Reviewer 1

The authors come to a sweeping conclusion about a technology using a version of the platform that is no longer available and has been superseded by two new generations of polymerase enzyme and sequencing chemistry. Since the paper focusses on a single technology, its conclusions using the P4-C2 chemistry are of very limited value to scientists currently working with the technology (current version: P6-C4). While it has to be appreciated that studies like these shoot at a moving target, the authors would at the very least have to make it very clear that their conclusions are only valid for the older version of the platform. The very strong statements of the current manuscript convey a message that ignores reported improvements.

Ideally, the authors should include additional analyses on the V1-V9 amplicon using the latest version of the PacBio platform. I believe this could be done without too much additional effort and greatly strengthen the impact of the paper. Alternatively, the authors need to place very visible disclaimers making it clear that all analyses are based on data generated with the P4-C2 chemistry, which is obsolete. Error rates will also be influenced by the initial amplification. Were the exact same conditions used to generate the amplicons used for PacBio sequencing that were also used in the publication by Kozich et al., that serves as the reference for the error rates of the MiSeq platform.

Reviewer 1 raised the important point that because the sequencing technology is a moving target, our conclusions were too strong and more importantly, out of date. In the last year we regenerated the data using the P6-C4 chemistry and the latest recommendations from PacBio on the appropriate methods for generating amplicon data. With that in mind we have performed a thorough rewrite of the paper that comes to a very different conclusion than the original manuscript. We appreciate the reviewer calling us on this point.

Reviewer 2

A general comment is, given the longer reads possible with this platform, perhaps there might not be need to focus on only one or a few commonly used hyper-variable region, but to focus on the nearly-full-length sequences. Considering the 3% divergence threshold commonly used for species definition, on the order of 40 errors could be permitted for a sequence length of 1500, and the goal of accurate genus classification and OTU clustering might be adequate. This may of course have other consequences when most analyses are done de novo, but I would suggest reference based analysis for chimera removal and OTU clustering. Although the authors have already done a very fine job with their assessment, perhaps they could address this.

Given the cost of generating sequences by this platform, we feel that the primary use case for the technology is in generating reference sequences. As we discussed in the previous version, it is debatable whether sequences with 40 errors would be appropriate as a reference. We do not foresee the utility of this approach for broad scale sequencing surveys when there are cheaper methods for generating deep samples of datasets.

Lines 140-141: Could the authors clarify the nature of the quality score from PacBio (Figure 1A does not appear to be comparable to that seen from other platforms)? How does it compare to that of Illumina/454 and how does mothur cope with the differences if there are any?

We have added text towards the end of the first paragraph of the Results and Discussion section that points out that the PacBio quality scores are not the traditional Phred scores. In contrast to what we previously

observed using the older chemistry, the new quality scores are not helpful in assessing the quality of the sequence data.

Lines 279-282: "We found that the frequency of the most abundant 1-nt error paralleled the number of sequences. There were two sequences in the V4 dataset that occurred 76 times, one sequence in the V1-V5 dataset that occurred 30 times, and one sequence in the V3-V5 dataset that occurred 17 times." Are there any general characteristics of the most abundant errors that could provide insight into why certain errors tend to occur? e.g. high AT content or high nucleotide diversity overall (or in a particular sliding window)?

Unfortunately, we were not able to identify any signatures beyond the various insertion, deletion, and substitution biases in the overall data set.

Figure 2 is bit difficult to follow with this "numbering" scheme, suggest to change to colored lines.

Part of the design of this figure was that the regions generally overlap on top of each other indicating that the effectiveness of the methods is reproducible. We have colored the numbers and increased the horizontal jitter.

Reviewer 3

Comments: In my first few readings of the manuscript it was unclear how the additional sequencing depth was taken into account. In the beginning of the results it states the error rates were ~1.8% on average. But it was unclear whether this is the baseline error rate for a single pass, or whether it was the result of taking a consensus of multiple passes. Since the introduction states the single pass error is ~15%, it must be the latter which makes sense given the results, but I'd recommend explaining this at the beginning of the results section.

It would also be helpful to explain how the multiple reads are used to reduce error rates. For example, if the single pass error is 15%, then the chances of seeing an incorrect nucleotide on 5 or more out of 10 passes would be something like 0.006%, which seems pretty good. But I'm guessing that non-random errors invalidate this simple estimate, but I cant tell for sure because there's no markings on the y-axis to figure 1C. If the high error rates stem from the fact that coverage rates are generally much lower than 10X, that would be good to know, because it means the platform could potentially be useful if read lengths are increased.

We have added text to the beginning of the first paragraph of the Results and Discussion section to indicate that ever read we analyzed had at least 3-fold coverage. As we go on to demonstrate, at some point (>10-fold coverage) increased coverage does not improve the error rate. This was a surprising result and we suspect that the non-random nature of the errors is responsible for the departure from the expected error rate.

We have added text to the legend indicating that the y-axis scales are the same in Figure 1B and 1C.

Finally, it was not clear if this study was based on a single sequencing run at a single facility. If so, it might be important to point out that results could vary from run to run, especially if read lengths/sequencing depth plays a big role.

It was based on a series of sequencing runs at a single facility. We have added text to the Conclusion section to indicate the value of sequencing mock communities in parallel with samples to make sure that the error rate is maintained regardless of the facility.

line 215: "whether" should be deleted.

This was removed in the rewrite

line 233: "one would unable…"

This has been corrected.

line 243: should say "when we achieved…".

This has been corrected.

lines 244ff.: in one place, they say they observed between 6 and 63.1 extra OTUs. in another place, between 7.4 and 86.8 extra OTUs. how do you have a fractional OTU?

Because this is based on rarefying the data, these numbers represent an average number of extra OTUs.

line 246: "sequence" should be plural.

This has been corrected.

Figure 2 is illegible. why not just have colored points or something? or add more left-right jitter?

This has been altered as indicated above.

line 280: if I understand correctly, it seems that some errors occurred once, but a small number occurred many times. That suggests the latter type may be errors during PCR or even cell division, which are not necessarily relevant to the sequencing methodology. If they are biases of the platform, however, that would be a major problem, so it would be good to have some clarification on this point.

Given the expected error rates of the PCR polymerase is at least one or two orders of magnitude lower than the sequencing error rate, we suspect that these errors are due to the platform.