# Developing a preprint culture in biology

Philippe Desjardins-Proulx,[1, 2, 3, *] Ethan P. White,[4] Joel J. Adamson,[5]
Timothée Poisot,[1, 2, 6] Karthik Ram,[7] and Dominique Gravel[1, 2]

[1] *Theoretical Ecosystem Ecology laboratory, Université du Québec à Rimouski, Canada.*
[2] *Quebec Center for Biodiversity Science, McGill University, Canada.*
[3] *Département des sciences biologiques, Université du Québec à Montréal, Canada.*
[4] *Departement of Bology, Utah State University, United-States of America.*
[5] *Ecology, Evolution and Organismic Biology, University of North Carolina at Chapel Hill, United-States of America*
[6] *International Network for Next-Generation Ecology.*
[7] *Environmental Science, Policy, and Management. University of California, Berkeley, United-States of America.*

## I. INTRODUCTION

Public preprint servers allow authors to make manuscripts publicly available before, or in parallel to, submitting them to journals for traditional peer-review. The rationale for preprint servers is fundamentally simple: to make the results available to the scientific community as soon as possible rather than to wait until the peer-review process is fully completed. The goal of arXiv and open preprint servers is not to circumvent the peer-review process; almost all manuscripts on arXiv are submitted to peer-review. Sharing ideas as quickly as possible using preprint servers has numerous advantages. These include rapid dissemination of work-in-progress to a wider audience, including readers in developing countries where access to subscription journals is often a limiting factor in scientific research. Furthermore, preprints provide the opportunity to solicit feedback from a larger pool of reviewers and can be seen as an integral aspect of a vigorous peer-review process [?]. Preprint servers therefore increase the number of opportunities for review and revision prior to publication, resulting in higher quality submissions and also could alleviate reviewer burden.

Preprints gained popularity 20 years ago with the advent of arXiv, an open preprint server widely used in physics and mathematics [?]. Preprints are also integral to the culture of other scientific fields. Paul Krugman noted that, in economics, the *traditional model of submit, get refereed, publish, and then people will read your work broke down a long time ago. In fact, it had more or less fallen apart by the early 80s* [?]. In addition to a section in arXiv, economists have also the RePEc (Research Papers in Economics) initiative, which aims to create an archive of working papers, manuscripts, and book chapters. Despite the success of this approach in other fields, most manuscripts in biology are not submitted to a public archive before being submitted to peer-review. In this article, we first highlight the advantages of open preprint servers for both scientists and publishers. We then debunk a few misconceptions, discuss the policies of major

publishers in biology, and briefly review the most popular open preprint servers currently available.

## II. THE CASE FOR PUBLIC PREPRINTS

The first and most often discussed advantage of arXiv and open preprints is speed (Figure 1). The time between submission and the official publication of a manuscript can be measured in months, sometime in years. For all this time, the research is known only to a select few: colleagues, editors, reviewers. Thus, the science cannot be used, discussed, or reviewed by the wider scientific community. This stage of manuscript preparation is a problem for both scientists and publishers. Manuscripts that are unknown cannot be used and thus take more time to be cited. It has been shown that high-energy physics, with its high arXiv submission rate, had the highest immediacy among physics and mathematics [?]. Immediacy measures how quickly articles are cited.

Furthermore, the review process as a whole is critically over-loaded, because as the number of active scientists increases, because the pressure to publish increases, and because of an effect dubbed "the tragedy of the reviewers commons" [?]. At the same time, rejection rates are high in most journals [? ?], and when not invited to submit a revision, authors must start the whole process all over again. Initiatives to reduce time from submission to publication have emerged across the scientific community. Rohr et al. [?] called for the recycling and reuse of peer-reviews: by attaching previous reviews and detailed replies to a new submission, both the editor and the referees can gauge the work done on the manuscript, and perhaps evaluate it with less prejudice. In a similar way, the *Peerage of Science* initiative allows authors to seek anonymous pre-review by their peers. Some journals now accept to publish papers which received good evaluations, with *Animal Biology* having recently accepted first a paper reviewed entirely with the *Peerage of Science* [?], effectively outsourcing the review process. A widespread use of preprint servers can achieve the same goal of reducing the time spent in review. By putting a manuscript for open comments and criticisms, the au-

---

* E-mail: philippe.d.proulx@gmail.com

thors will receive valuable feedback and can improve the version which will be submitted. With a rich enough community of scientists depositing preprints, and commenting on them, the process of an open pre- review can become widespread and will overall increase the quality of first submissions.
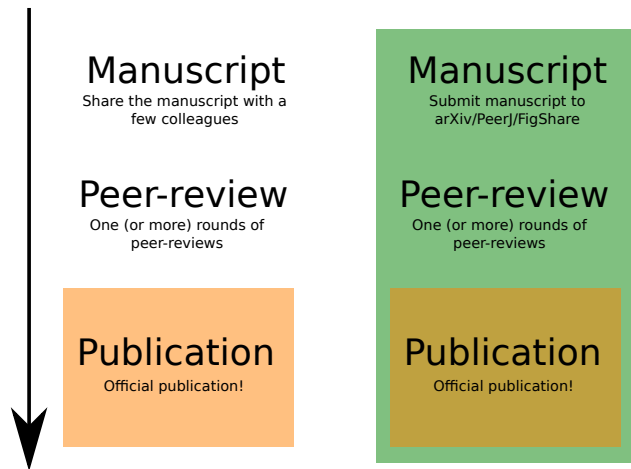


FIG. 1. It can take several months, and even a few years, before a submitted paper is officially published and citable. The average time to publication varies greatly between journals and can be as low as 104 days (Evolution for 2011) to 213 (PLOS One in 2010). Meanwhile, few people are aware of the research that has been done since, typically, only close colleagues are given access to the preprints. With public preprint servers, the science is immediately available and can be openly discussed, analyzed, and integrated into current research. It benefits both science and publishers. Both want the papers to be well-known and cited, and public preprints make it possible to integrate research even before publication, greatly improving immediacy.

Public preprint servers offer a much fairer way to establish intellectual priority by making the work available when done. Some manuscripts will spend much more time in the review process than others. Surprisingly, there is a perception in biology that public preprints make it easier to steal ideas, as if a scientific idea was only complete once published in a peer-reviewed journal [**?** ]. Ironically, mathematicians and physicists have embraced arXiv in part to establish priority in a fair way [**?** **?** ].

Prepublication reviews by a small network of colleagues is an important part of the scientific process, attested by the fact that nearly all published papers acknowledge comments by people not listed as co-authors. Preprint servers simply offer a way to extend this network of colleagues to the entire scientific community. It ensures that science is not constrained by small networks of scientists exchanging ideas. Paul Ginsparg created arXiv.org in part for democratic reasons: he wanted everyone from students in small universities to Ivy-League professors to have access to the most recent scientific

*ideas.* Ginsparg's revolutionary idea was simply to use the power of the internet for preprints, not just for the end product, so the scientific process can be open as soon as possible.

## III. PREPRINTS IN BIOLOGICAL SCIENCES

Submitting to a preprint server is becoming more common in biology, even though it still involves a minority of papers. The quantitative biology section in arXiv is experiencing faster growth in submissions than any other field [**?** ]. Most scientific journals are preprint-friendly: Nature, PLOS, BMC, PNAS, Science (mostly) I, and all the journals from Elsevier and Springer. The Ecological Society of America and the Genetics Society of America recently changed their policies to allow public preprints. Few scientific publications will not consider a manuscript submitted to a public preprint server. Still, a few journals adopt a "by default" hostile attitude towards preprints, mostly due to the lack of clear policy of the publishers, or perhaps because a preprint culture has not developed in biology and the practice is still considered unusual. As an example, Wiley-Blackwell, which publishes some leading journals, has no official policy on the subject I.

Part of the hostility to preprints servers comes from a certain interpretation of the Ingelfinger rule: scientists should not publish the same manuscript twice [**?** ]. However, papers submitted to arXiv or other preprints servers are only published in one peer-reviewed journal. A preprint is just a preprint, it is a document that allows ideas to spread and be discussed, but it is not yet formally validated by the peer-review system, which is why the majority of publishers does not see arXiv and similar services as a violation of the Ingelfinger rule. *Nature* responded to the rumour that they refused manuscripts submitted to arXiv by saying that "*Nature* never wishes to stand in the way of communication between researchers. We seek rather to add value for authors and the community at large in our peer review, selection and editing" [**?** ].

## IV. CURRENT OFFERINGS

We briefly discuss the main options to submit preprints to open servers: arXiv.org, Figshare, and the upcoming PeerJ and F1000Research.

### A. arXiv

arXiv (http://arxiv.org/) is the most widely-used preprint server today, and its use is almost universal in some branches of mathematics and physics. arXiv provides a reliable citation system for all eprints and is especially popular in high-energy physics. Physicist

| Publisher | Policy |
|---|---|
| Springer | Accept |
| BMC | Accept |
| Elsevier | Accept |
| Nature Publishing Group | Accept |
| Public Library of Science | Accept |
| Genetics Society of America | Accept |
| Royal Society | Accept |
| National Academy of Science (USA) | Accept |
| Ecological Society of America | Accept |
| Oxford Journals | Accept |
| Science | Ambiguous |
| Wiley-Blackwell | No general policy |
| British Ecological Society | No answer to our query |

TABLE I. Policies for important publishers in biology. Some publishers tolerate preprints except for a few of their medical journals, e.g.: Journal of the National Cancer Institute from Oxford and The Lancer from Elsevier.

Paul Ginsparg created arXiv in 1991 for theoretical high-energy physicists to communicate preprints via email and ftp, and soon thereafter adopted the newly created world-wide web[**?** ]. arXiv now receives over 7 000 submissions per month (http://arxiv.org/show_monthly_submissions) and divides its submissions into subcategories of physics, mathematics, computer science, quantitative biology, finance and statistics. The quantitative biology category includes subcategories for Populations and Evolution, Quantitative Methods and other categories that may be of interest to biologists.

Submission to arXiv is fully automated. Authors can submit TEX/LATEX documents that are compiled on the server or directly submit in PDF/PS format (for example, as exported by a word processor). A moderation system was put in place in 2004: papers must be categorized by an endorser. At least one author of a paper must be an endorser that has previously submitted a paper or has received permission to submit to a particular category. Many authors in mathematics and physics submit papers as soon as they are ready for review by colleagues, although another popular option is submitting simultaneously to a journal and arXiv.

Authors must either have their arXiv submission available under an open license or grant arXiv a non-exclusive and irrevocable license to distribute the work. In either case, arXiv does not require copyright transfer and only requires the rights to distribute submitted articles in perpetuity. Thus, submitting to arXiv does not in itself prevent the authors from transferring their rights to a publisher, which is why most publishers tolerate arXiv, even though they ask the authors to transfer their right to them upon acceptance of the article.

Most papers posted to arXiv are eventually printed in journals but there are notable exceptions, such as Perelman's landmark paper leading to the proof of the Poincaré conjecture [**?** ]. However, arXiv has never sought to replace scientific journals and explicitly states that it serves a different function as "an openly accessible, moderated repository for scholarly articles in specific scientific disciplines." arXiv is now administered by the Cornell University Libraries, with funding coming from voluntary pledges by academic institutions along with matching funds from the Simons Foundation [**?** ]. One-hundred twenty six of the top two-hundred institutions in terms of downloads have provided the operating budget for arXiv over the next five years. This plan reduces the financial burden on Cornell University and transfers governance to a collaborative community in accordance with arXiv's key principles. arXiv takes numerous measures to ensure that the repository will remain permanently available and submissions will be readable.

## B. FigShare

FigShare (http://figshare.com/) is an open server that allows scientists to submit any research output: manuscript, figures, datasets, videos, theses, presentations, and so on. There are no rules to limit what constitutes a research output and, unlike arXiv, there is no endorser system. All FigShare content has a unique digital object identifier (DOI) like any journal article, thus offering a permanent and stable link to the content. A flexible tag system is used to classify each item. All content can be commented and is licensed under the Creative Commons (CC-BY) license, except datasets which are published under CC0. The CC-BY license grants rights to *copy, distribute, display and perform the work and make derivative works based on it only if they give the author or licensor the credits in the manner specified by these.* CC0 is the most permissive license and effectively puts the work in public domain (no rights reserved) or, if it is not possible in the given jurisdiction, provides a simple permissive license.

One of the biggest advantage of FigShare over arXiv is that is it not limited to quantitative sciences. arXiv.org has sections on quantitative biology but might not be appropriate for non-quantitative work. With its flexible approach to preprints, FigShare offers an important alternative to arXiv for empirical biologists. Furthermore, by allowing all types of content, FigShare arguably pro-

vides an archive for early results (e.g.: figures, lab presentations).

### C.   PeerJ

PeerJ (https://peerj.com/) is a new publishing system that combines both a preprint server, and a peer reviewed journal. It is focused on the biological and medical sciences, which may help overcome the perception that preprints do not have a home in biology. PeerJ allows commenting on posted preprints, improving the potential for pre-publication dialog. In addition, preprints can be made private if the authors choose, and shared only with selected colleagues. While this reduces some of the benefits of preprints described above, it may allow some researchers who would not otherwise post preprints to begin to explore the possibility in a manner appropriate to their current circumstances.

In contrast to other preprint servers users cannot post unlimited public preprints for free. One preprint per year can be posted for free and a onetime (i.e. lifetime) fee of 99 dollars allows the posting of unlimited public preprints. It is also worth noting that the preprint server is not tied to the journal, so preprints can be posted regardless of where they will eventually be submitted for publication.

PeerJ uses the CC-BY-SA 3.0 license, which is similar to the CC-BY license used by Figshare but adds the *Share-alike* (sa) restriction that derivate works need a license identical to the license that governs the original work.

### D.   F1000Research

F1000Research is not a public preprint server like the previous three servers. Whereas arXiv, Figshare, and PeerJ offer an option to submit a manuscript without having it reviewed, papers submitted to F1000Research will eventually be reviewed. Thus, F1000Research offers a hybrid model with publicly available manuscripts at time of submission and standard peer-reviews. Manuscripts are considered "accepted" and will only be indexed after two positive referee response. F1000Research works closely with data providers to integrate raw data to the paper. For instance, upon submitting a paper, authors are asked to upload their data, which are then integrated in *e.g.* FigShare widgets, the DOI of which are given in the paper when the data are first mentioned. The licensing of the data is similar to the one used by FigShare, meaning that the articles are free to access, and can be redistributed readily. By putting much effort in integrating data to the paper, F1000Research is working to make science more reproducible and open.

### E.   Github

This manuscript was developed entirely as an open project on GitHub. GitHub is one of several hosting services for collaborative development using the git version control system (VCS). git is a decentralized revision control system created by Linus Torvalds and is used primarily to develop software, including the Linux kernel. Git provides powerful features that allow numerous contributers to work asynchronously on the same project, often in parallel branches, all of which can be effortlessly merged and version controlled. Although created primarily for software development, git is ideal for academic research since it provides a way to version control and collaborate on every step of the manuscript development process, from data manipulation and analysis to writing and revision. For example, during the development of this manuscript, each author would clone the project (*i.e.* make a personal copy), modify it, and then merge their changes into a master branch.This takes the preprint process to an entirely new level, where the entire writing process is open from the beginning.

## V.   CONCLUSION

Open preprint servers offer a great opportunity for open science, especially if the community embraces the idea of discussing preprints. Initiatives like Haldane's Sieve (http://haldanessieve.org/), a new blog discussing arXiv papers in population genetics, will help make arXiv attractive for scientists looking to promote their work. These initiatives are important to fully exploit the potential of open preprints servers. Posting preprints online increases the community of available informal peer reviewers, and uses the internet for its original community-building purposes.

Preprint servers also facilitate communication between disciplines, bridging cultural as well as geographic divides. Examples include a recent series of papers on the theory of natural selection that was posted to arXiv simultaneously with its publication in the *Journal of Evolutionary Biology* [? ? ? ? ]. Other submissions in this category include evolutionary and ecological theory by authors trained in physics and computer science. Since authors in these fields regularly check arXiv, submitting preprints may be the most effective way biologists can help others avoid "repeated work" [? ]  and form a synthetic community of evolutionary theorists from disparate backgrounds. The advantages are clear and the costs are low.

## VI.   FUNDING

## VII.  ACKNOWLEDGEMENTS

We thank Carl Boettiger for helpful comments on an earlier version of this manuscript.