Editor comments:

Overall, the reviewers and I agree that the manuscript is novel, timely, and worth publishing. It is extremely well written, with excellent examples of current and potential applications and outcomes of the use of preprints. With one example, "Arsenic Life" (mentioned kindly), serving as the "poster child" in support of preprints. At the outset, I wondered if he had subjected this paper to the Preprint process (he did) so it includes a "real time" test of the concepts he develops and discusses. The author is to be commended for taking on this study, beyond the call of scientific research papers, and I am sure it will have an impact on thinking about, discussion of, and ultimately the use of preprints in microbial sciences.

Main issues with the manuscript by Reviewer 1 lie primarily with documenting references for many of his assertions. Reviewer 2 comments are very thoughtful, though a bit more philosophical. All points should be considered in a revision, including added citations, and brief additions to the 'history' of preprints. I think an extended discussion of the role of preprints in physics is beyond the scope of this paper. All comments - accepted or not - should be addressed both in a response to reviews with changes to the manuscript noted as well as arguments against changes.

Thank you for the kind comments and useful feedback from the two reviewers. I am excited to learn that the manuscript received such a positive reaction. You will see my comments in a normal font interleaved with the reviewers' comments in a bold font. In addition to responding to the reviewer comments I have made several additional changes to the manuscript. First, the original version of the manuscript used preprints that were posted by December 31, 2016. The revised manuscript uses those preprints posted prior to April 17, 2017. Second, on March 24, 2017, NIH released a notice (NOT-OD-17-050) addressing preprints. The previous version of the manuscript intimated that there might be something in the works at NIH and this notice addresses many of those comments.

Reviewer 1

1. This is overall an excellent and very useful paper on preprints in one general area (microbiology)

2. I believe the overall analysis is sound and the claims are generally supported by the evidence.

3. There are a few areas of the manuscript that could use modification - mostly just additional detail. I provide line by line comments for any part of the manuscript that I believe could use some modification below.

4. It would be good if all references to web sites include a data when the web site was accessed and in many cases it would be good to include some information from the site as supplemental material. After all, web sites can change and referencing them is tricky. More detail in line by line comments below.

5. The posting of a reproducible version of the analysis in Github is great and I would recommend highlighting this in the text when the analysis is first mentioned.

I appreciate the feed back and have done my best to incorporate your comments in the revised manuscript. I added a sentence to the end of the Acknowledgements section that citations of information contained within any webpages was current as of April 21, 2017; I thought it was too clunky and redundant to repeat this accession date repeatedly. Regarding the last point, I have added a reference to the paper's GitHub repository to L248.

Lines 18 and 19.

"Preprints were initially adopted among physicists and biologists in the 1960s as a method of sharing interesting research amongst colleagues (5)"

I think it would be good to discuss a little bit more the history of preprints and of sharing manuscripts prior to publication. The text here makes it seem like preprint sharing started in the 1960s when I believe this is untrue. The reference cited here appears to focus on

organized efforts to share preprints in the 60s (e.g. an attempt by NIH and efforts by the physics community). While important these were not the beginning of preprints and thus the text is a bit inaccurate

In addition, I think it is not quite accurate to say this was a means of "sharing interesting research among colleagues". The efforts in the 60s were attempts to make preprints more broadly available and with less of a bias where some people saw preprints but others did not.

Furthermore, I think there were probably many reasons why people did this including establishing priority and staking out territory (i.e., it was not just about sharing interesting research among colleagues).

I have modified the text here to make clear that it was the first formal structure (as opposed to people informally sending manuscripts to each other) with the goal of broadly disseminating paper-based preprints. I also added a reference to Till (2001), which has a nice overview of predecessors of electronic preprint servers.

Line 36

"preprints are typically publicly available in about 24 hours"

Is there data supporting this claim

This has been my experience based on posting 7 preprints and is the specific claim made by *bioRxiv*. I have added a clause to the beginning of this sentence indicating that these observations are taken from information provided by the *bioRxiv* and *PeerJ Preprints* websites.

Lines 42-45

"This difference can be meaningful to authors since some journals, including the American Society for Microbiology (ASM) Journals, will only accept submissions that have been posted on preprint servers hosted by non-profit organizations"

Is there a reference for this claim?

I have added a link to the *mBio* Instructions to Authors which states the global ASM Policy:

> **Preprint policy.** ASM Journals will consider for publication manuscripts that have been posted in a recognized not-for-profit preprint archive provided that upon acceptance of the manuscript for publication, the author is still able to grant ASM copyright or agree to the terms of an Open Access license and pay the associated fee. It is the responsibility of authors to inform the journal at the time of submission if and where their article has been previously posted, and if the manuscript is accepted for publication in an ASM journal, authors are required to update the preprint with a citation to the final published article that includes the DOI along with a link.

Lines 48-49

"their work to a journal as journals will not consider manuscripts posted as preprints under a CC-BY license (e.g. Proceedings of the National Academy of Sciences)."

Should this read "to a journal as SOME journals"

This has been clarified since *PNAS* is the only journal I have been able to find with this restriction.

Lines 53-55.

"Compared to the bioRxiv site, the PeerJ Preprint site is more fluid, gives readers the ability to"follow" a preprint, and provides better access to article keywords and the ability to search preprints."

This should include a date of accessing the sites and I would recommend perhaps submit screenshots as supplemental material.

I have added Figures S1 and S2, which provide screenshots of two preprints at the different sites.

Lines 55-57

"With broader acceptance of preprints by traditional journals, many journals, including all of the ASM journals, have established mechanisms to directly submit manuscripts that are

posted as preprints on bioRxiv."

and

Lines 57-58.

"The only direct submission mechanism for manuscripts submitted at PeerJ Preprint is to the PeerJ journal"

A reference / web link would be useful here. Plus a date when these policies were examined would be helpful. I would recommend including the text of the policies as supplemental material if possible.

I have added a sentence to the end of the manuscript indicating that the policies were current as of April 21, 2017. I think it would be inappropriate to copy the text of policies as supplemental material since it is not my own work.

Lines 68-71.

"A relatively new example of what this might look like is PrePubMed (http://www.prepubmed.org/), which seeks to index preprints from numerous sources. A more organized effort is being initiated with funding through ASAPbio to create a"Central Service" that would aggregate preprints in the life sciences (http://asapbio.org)."

It would be good to include the date of access of these sites and if possible some additional information in supplemental material such as text from the site or screen captures to support the text.

I added a sentence to the end of the Acknowledgements section that citations of information contained within any webpages was current as of April 21, 2017

Lines 72-73

There is also hope that the National Institutes of Health (NIH) will renew their interest in indexing preprints as separate research products than peer-reviewed publications.

Please clarify. Hope from whom?

This section has been rewritten in light of the NIH Notice on preprints (NOT-OD-17-050).

Lines 81-83.

"First, a significant amount of attention has to be given to the potential dual use research of concern (DURC) since posted results in microbiology research could offer insights to individuals seeking to engage in terrorist activities"

DURC is not just about terrorist activities - it is about any possible misuse of research information.

This has been corrected.

Lines 83-85

"Second, for researchers engaging in research that involves human subjects it is critical that assurances be made that institutional review boards have been consulted and have approved of the research."

I would suggest adding a line here about animal research too and how it is critical that appropriate reviews were done for such research.

A comment about appropriate oversight for all vertebrate research has been added to this sentence.

Lines 89-90

"Again, while hoping to maintain the efficiency of the preprint format, traditional microbiology journals have policies for these issues in place that should be easy to implement by preprint servers"

Could you clarify or provide examples of what you mean by policies.

I have clarified this sentence to indicate that I meant that most journals have screening procedures and oversight committees that look for problems relating to ethics, DURC, etc.

Lines 100-101.

“Each take a generally permissive stance towards posting of preprints prior to submission.”

Given the following sentences about journal policies changing it would be good to mention how the journal policy was inferred and if possible include references and or web links with dates and such.

Links to two sites that aggregate journal policies are provided in the sentence that follows this one.

Line 145-146.

“Considering the preprint is a citable work with a DOI, it would, in fact, be the preprint author that scooped the second.”

I personally agree with this statement but many in the community do not. This only works if people view preprints as valid publications. There are many examples where peer reviewed papers have been published claiming priority on some novel finding when preprints existed on the topic and where the authors of the peer reviewed paper and sometimes the editors of the journal have stated something akin to “we do not consider preprints valid papers”.

I have added a sentence in this section indicating that it remains to be determined whether the journals that claim to apply “scoop protection” will stand by these policies.

Lines 160-162.

“Some fear that the use of preprints will allow scientists to circumvent page limits by posting preliminary manuscripts.”

It would be helpful to reference examples of people expressing such fear if they are available online anywhere.

I have added a citation to the FASEB letter in response to the NIH’s RFI regarding the use of preprints.

Lines 173-176

“In fact, several funding agencies including the Wellcome Trust and the UK Medical Research Council encouraging fellowship applicants to include preprints in their materials; meanwhile, the NIH is in the process of soliciting input from the scientific community on their role in grant applications.”

References / links would be very useful here.

This section has been rewritten in light of the NIH Notice on preprints (NOT-OD-17-050).

Lines 193-195

“Any manuscript that was published went through several month delays in releasing information to health care workers, the public, and scientists needing to learn new methods to study a previously obscure virus.”

I think the wording here and in a few other places in the paper is a bit confusing or awkward. Preprints are published too. So I would suggest using some alternative wording to “published” for papers that have gone through peer review vs. preprints for those that have not.

I have edited the manuscript to make this clear. I am trying to use the NIH definition of preprints that these are documents that are not final. I would take “published” to be “final” and since preprints are interim research products that are “posted”. I get that this is a nuance that is likely lost on many people. I have tried to take the suggested approach of differentiating between what is peer-reviewed and what is not.

Lines 207-214

Discussion of NY Subway metagenomics.

I think an important part of the story has been left out. On February 17, 2015 Chris Mason wrote a long blog post (at the request of this reviewer) for microBEnet about their paper and about some of the challenges in inferring the presence of anthrax in the subway. See https://www.microbe.net/2015/02/17/the-long-road-from-data-to-wisdom-and-from-dna-to-pathogen/

The focus of this manuscript and the section in question is on how preprints have been used to help move the scientific record. I appreciate Chris Mason's comments on the microBEnet blog, but I have also pointed out his response and correction to the scientific record, which was at least partially motivated by Petit et al's preprint posted on Zenodo.

Lines 265-267

"Although the hosted commenting is only one mechanism for peer review, this result was somewhat disturbing since the preprint model implicitly depends on people's willingness to offer others feedback."

Any information you could provide on the relative balance of commenting in other systems (e.g., Twitter, Facebook, Blogs) would be very helpful here. Or perhaps at least refer to the discussion later in the article

I have added a sentence to this section to indicate that although twitter, facebook, and blogs are a mechanism for providing feedback these comments are often lost since they are not directly connected to the preprint.

Lines 301-303

"Although it is impossible to quantify the quality or impact of research with individual metrics, it is clear that preprints and the publications that result from them are broadly accepted by the microbiology community"

I would love for this to be true. But I am not sure I completely follow the logic of this claim here. Could it not be that the citations for preprints are mostly coming from the authors of those preprints? I know we in my lab cite our preprints quite a bit. Or could it be that a subset of the community cites preprints a lot but others do not?

I have edited this sentence to make it more clear that the metrics I presented indicate that the science presented in preprints are accepted by microbiologists on a level comparable to more traditional publications.

Reviewer #2

One of the first and most acute causes of culture shock in physicists who have fallen among biologists is the relative role played by preprint servers in the two sciences. The arxiv has been a standard part of physics for years; a place where preliminary results can be displayed, critiqued and refined. In contrast biology preprint servers like biorxiv have only recently begun to attract attention and widespread use. For those who don't yet know about preprints, this perspective is a good introduction.

The MS is very strong on educating the reader about preprints and their advantages for an individual researcher. But less so when it comes to the advantages to the field in general. I think this is important, because the value to the individual and the field are entwined with how preprints are viewed, and the respect in which they are held, and this article is an opportunity to push the dial in the direction of more respect.

Physics and biology are very different enterprises. The position of arxiv in the former is not (only) due to a more open-minded attitude among physicists, but the history of the science and the esteem in which preprints are held. So in a highly competitive field such as Fermi gas microscopy, researchers upload their work as soon as possible and their priority is recognized. Similarly, dense theoretical work can be trialed online in front of a tiny community of peers who can suggest alternatives. These may be extremes/caricatures, but it is true that the important factor is the status of preprints in the community. And while preprints have all the advantages stated in this article, unless their status changes they will remain underused.

In a high status preprint world, as the MS states, publishing on a preprint server does get your paper out and could be seen as 'scooping' competitors. But we are a very long way from that at present. Why cite a preprint that came out well over a year ago, and is accepted but still waiting to see the light of day, when there's an alternative in a glamor journal that came out a few months later? The preprint might get some nice comments on twitter, but nobody is going to see it as having priority.

Also in a high status preprint world, folks would be more likely to comment. At the moment, only a few preprints get many comments, which limits the value of the approach quite a lot. I know the MS says this, but it could say more about it as without this many of the advantages claimed for preprints fall away - so what if you can show your preliminary findings to the world if the world doesn't give you feedback? I suspect it is also true that the numbers of comments are not only down to the inherent interest of the content, but also the profile of the authors. If Nick Loman (for example) puts something on twitter about his preprint, it will get looked at more. Similarly a Very Eminent Scientist* has tried to convince me that the large amount of attention and comments their preprints receive is wholly down to their extraordinary high quality, and has nothing to do with the fact that they are member of the National Academy with a very high profile. Such distortions will never go away, but might be less likely if preprints were consumed with the same interest as 'proper' journals.

So I urge some rewriting to raise these issues -perhaps explicitly recommending that funders mandate that papers should be put up as preprints at the same time they are initially submitted for publication. I have the following specific points I noted as I was reading

Line 14 - these other things like blog posts seem interesting, but quite different from preprints. I'd suggest ditching them or making the focus clearer

This has been edited as suggested.

64 repetitions of landscapes fractured etc

This sentence has been edited to minimize the repetition.

82 and beyond. Several issues are raised here that are never fully dealt with. The most important is DURC. I think you should say a bit more about how this can be handled responsibly, and the potential for preprints to be subject to minimal review to prevent some of these things.

I have clarified this point in the second to last sentence of the paragraph (LXXX).

186 Could preprints not attract adversarial comments as well? I am not sure scientists are fully free from the attributes of BTL commentators. What about the possibility of vexatious criticisms? This could be especially true for 'political' topics.

I agree that it may still be adversarial; however, given the public and free nature of preprints, it is far less likely for responses to preprints to be adversarial.

249 I wonder if preprints are more common in fields with a high proportion of physics types?

I suspect they are more common among groups of biologists where there is a culture of openness. For example, the widespread use of open source software among those in bioinformatics and genomics likely translates to openness in research. Unfortunately, I don't have direct data to test this and I think it is probably beyond the scope of the current manuscript.

286 the public, or other scientists?

Public. The metric incorporates inclusion in popular media and does not incorporate the peer reviewed literature. I have edited this sentence to make both points more clear.

294 impact factors are 'more traditional' than altmetric, but more 'controversial'? I have to say this made me giggle. I also confess I have no idea how to interpret these scores so this bit was useful to me.

I have edited this sentence to make my intention more clear: "A controversial, yet more traditional metric of impact has been the number of citations an article receives."

301 The comparisons with mBio are really neat, but there is a skew evident in the comparison of the means and median citations. It looks like the majority of preprints get very few citations, but a few get many more. The mBio citations look more normally distributed. Can you say more about this?

This is one of the big problems with Journal Impact Factors (JIFs), since they focus on the mean, rather than the median. Considering the mean is more sensitive to outliers, it is really an

inappropriate metric for citation counts, which are typically zero-inflated and long tailed. I have included the means to provide a basis for comparison to JIFs.

321 I think this is a bit misleading as review is still needed before publication. Sidestepping is not possible.

I have edited this sentence to make my intent more clear, “we were able to publicize our results prior to lengthy review processes”.

365 brining? :)

This has been corrected to “bringing”