

Preprinting Microbiology

Patrick D. Schloss[†]

[†] To whom correspondence should be addressed: pschloss@umich.edu; Department of Microbiology and Immunology, University of Michigan, Ann Arbor, MI

Format: Perspective

Counts: ~5600 words plus 50 references, 1 figure, and a 150 word abstract

1 **Abstract**

2 The field of microbiology has experienced significant growth due to transformative advances in
3 technology and the influx of scientists driven by a curiosity to understand how microbes sustain
4 myriad biochemical processes that maintain the Earth. With this explosion in scientific output, a
5 significant bottleneck has been the ability to rapidly disseminate new knowledge to peers and the
6 public. Preprints have emerged as a tool that a growing number of microbiologists are using to
7 overcome this bottleneck. Posting preprints can help to transparently recruit a more diverse pool
8 of reviewers prior to submitting to a journal for formal peer-review. Although use of preprints is
9 still limited in the biological sciences, early indications are that preprints are a robust tool that can
10 complement and enhance peer-reviewed publications. As publishing moves to embrace advances
11 in internet technology, there are many opportunities for preprints and peer-reviewed journals to
12 coexist in the same ecosystem.

13 **Background.** A preprint is an interim research product that is made publicly available before
14 going through an official peer-review process with the goals of soliciting feedback, accelerating
15 dissemination of results, establishing priority, and publicizing negative results (1–5). Authors can
16 post their manuscript to a preprint server for others to read, share, and comment. In the 1960s,
17 Information Exchange Groups were the first formal attempt to broadly disseminate paper-based
18 preprints among physicists and biologists (6, 7). Although the biological community’s commitment
19 to preprints waned by 1967, the physics community persisted and eventually adopted what is
20 now the *arXiv* (pronounced “archive”) preprint server that was hosted at the Los Alamos National
21 Laboratories from 1991 to 1999 and then at Cornell University (8). For some physicists and
22 mathematicians, posting a preprint to *arXiv* optionally followed by submission to a peer-reviewed
23 journal has become a standard publication pathway. Although *arXiv* has hosted a number of
24 computational biology preprints, the server has not drawn widespread attention from biologists.
25 Among proponents of *arXiv*, preprints have aided in the development of research communication
26 by accelerating the release of science and helping authors reach a wider audience for critique
27 and establishment of priority (9). Considering the broadening adoption of preprints among
28 microbiologists, I sought to explore the specific uses of and concerns regarding preprints.

29 **Landscape of preprint servers.** In 2013, two preprint servers, the *bioRxiv* (pronounced
30 “bio-archive”) and *PeerJ Preprints*, were launched as preprint servers for biologists that would
31 parallel *arXiv* (10). According to information provided on the *bioRxiv* and *PeerJ Preprints* websites
32 and my personal experiences, both platforms offer similar features: preprint posting is free; each
33 preprint receives a digital object identifier (DOI) that facilitates the ability to cite preprints in other
34 scholarly work; if the preprint is ever published, the preprint is linked to the published version; the
35 submission process for both options is relatively simple allowing authors to upload a PDF version
36 of their preprint and supplemental materials; preprints are typically publicly available in about
37 24 hours; they have built-in venues for authors to discuss their research with people who leave
38 comments on the manuscript; preprints undergo a basic screening process to remove submissions
39 with offensive or non-scientific content; and the sites provide article-level metrics indicating the
40 number of times an abstract has been accessed or the preprint has been downloaded. There
41 are several important differences between the two options. First, *PeerJ Prints* is a for-profit

42 organization and *bioRxiv* is a non-profit organization sponsored by Cold Spring Harbor Laboratory.
43 This difference can be meaningful to authors since some journals, including the American Society
44 for Microbiology (ASM) Journals, will only accept submissions that have been posted on preprint
45 servers hosted by not-for-profit organizations (11). Second, preprints at *PeerJ Preprints* are posted
46 under the Creative Commons Attribution License (CC-BY) and *bioRxiv* preprints can be posted
47 under one of four CC-BY licenses or with no permission for reuse. This can be relevant for authors
48 hoping to submit their work to *Proceedings of the National Academy of Sciences* as the journal
49 will not consider manuscripts posted as preprints under a CC-BY license. The NIH encourages
50 authors to post preprints under the CC-BY or public domain licenses (5). The flexibility of the
51 *bioRxiv* licensing empowers authors to choose the model that best suits them, while ensuring
52 the rapid posting of their research results; however, it is important to provide clear information to
53 authors on the legal and practical tradeoffs of each option. A cosmetic, but still relevant difference
54 is the layout and feel of the two websites. Compared to the *bioRxiv* site (Figure S1), the *PeerJ*
55 *Preprint* site is more fluid, gives readers the ability to “follow” a preprint, and provides better access
56 to article keywords and the ability to search preprints (Figure S2). With broader acceptance of
57 preprints by traditional journals, many journals, including all of the ASM journals, have established
58 mechanisms to directly submit manuscripts that are posted as preprints on *bioRxiv*. The only direct
59 submission mechanism for manuscripts submitted at *PeerJ Preprint* is to the *PeerJ* journal. In
60 many ways, preprint servers have taken on the feel of a journal. As adoption of this approach
61 expands, it is likely that the features of these sites will continue to improve. It is also worth noting
62 that numerous other opportunities exist for other forms of interim research products (e.g. blog posts,
63 videos, protocols, etc.) to obtain DOIs that make the work citable. As these possibilities increase,
64 the preprint landscape risks becoming fractured.

65 One solution to the fracturing of the preprint landscape would be the creation of indexing sites that
66 allow a user to easily search for content across multiple preprint servers. Several examples of these
67 efforts already exist and it is likely that these interfaces and their ability to span the landscape will
68 improve. For example, although Google Scholar includes preprints hosted at *bioRxiv* and *PeerJ*
69 *Preprints* in their search results, PubMed and Web of Science do not. A relatively new example of
70 what this might look like is PrePubMed (12), which seeks to index preprints from numerous sources.

71 A more organized effort is being initiated with funding through ASAPbio to create a “Central Service”
72 that would aggregate preprints in the life sciences (13). As preprint servers and other content
73 providers begin to look and act like traditional journals by incorporating features and interfaces, it is
74 important to value the strength of the preprint - that of an interim research product that is nimble
75 and quickly posted. It is therefore essential to balance the requirements placed on authors for
76 features associated with preprints with the efficiency of the preprint format.

77 ***Specific challenges for microbiology.*** Although preprints offer an efficient and novel venue for
78 disseminating microbiology research, there are several considerations that the scientific community
79 and those that oversee preprint servers must consider. It is critical that assurances be given that
80 policies are in place to address these issues and that these policies are made transparent. First,
81 attention has to be given to dual use research of concern (DURC) since microbiology-related
82 research could offer insights to individuals seeking to engage in inappropriate activities. Second,
83 for researchers engaging in research that involves human subjects and other vertebrates, it is
84 critical that assurances be made that institutional oversight committees have been consulted
85 and have approved of the research. Third, there is significant concern regarding researchers
86 disclosing potential conflicts of interest that could affect a project’s experimental design, analysis,
87 and interpretation of results. Finally, recent expansions in scientific publishing have revealed
88 numerous cases of plagiarism or misconduct. Again, while hoping to maintain the efficiency of
89 the preprint format, traditional microbiology journals have screening procedures and oversight
90 committees that address these issues. Similar efforts need to be implemented by preprint servers.
91 As preprint usage continues to expand many of these problems may also grow similar to the
92 experiences within the traditional publishing industry has expanded.

93 ***Acceptance of preprints by journals.*** An early controversy encountered by researchers
94 interested in posting their work as preprints as a stage in disseminating their research was whether
95 it constituted prior publication (14). The broad consensus of the International Committee of Medical
96 Journal Editors and numerous journals is that preprints do not constitute prior publication (15).
97 This consensus is reflected in the current policies of journals that commonly publish microbiology
98 research including those published by ASM, the Microbiology Society, International Society for
99 Microbial Ecology, PLOS, the *Proceedings of the National Academy of Science*, *Science*, *Nature*,

100 *Journal of Infectious Diseases*, and Cell press. Each take a generally permissive stance towards
101 posting of preprints prior to submission. Comprehensive lists of journals' attitudes towards preprints
102 are available online and are regularly updated (16, 17). Considering the relatively fluid nature of
103 many of these policies and the journals' specific policies, prospective authors should be aware of
104 the positions taken by the journals where they may eventually submit their work.

105 ***Preprints and peer-review.*** The use of preprints for citations in other scientific reports and grant
106 proposals has recently been called into question (18). It is important to note that the peer-review
107 process was adapted to the technologies and trends that have evolved over the past 100 years.
108 The formal peer-review system that most journals currently use was not developed until the end of
109 the 1800s with the advent of typewriters and carbon paper (19). Editorial decisions were typically
110 made by a single person or a committee (i.e. the editorial board) who had an expertise that covered
111 the scope of the journal. As science became more specialized, new journals would form to support
112 and provide a source of validation to the new specialty. The growth in science in the mid 1900s
113 resulted in a shift from journals struggling to find sufficient numbers of manuscript to publish to
114 having too many manuscripts submitted. It has been argued that the widespread adoption of
115 decentralized peer-review was due to the increased specialization and to deal with the large number
116 of manuscript submissions (20). Peer-review did not achieve widespread use at many journals,
117 including the *Journal of Bacteriology*, until the 1940s and 1950s. Thus the "tradition" of peer-review
118 is only 70 years old. Given the rapid advances in communication technology and even greater
119 specialization within microbiology, it is worth pondering whether the current scientific publishing
120 system and peer-review system, in particular, need to continue to adapt with our science.

121 Communicating research has traditionally been done within research group meetings, departmental
122 seminars, conferences, and as publications. Along this continuum, there is an assumption that
123 the quality of the science has been improved because it has been vetted by more experts in the
124 field. The public dissemination of one's research is a critical component of the scientific method. By
125 describing their research, scientists subject their work to formal and informal peer-review. Their
126 research is scrutinized, praised, and probed to identify questions that help seed the next iteration
127 of the scientific method. A common critique of more modern approaches to publishing has been
128 an inability to assess the quality of the science without the validation of peer-review. Attached

129 to assertions of the validity of the research has been assertions of the impact and robustness of
130 the research. These are all quality assessments that many acknowledge are difficult to assess
131 by the traditional peer-review process. This has led to some journals, most notably *PLOS ONE*,
132 calling for referees to place a reduced emphasis on the perceived impact or significance of the
133 work. It has also led to the call for replacing or complementing pre-publication peer-review with
134 post-publication peer-review using PubMed Commons, PubPeer, journal-based discussion forums,
135 F1000Research, and other mechanisms. Alas if scientists are going to depend on post-publication
136 peer-review or informal methods of peer-review for documents like preprints, they must be willing to
137 provide constructive feedback on the work of others.

138 ***Preprints have the potential to change the advancement of science.*** Preprints are often
139 viewed as existing in a state of scientific limbo. As noted above, they represent a formal
140 communication, but an interim one, not officially published. As the use of preprints grows and
141 scientists' perceptions of preprints matures, there are a number of issues that will need to be
142 addressed.

143 First, a common concern is that if a researcher posts their work as a preprint, it will be “scooped”
144 by another researcher and the preprint author will lose their ability to claim primacy or their ability
145 to publish the work in a journal. Considering the preprint is a citable work with a DOI, it would, in
146 fact, be the preprint author that scooped the second. Furthermore, a preprint could prevent getting
147 scooped since a preprint would indicate to others in the field that the work had already been done,
148 which would prevent wasted time and effort. The use of preprints uncouples the communication
149 of the discovery from the relevance of the discovery, which will come later based on peer-review,
150 comments from other scientists at meetings or online, and eventually citations. A growing number of
151 scientific societies and journals, including ASM view preprints as citable and as having a legitimate
152 claim to primacy (1, 21–23); however, it remains to be determined whether the journals will stand
153 by these policies. Some scientists worry that with such protection a researcher can make a claim
154 without valid data to support their claims (3). This is possible; however, it is also the responsibility
155 of the scientific community to utilize the peer-review mechanisms that are available to comment on
156 those preprints pointing out methodological problems or to indicate that they are speaking beyond
157 the data. As preprints gain broader adoption, the tension between establishing primacy and the

158 completeness of the preprint may test the policies of preprint-friendly journals.

159 A second area of concern is whether a preprint can be used to support a grant proposal. Given the
160 length limitations placed on grant proposals by funding agencies, there is a push to cite previous
161 work to indicate a research team's competence in an area or to provide preliminary data. The
162 National Institutes of Health (NIH) recently released a notice clarifying their position on the use
163 of preprints and synthesizing feedback that they received as part of a request for information (5).
164 In this notice, the NIH indicated that preprints can be cited anywhere that other research is cited
165 including research plans, bibliographies, biosketches, and progress reports. Some fear that the
166 use of preprints will allow scientists to circumvent page limits by posting preliminary manuscripts
167 (24). One would hope that both consumers of preprints and grant proposal reviewers would be able
168 to differentiate between someone trying to game the system and someone that is using preprints
169 as a mechanism to improve their science. This would be greatly facilitated by following the NIH
170 recommendation of using preprints as evidence for research progress, but providing an indication
171 that the preprints are not peer-reviewed publications (5). This would help review panels in rendering
172 their decisions and help authors substantiate their preliminary data.

173 A third concern is what role preprints should have in assessing a scientist's productivity. Clearly
174 use of publication metrics as an indicator of a scientist's productivity and impact is a contentious
175 topic without even discussing the role of preprints. Regardless, given the propensity for researchers
176 to list manuscripts as being "in preparation" or "in review" on an application or curriculum vitae,
177 listing them instead as preprints that can be reviewed by a committee would significantly enhance
178 an application and a reviewer's ability to judge the application. In fact, several funding agencies
179 including the NIH, Wellcome Trust, UK Medical Research Council encouraging fellowship applicants
180 to include preprints in their materials (5).

181 Beyond these concerns, preprints are also causing some to change their publication goals. Some
182 authors are explicitly stating that a preprint will not be submitted to a journal (25). Although these
183 authors may be a minority of those who post preprints, such an approach may be attractive to those
184 who need to cite a report of a brief research communication, a critique of another publication, or
185 negative results. It is clear that the adoption of preprints will challenge how scientists interact and

186 evaluate each other's work. There is great potential to empower researchers by controlling when a
187 citable piece of work is made public.

188 **Microbiology anecdotes.** The peer-review editorial process can be lengthy and adversarial.
189 Because preprints are public and freely available they represent a rapid and potentially collaborative
190 method for disseminating research. Several anecdotes from the microbiology literature are
191 emblematic of benefits of the rapid release cycle that is inherent in the use of preprints.

192 First, preprints have proven useful for rapidly disseminating results for disease outbreaks and new
193 technologies. Prior to the recent Zika virus outbreak there were approximately 50 peer-reviewed
194 publications that touched on the biology and epidemiology of the virus; as of April 2017 the number
195 of Zika virus-related peer-reviewed publications was over 2,300. During the recent outbreak, more
196 than 150 Zika virus-related preprints have been posted at *bioRxiv*. Any manuscript that was formally
197 published went through several month delays in releasing information to health care workers, the
198 public, and scientists needing to learn new methods to study a previously obscure virus. In contrast,
199 those that posted their work as a preprint were able to disseminate their methods and results
200 instantly. Another interesting use of preprints to disseminate new information about Zika virus has
201 been the posting of a preprint describing the Zika virus outbreak in the US Virgin Islands that will be
202 continually updated as new data and analyses are performed (26). Over the last several years there
203 have also been rapid advances in DNA sequencing technologies that have fundamentally changed
204 how microbial science is performed. One notable technology, the Minlon sequencing platform from
205 Oxford Nanopore, has received considerable attention from researchers who have posted more
206 than 110 preprints describing new Minlon-based methods and results to preprint servers. For such
207 a rapidly developing technology, the ability to share and consume methods from other scientists
208 has created a feed forward effect where the technology has likely advanced at a faster rate than it
209 otherwise would have.

210 Second, preprints have proven useful for rapidly correcting the scientific literature. On February
211 9, 2015, *Cell Systems* published a study that collected and analyzed metagenomic sequence
212 data from the New York City subway system and reported finding *Yersinia pestis* and *Bacillus*
213 *anthracis* (27). Because of the focus on these two bioterrorism agents, this study generated a

214 considerable amount of media attention. On April 25, 2015, Petit et al. (28) posted a preprint to
215 Zenodo demonstrating that there was no evidence for *B. anthracis* in the dataset. On July 29,
216 2015, a critique was published by *Cell Systems* along with a response from the original authors
217 offering a correction to their manuscript (29, 30). A second anecdote of using preprints to aid
218 in post-publication peer-review surrounds the publishing of a draft tardigrade genome in *The*
219 *Proceedings of the National Academy of Sciences*. On November 23, 2015 a study by Boothby et
220 al. (31) was published online. The authors claimed that 17.5% of its genes came from bacteria,
221 archaea, fungi, plants, and viruses. Another group had been analyzing sequence data from a
222 parallel tardigrade genome sequencing project and did not observe the same result. A week later,
223 on December 1, 2015, the second group posted a preprint comparing the two genome sequences
224 and demonstrating that the exciting claims of horizontal gene transfer were really the product of
225 contaminants (32); this analysis would eventually be peer-reviewed and published online by the
226 original journal on March 24, 2016 followed by a rebuttal by the original authors on May 31, 2016
227 (33, 34). Two other analyses of the original data were peer-reviewed and published in May 2016
228 and a third was posted as a preprint on February 2, 2016 (35–37). Both of these anecdotes
229 underscore the value of having a rapid posting cycle to correcting errors in the scientific literature
230 and that results posted to preprint servers were able to correct the record within weeks of the initial
231 publication while the traditional peer-review path took six months in both cases. A final notable case
232 where preprints have accelerated the correction of the scientific record was a preprint posted by Bik
233 et al. (38) reporting numerous cases of image manipulation in peer-reviewed studies. This was a
234 case where a journal may have been reluctant to publish the findings because it could have put the
235 journal in a bad light. Posting the manuscript as a preprint removed potential conflicts of interests
236 from journals that could have hindered its ability to be formally published in a journal. After the
237 preprint was posted on April 20, 2016 it was peer-reviewed and published in *mBio* on June 7, 2016
238 (39). Instead of using preprints to react to published papers that have been through peer-review, it
239 would be interesting to consider how the editorial process for these examples and the infamous
240 “Arsenic Life” paper (40) would have been different had they initially been posted as preprints.

241 ***Metrics for microbiology-affiliated preprints.*** To analyze the use of preprints, I downloaded
242 the *bioRxiv* on April 17, 2017. I chose to analyze *bioRxiv* preprints because these preprints are

243 amenable for submission to ASM journals and there were 9,780 *bioRxiv* preprints compared to the
244 2,911 preprints that were available at *PeerJ Preprint* on the same date. The code used to analyze
245 these preprints and the rest of this manuscript are available as a reproducible GitHub repository at
246 http://www.github.com/SchlossLab/Schloss_PrePrints_mBio_2017. Among the 9,780 preprints on
247 bioRxiv, 483 were assigned by the authors into the Microbiology category. One limitation of the
248 *bioRxiv* interface is the inability to assign manuscripts to multiple categories or to tag the content
249 of the preprint. For example, this manuscript could be assigned to either the Microbiology or the
250 Scientific Communication and Education categories. To counter this limitation, I developed a more
251 permissive approach that classified preprints as being microbiology-affiliated if their title or abstract
252 had words containing *yeast*, *fung*, *viral*, *virus*, *archaea*, *bacteri*, *microb*, *microorganism*, *pathogen*,
253 or *protist*. I identified 1,617 additional manuscripts that I considered microbiology-affiliated. These
254 microbiology-affiliated preprints were primarily assigned to the Evolutionary Biology (N=283),
255 Bioinformatics (N=237), or Genomics (N=231) categories.

256 As the total number of preprints has grown exponentially since the creation of *bioRxiv*, submission
257 of microbiology-affiliated preprints has largely followed this growth (**Figure 1A**). Although preprints
258 are still relatively new, the collection of microbiology-affiliated preprints indicates widespread
259 experimentation with the format and considerable geographic diversity. Reflecting the relative
260 novelty of preprints, 1,484 (85.5%) corresponding authors who submitted a microbiology-affiliated
261 preprint (N=1,735 total) have posted a single preprint and 4.6% have posted 3 or more preprints.
262 Corresponding authors that have posted microbiology-affiliated preprints are from 67 countries
263 and are primarily affiliated with institutions in the United States (46.2% of microbiology-affiliated
264 preprints), United Kingdom (12.9%), and Germany (4.6%). As the preprint format matures, it will be
265 interesting to see whether the fraction of authors that post multiple preprints increases and whether
266 the geographic diversity amongst those authors is maintained.

267 As stated above, preprints offer researchers the opportunity to improve the quality of their work by
268 adding a more formal and public step to the scientific process. Among the microbiology-affiliated
269 preprints, 197 (9.3%) had been commented on at least once and only 48 (2.3%) more than three
270 times using the *bioRxiv*-hosted commenting feature. Although the hosted commenting is only
271 one mechanism for peer-review, this result was somewhat disturbing since the preprint model

272 implicitly depends on people's willingness to offer others feedback. In spite of the lack of tradition
273 within the scientific community to comment publicly online about colleagues' research results, I
274 am optimistic that this will change given the possibilities of new media (e.g. Twitter, Facebook,
275 blogs); the advantage of the centralized commenting is that it is easier for the authors and others
276 to integrate the feedback with the preprint. It is possible that incentives for open commenting
277 and reviewing could shift the trend. Importantly, authors do appear to be incorporating feedback
278 from colleagues or editorial insights from journals as 545 (25.9%) microbiology-affiliated preprints
279 were revised at least once. Among the preprints posted prior to January 1, 2016, 31.3% of the
280 Microbiology category preprints, 35.6% of the microbiology-affiliated preprints, and 33.6% of all
281 preprints have been published. As noted above, not all authors submit their preprints to journals.
282 This would indicate that the "acceptance rates" are actually higher. Regardless, considering that
283 these acceptance rates are higher than many peer-reviewed journals (e.g. approximately 20% at
284 ASM Journals), these results dispel the critique that preprints represent overly preliminary research.

285 Measuring the impact and significance of scientific research is notoriously difficult. Using several
286 metrics I sought to quantify the effect that broadly defined microbiology-affiliated preprints have
287 had on the work of others. Using the download statistics associated with each preprint, I found
288 that the median number of times an abstract or PDF had been accessed was 922 (IQR: 601 to
289 1446) and 301 (IQR: 155 to 549), respectively. These values represent two aspects of posting a
290 preprint. First, they reflect the number of times people were able to access science before it was
291 published. Second, they reflect the number of times people were able to access a version of a
292 manuscript that is published behind a paywall. To obtain a measure of a preprint's ability to garner
293 attention and engage the general public, I obtained the Altmetric Attention Score for each preprint
294 (**Figure 1B**). The Altmetric Attention Score measures the number of times a preprint or paper is
295 mentioned in social media, traditional media, Wikipedia, policy documents, and other sources; it
296 does not include the number of citations (41). A higher score indicates that a preprint received
297 more attention. Microbiology-affiliated preprints have had a median Altmetric Attention Score of
298 7.6 (IQR: 3.2 to 16.6) and those of all preprints hosted at *bioRxiv* have had a median score of 7.7
299 (IQR: 3.1 to 16.2). For comparison, the median Altmetric Attention Score for articles published in
300 *mBio* published since 2013 was 5.0 (IQR: 1.5 to 14.5). Of all scholarship tracked by Altmetric, the

301 median Altmetric Attention Score for preprints posted at *bioRxiv* ranks at the 87 percentile (IQR: 75
302 to 94). A controversial, yet more traditional metric of impact has been the number of citations an
303 article receives. I obtained the number of citations for the published versions of manuscripts that
304 were initially posted as preprints. To allow for a comparison to traditional journals, I considered the
305 citations for preprints published in 2014 and 2015 as aggregated by Web of Science (**Figure 1C**).
306 Among the preprints that were published and could be found in the Web of Science database, the
307 median number of citations was 9 (IQR: 3-19; mean: 17.1). For comparison, among the papers
308 published in *mBio* in 2014 and 2015, the median number of citations was 6 (IQR: 3-11; mean: 8.5).
309 Although it is impossible to quantify the quality or impact of research with individual metrics, it is
310 clear that the science presented in preprints and the publications that result from them are accepted
311 by the microbiology community at a level comparable to more traditionally presented research.

312 ***Preprints from an author's perspective.*** Posting research as a preprint gives an author great
313 control over when their work is made public. Under the traditional peer-review model, an author
314 may need to submit and revise their work multiple times to several journals over a long period
315 before it is finally published. In contrast, an author can post the preprint at the start of the process
316 for others to consume and comment on as it works its way through the peer-review process. A first
317 example illustrates the utility of preprints for improving access to research and the quality of its
318 reporting. In 2014, my research group posted a preprint to *PeerJ Preprints* describing a method
319 of sequencing 16S rRNA gene sequences using the Pacific Biosciences sequencing platform
320 (42). At the same time, we submitted the manuscript for review at *PeerJ*. While the manuscript
321 was under review, we received feedback from an academic scientist and from scientists at Pacific
322 Biosciences that the impact of the results could be enhanced by using a recently released version
323 of the sequencing chemistry. Instead of ignoring this feedback and resubmitting the manuscript to
324 address the reviews, we generated new data and submitted an updated preprint a year later with a
325 simultaneous submission to *PeerJ* that incorporated the original reviews as well as the feedback we
326 received from the academic scientist and Pacific Biosciences. It was eventually published by *PeerJ*
327 (43, 44). Since 2015, we have continued to post manuscripts as preprints at the same time as we
328 have submitted manuscripts. Although the feedback to other manuscripts has not always been as
329 helpful as our initial experience, in each case we were able to publicize our results prior to lengthy

330 peer-review processes by immediately making our results available; in one case our preprint was
331 available 7 months ahead of the final published version (45, 46). As another example, I posted a
332 preprint of the current manuscript to *bioRxiv* on February 22, 2017. I then solicited feedback on
333 the manuscript using social media. On March 14, 2017 I incorporated the comments and posted a
334 revised preprint and submitted the manuscript to *mBio*. During that time, the abstract was accessed
335 189 times and the PDF was accessed 107 times. This process engaged 3 commenters on *bioRxiv*,
336 61 people either tweeted or re-tweeted the preprint on Twitter, 2 people on the manuscript's GitHub
337 repository, 1 person on a blog, and 2 via email. Compared to the two scientists that eventually
338 reviewed the manuscript, the preliminary round of informal peer-review engaged a much larger and
339 more diverse community than had I foregone the posting of a preprint. By the time that the final
340 version of the manuscript was submitted on April 21, 2017, the preprint version of this manuscript
341 had an Attention Score of 58, which placed it in the top 5% of all research scored by Altmetric and
342 the abstract and PDF had been accessed 2,152 and 512 times, respectively. Although there are
343 concerns regarding the quality of the science posted to a preprint server, I contend that responsible
344 use of preprints as a part of the scientific process can significantly enhance the science.

345 ***Preprints from a publisher's perspective.*** A lingering question is what role traditional journals
346 will have in disseminating research if there is broad adoption of preprints. Edited peer-reviewed
347 journals offer and will continue to offer significant added value to a publication. A scholarly
348 publishing ecosystem in which preprints coexist with journals will allow authors to gain value from
349 the immediate communication of their work associated with preprints and also benefit from the
350 peer-reviewed, professionally edited publication that publishers can provide.

351 The professional copyediting, layout, and publicity that these publishers offer are also unique
352 features of traditional journals. An alternative perspective is that preprints will eventually replace
353 traditional journals. Certainly, this is a radical perspective, but it does serve to motivate publishers
354 to capture the innovation opportunities offered by preprints. By adopting preprint-friendly policies,
355 journals can create an attractive environment for authors. As discussed above, a growing number
356 of journals have created mechanisms for authors to directly submit preprints to their journals. An
357 example is offered by the ASM, which earlier this year launched a new venture from *mSphere*.
358 mSphereDirect is a publication track of the journal that capitalizes on the opportunity offered to

359 couple preprints with rigorous peer-review. mSphereDirect actively encourages authors to post
360 their manuscripts as preprints as part of an author-driven editorial process where an editorial
361 decision is rendered within five days and publication in *mSphere* within a month (47). As the
362 mSphereDirect mechanism evolves and is perhaps adopted by other journals, it will be interesting
363 to see whether public feedback on preprints will be used to further streamline the editorial process.
364 ASM is developing a new platform, MicroNow, which will help coalesce specific communities within
365 the microbial sciences, further enhancing the use of preprints as well as published articles (Stefano
366 Bertuzzi, personal communication). In addition to integrating preprints into the traditional editorial
367 process, several professional societies have also explicitly supported citation of preprints in their
368 other publications and recognize the priority of preprints in the literature (21–23). These are policies
369 that empower authors and make specific journals more attractive. Other practices have great
370 potential to improve the reputation of journals. As measured above, preprints are able to garner
371 attention on par with papers published in highly selective microbiology journals. Thus, it is in a
372 journal's best interest to recruit these preprints to their journals. Several journals including *PLOS*
373 *Genetics* and *Genome Biology* have publicly stated that they scout preprints for this purpose (48,
374 49). Preprints can also be viewed as a lost opportunity to journals. A preprint that garners significant
375 attention may be ignored when it is finally published, bringing little additional attention to the journal.
376 Going forward, there will likely be many innovative approaches that publishers develop to benefit
377 from incorporating preprints into their process and whether publishers' influence is reduced by the
378 widespread adoption of preprints.

379 **Conclusions.** Since the first microbiology-affiliated preprint was posted on bioRxiv in November
380 2013 (50), an increasing number of microbiologists are posting their unpublished work to preprint
381 servers as an efficient method for disseminating their research prior to peer-review. A number of
382 critical concerns remain about how widespread their adoption will be, how they will be perceived by
383 traditional journals and other scientists, and whether traditional peer-review will adapt to the new
384 scientific trends and technologies. Regardless, preprints should offer a great opportunity for both
385 scientists and journals to publish high quality science.

386 **Acknowledgements**

387 I am grateful to Stefano Bertuzzi and Lynn Enquist for their helpful comments on earlier versions of
388 this manuscript and to the numerous individuals who provided feedback on the preprint version
389 of the manuscript. This work was supported in part by funding from the National Institutes of
390 Health (P30DK034933). I appreciate the support of Altmetric, Inc and Thompson Reuters who
391 provided advanced programming interface (API) access to their databases. The workflow utilized
392 commands in GNU make (v.3.81), GNU bash (v.4.1.2), and R (v.3.3.3). Within R I utilized the
393 cowplot (v.0.7.0), dplyr (v.0.5.0), ggplot2 (v.2.2.1), httr (v.1.2.1), RCurl (v.1.95-4.8), rentrez (v.1.0.4),
394 RJSONIO (v.1.3-0), rvest (v.0.3.2), sportcolors (v.0.0.1), and tidyr (v.0.6.1) packages. All journal
395 policies and the information cited using webpage links was current on April 21, 2017.

396 **Figure 1. Summary of microbiology-affiliated preprints since the creation of *bioRxiv*.** The
397 total number of preprints posted for each quarter ending March 31, 2017 has largely tracked the
398 overall submission of preprints to *bioRxiv* (A). The Altmetric attention scores of preprints posted to
399 *bioRxiv* are similar to those published in *mBio* since November 2013 indicating preprints engender
400 a similar level of attention (B). The number of times preprints that were published in 2014 and 2015
401 have been cited is similar to the number of citations for papers published in *mBio* in 2014 and
402 2015 indicates that published preprints are frequently cited (C). Regions with common background
403 shading in A are from the same year. The vertical lines in B and C indicate the median Altmetric
404 impact score and the median number of citations.

405 **Supplemental Figure 1. Screen shot of the preprint for this manuscript at bioRxiv.**

406 **Supplemental Figure 2. Screen shot of a preprint by the author hosted at PeerJ Preprints.**

407 **References**

- 408 1. **Vale RD**. 2015. Accelerating scientific publication in biology. *Proceedings of the National*
409 *Academy of Sciences* **112**:13439–13446. doi:[10.1073/pnas.1511912112](https://doi.org/10.1073/pnas.1511912112).
- 410 2. **Desjardins-Proulx P, White EP, Adamson JJ, Ram K, Poisot T, Gravel D**. 2013. The case
411 for open preprints in biology. *PLoS Biology* **11**:e1001563. doi:[10.1371/journal.pbio.1001563](https://doi.org/10.1371/journal.pbio.1001563).
- 412 3. **Berg JM, Bhalla N, Bourne PE, Chalfie M, Drubin DG, Fraser JS, Greider CW, Hendricks**
413 **M, Jones C, Kiley R, King S, Kirschner MW, Krumholz HM, Lehmann R, Leptin M, Pulverer**
414 **B, Rosenzweig B, Spiro JE, Stebbins M, Strasser C, Swaminathan S, Turner P, Vale RD,**
415 **VijayRaghavan K, Wolberger C**. 2016. Preprints for the life sciences. *Science* **352**:899–901.
416 doi:[10.1126/science.aaf9133](https://doi.org/10.1126/science.aaf9133).
- 417 4. **Bhalla N**. 2016. Has the time come for preprints in biology? *Molecular Biology of the Cell*
418 **27**:1185–1187. doi:[10.1091/mbc.e16-02-0123](https://doi.org/10.1091/mbc.e16-02-0123).
- 419 5. Reporting preprints and other interim research products. [https://grants.nih.gov/grants/guide/](https://grants.nih.gov/grants/guide/notice-files/NOT-OD-17-050.html)
420 [notice-files/NOT-OD-17-050.html](https://grants.nih.gov/grants/guide/notice-files/NOT-OD-17-050.html).
- 421 6. 1966. Preprints galore. *Nature* **211**:897–898. doi:[10.1038/211897a0](https://doi.org/10.1038/211897a0).
- 422 7. **Till JE**. 2001. Predecessors of preprint servers. *Learned Publishing* **14**:7–13. doi:[10.1087/09531510125100214](https://doi.org/10.1087/09531510125100214).
- 423 8. **Ginsparg P**. 2011. ArXiv at 20. *Nature* **476**:145–147. doi:[10.1038/476145a](https://doi.org/10.1038/476145a).
- 424 9. **Shuai X, Pepe A, Bollen J**. 2012. How the scientific community reacts to newly
425 submitted preprints: Article downloads, twitter mentions, and citations. *PLoS ONE* **7**:e47523.
426 doi:[10.1371/journal.pone.0047523](https://doi.org/10.1371/journal.pone.0047523).
- 427 10. **Callaway E**. 2013. Biomedical journal and publisher hope to bring preprints to life. *Nature*

- 428 Medicine **19**:512–512. doi:[10.1038/nm0513-512](https://doi.org/10.1038/nm0513-512).
- 429 11. mBio instructions to authors. http://mbio.asm.org/site/misc/journal-ita_edi.xhtml.
- 430 12. PrePubMed. <http://www.pubmed.org>.
- 431 13. ASAPbio awarded \$1 million from Helmsley Charitable Trust for next-generation life sciences
432 preprint infrastructure. <http://asapbio.org/feb23>.
- 433 14. **Angell M, Kassirer JP**. 1991. The ingelfinger rule revisited. *New England Journal of Medicine*
434 **325**:1371–1373. doi:[10.1056/nejm199111073251910](https://doi.org/10.1056/nejm199111073251910).
- 435 15. Recommendations for the conduct, reporting, editing, and publication of scholarly work in
436 medical journals. [http://www.icmje.org/news-and-editorials/icmje-recommendations_annotated_](http://www.icmje.org/news-and-editorials/icmje-recommendations_annotated_dec16.pdf)
437 [dec16.pdf](http://www.icmje.org/news-and-editorials/icmje-recommendations_annotated_dec16.pdf).
- 438 16. List of academic journals by preprint policy. [https://en.wikipedia.org/wiki/List_of_academic_](https://en.wikipedia.org/wiki/List_of_academic_journals_by_preprint_policy)
439 [journals_by_preprint_policy](https://en.wikipedia.org/wiki/List_of_academic_journals_by_preprint_policy).
- 440 17. Publisher copyright policies & self-archiving. <http://www.sherpa.ac.uk/romeo/index.php>.
- 441 18. **Drubin DG**. 2016. The mismeasure of scientific research articles and why MBoC quickly
442 embraced preprints. *Molecular Biology of the Cell* **27**:3181–3182. doi:[10.1091/mbc.e16-09-0651](https://doi.org/10.1091/mbc.e16-09-0651).
- 443 19. **Spier R**. 2002. The history of the peer-review process. *Trends in Biotechnology* **20**:357–358.
444 doi:[10.1016/s0167-7799\(02\)01985-6](https://doi.org/10.1016/s0167-7799(02)01985-6).
- 445 20. **Burnham JC**. 1990. The evolution of editorial peer review. *JAMA: The Journal of the American*
446 *Medical Association* **263**:1323. doi:[10.1001/jama.1990.03440100023003](https://doi.org/10.1001/jama.1990.03440100023003).
- 447 21. **Pulverer B**. 2016. Preparing for preprints. *The EMBO Journal* **35**:2617–2619.
448 doi:[10.15252/emboj.201670030](https://doi.org/10.15252/emboj.201670030).
- 449 22. **Loew LM**. 2016. Peer review and bioRxiv. *Biophysical Journal* **111**:E01–E02.

450 doi:[10.1016/j.bpj.2016.06.035](https://doi.org/10.1016/j.bpj.2016.06.035).

451 23. **Vale RD, Hyman AA.** 2016. Priority of discovery in the life sciences. *eLife* **5**.
452 doi:[10.7554/elife.16931](https://doi.org/10.7554/elife.16931).

453 24. Interim research product rfi response from faseb. [http://www.faseb.org/Portals/2/PDFs/opa/](http://www.faseb.org/Portals/2/PDFs/opa/2016/Interim%20Research%20Product%20RFI.pdf)
454 [2016/Interim%20Research%20Product%20RFI.pdf](http://www.faseb.org/Portals/2/PDFs/opa/2016/Interim%20Research%20Product%20RFI.pdf).

455 25. **Chawla DS.** 2017. When a preprint becomes the final paper. *Nature*. doi:[10.1038/nature.2017.21333](https://doi.org/10.1038/nature.2017.21333).

456 26. **Black A, Potter B, Dudas G, Feldstein L, Grubaugh ND, Andersen KG, Ellis BR, Ellis**
457 **EM, Bedford T.** 2017. Genetic characterization of the zika virus epidemic in the us virgin islands.
458 bioRxiv 113100.

459 27. **Afshinnekoo E, Meydan C, Chowdhury S, Jaroudi D, Boyer C, Bernstein N, Maritz JM,**
460 **Reeves D, Gandara J, Chhangawala S, Ahsanuddin S, Simmons A, Nessel T, Sundaresh**
461 **B, Pereira E, Jorgensen E, Kolokotronis S-O, Kirchberger N, Garcia I, Gandara D, Dhanraj**
462 **S, Nawrin T, Saletore Y, Alexander N, Vijay P, Hénaff EM, Zumbo P, Walsh M, O'Mullan**
463 **GD, Tighe S, Dudley JT, Dunaif A, Ennis S, O'Halloran E, Magalhaes TR, Boone B, Jones**
464 **AL, Muth TR, Paolantonio KS, Alter E, Schadt EE, Garbarino J, Prill RJ, Carlton JM, Levy**
465 **S, Mason CE.** 2015. Geospatial resolution of human and bacterial diversity with city-scale
466 metagenomics. *Cell Systems* **1**:72–87. doi:[10.1016/j.cels.2015.01.001](https://doi.org/10.1016/j.cels.2015.01.001).

467 28. **Petit III RA, Ezewudo M, Joseph SJ, Read TD.** 2015. Searching for anthrax in the New York
468 City subway metagenome. doi:[10.5281/zenodo.17158](https://doi.org/10.5281/zenodo.17158).

469 29. **Ackelsberg J, Rakeman J, Hughes S, Petersen J, Mead P, Schriefer M, Kingry L,**
470 **Hoffmaster A, Gee JE.** 2015. Lack of evidence for plague or anthrax on the new york city subway.
471 *Cell Systems* **1**:4–5. doi:[10.1016/j.cels.2015.07.008](https://doi.org/10.1016/j.cels.2015.07.008).

472 30. **Afshinnekoo E, Meydan C, Chowdhury S, Jaroudi D, Boyer C, Bernstein N, Maritz JM,**
473 **Reeves D, Gandara J, Chhangawala S, Ahsanuddin S, Simmons A, Nessel T, Sundaresh B,**
474 **Pereira E, Jorgensen E, Kolokotronis S-O, Kirchberger N, Garcia I, Gandara D, Dhanraj S,**
475 **Nawrin T, Saletore Y, Alexander N, Vijay P, Hénaff EM, Zumbo P, Walsh M, O'Mullan GD,**

- 476 **Tighe S, Dudley JT, Dunaif A, Ennis S, O’Halloran E, Magalhaes TR, Boone B, Jones AL,**
477 **Muth TR, Paolantonio KS, Alter E, Schadt EE, Garbarino J, Prill RJ, Carlton JM, Levy S,**
478 **Mason CE.** 2015. Modern methods for delineating metagenomic complexity. *Cell Systems* 1:6–7.
479 doi:[10.1016/j.cels.2015.07.007](https://doi.org/10.1016/j.cels.2015.07.007).
- 480 31. **Boothby TC, Tenlen JR, Smith FW, Wang JR, Patanella KA, Nishimura EO, Tintori SC, Li**
481 **Q, Jones CD, Yandell M, Messina DN, Glasscock J, Goldstein B.** 2015. Evidence for extensive
482 horizontal gene transfer from the draft genome of a tardigrade. *Proceedings of the National*
483 *Academy of Sciences* 112:15976–15981. doi:[10.1073/pnas.1510461112](https://doi.org/10.1073/pnas.1510461112).
- 484 32. **Koutsovoulos G, Kumar S, Laetsch DR, Stevens L, Daub J, Conlon C, Maroon H, Thomas**
485 **F, Aboobaker A, Blaxter M.** 2016. No evidence for extensive horizontal gene transfer in the
486 genome of the tardigrade *hypsibius dujardini*. *bioRxiv* 033464. doi:[10.1101/033464](https://doi.org/10.1101/033464).
- 487 33. **Boothby TC, Goldstein B.** 2016. Reply to bemm et al. and arakawa: Identifying foreign genes
488 in independent *Hypsibius dujardini* genome assemblies. *Proceedings of the National Academy of*
489 *Sciences* 113:E3058–E3061. doi:[10.1073/pnas.1601149113](https://doi.org/10.1073/pnas.1601149113).
- 490 34. **Koutsovoulos G, Kumar S, Laetsch DR, Stevens L, Daub J, Conlon C, Maroon H, Thomas**
491 **F, Aboobaker AA, Blaxter M.** 2016. No evidence for extensive horizontal gene transfer in the
492 genome of the tardigrade *Hypsibius dujardini*. *Proceedings of the National Academy of Sciences*
493 113:5053–5058. doi:[10.1073/pnas.1600338113](https://doi.org/10.1073/pnas.1600338113).
- 494 35. **Bemm F, Weiß CL, Schultz J, Förster F.** 2016. Genome of a tardigrade: Horizontal
495 gene transfer or bacterial contamination? *Proceedings of the National Academy of Sciences*
496 113:E3054–E3056. doi:[10.1073/pnas.1525116113](https://doi.org/10.1073/pnas.1525116113).
- 497 36. **Arakawa K.** 2016. No evidence for extensive horizontal gene transfer from the draft
498 genome of a tardigrade. *Proceedings of the National Academy of Sciences* 113:E3057–E3057.
499 doi:[10.1073/pnas.1602711113](https://doi.org/10.1073/pnas.1602711113).
- 500 37. **Delmont TO, Eren AM.** 2016. Identifying contamination with advanced visualization
501 and analysis practices: Metagenomic approaches for eukaryotic genome assemblies.

502 [doi:10.7287/peerj.preprints.1695v1](https://doi.org/10.7287/peerj.preprints.1695v1).

503 38. **Bik EM, Casadevall A, Fang FC**. 2016. The prevalence of inappropriate image duplication in
504 biomedical research publications. bioRxiv 049452. doi:[10.1101/049452](https://doi.org/10.1101/049452).

505 39. **Bik EM, Casadevall A, Fang FC**. 2016. The prevalence of inappropriate image duplication in
506 biomedical research publications. mBio 7:e00809–16. doi:[10.1128/mbio.00809-16](https://doi.org/10.1128/mbio.00809-16).

507 40. **Wolfe-Simon F, Blum JS, Kulp TR, Gordon GW, Hoefft SE, Pett-Ridge J, Stolz JF, Webb
508 SM, Weber PK, Davies PCW, Anbar AD, Oremland RS**. 2010. A bacterium that can grow by
509 using arsenic instead of phosphorus. Science 332:1163–1166. doi:[10.1126/science.1197258](https://doi.org/10.1126/science.1197258).

510 41. How is the Altmetric Attention Score calculated? [https://help.altmetric.com/support/solutions/
511 articles/6000060969-how-is-the-altmetric-attention-score-calculated-](https://help.altmetric.com/support/solutions/articles/6000060969-how-is-the-altmetric-attention-score-calculated-).

512 42. **Schloss PD, Westcott SL, Jenior ML, Highlander SK**. 2015. Sequencing 16S rRNA gene
513 fragments using the PacBio SMRT DNA sequencing system. doi:[10.7287/peerj.preprints.778v1](https://doi.org/10.7287/peerj.preprints.778v1).

514 43. **Schloss PD, Jenior ML, Koumpouras CC, Westcott SL, Highlander SK**. 2016.
515 Sequencing 16S rRNA gene fragments using the PacBio SMRT DNA sequencing system.
516 doi:[10.7287/peerj.preprints.778v2](https://doi.org/10.7287/peerj.preprints.778v2).

517 44. **Schloss PD, Jenior ML, Koumpouras CC, Westcott SL, Highlander SK**. 2016. Sequencing
518 16S rRNA gene fragments using the PacBio SMRT DNA sequencing system. PeerJ 4:e1869.
519 doi:[10.7717/peerj.1869](https://doi.org/10.7717/peerj.1869).

520 45. **Baxter NT, Koumpouras CC, Rogers MA, Ruffin MT, Schloss P**. 2016. DNA from fecal
521 immunochemical test can replace stool for microbiota-based colorectal cancer screening. bioRxiv
522 048389. doi:[10.1101/048389](https://doi.org/10.1101/048389).

523 46. **Baxter NT, Koumpouras CC, Rogers MAM, Ruffin MT, Schloss PD**. 2016. DNA from fecal
524 immunochemical test can replace stool for detection of colonic lesions using a microbiota-based
525 model. Microbiome 4. doi:[10.1186/s40168-016-0205-y](https://doi.org/10.1186/s40168-016-0205-y).

526 47. **Imperiale MJ, Shenk T, Bertuzzi S**. 2016. mSphereDirect: Author-initiated peer review of

527 manuscripts. *mSphere* **1**:e00307–16. doi:[10.1128/msphere.00307-16](https://doi.org/10.1128/msphere.00307-16).

528 48. **Vence T.** 2017. Journals seek out preprints. *TheScientist*.

529 49. **Barsh GS, Bergman CM, Brown CD, Singh ND, Copenhaver GP.** 2016. Bringing PLOS
530 genetics editors to preprint servers. *PLOS Genetics* **12**:e1006448. doi:[10.1371/journal.pgen.1006448](https://doi.org/10.1371/journal.pgen.1006448).

531 50. **Hockett KL, Nishimura MT, Karlsrud E, Dougherty K, Baltrus DA.** 2013. Interactions
532 between genome architecture and virulence genes in *pseudomonas syringae*, strain cc1557 as a
533 model. bioRxiv 000869.

