MCC metric at 3% OTU differences. So the result reported here that Optifit/Opticlust scores higher on that metric is circular. Vsearch isn't trying to create OTUs that optimize that metric. Why should one should care about MCC at all when considering predictive accuracy of the kind evaluated here? If another method has a low 3%-MCC, but better model performance, that would be the approach of choice. 3%-MCC score is irrelevant.**

As described in our previous papers addressing clustering quality (e.g., 10.1128/AEM.02810-10, 10.1128/mSystems.00027-16, 10.7717/peerj.1487) there have been many approaches to evaluating the quality of de novo OTUs. The innovation of using MCC scores is that it is an objective metric of whether OTUs defined at a certain threshold represents clusters of sequences at that threshold. So, if someone says that they are using 3% OTUs, MCC will indicate how well that threshold is being applied. Whether a specific threshold represents biology is a question that is far bigger than what we are addressing in this study. Again, we have looked at other levels of taxonomic resolution and found that 3% OTUs perform well for training machine learning models to diagnose colonic SRNs. We included the MCC scores for the different clustering strategies used in this study because we felt it helps explain the variation in modelling performance. It is entirely possible that poor clusters or different clusters (e.g., ASVs, phyla) could do a better job of training a model, but that is not what we have found in this application.