

## Revisiting authorship, and JOSS software publications

We are slowly working towards a v2.0 release of [sourmash](#), our software for MinHash and modulo hash analysis of genomic data, and the question of proper authorship is once again on my mind!

Wed 16 January 2019

By [C. Titus Brown](#)

In [science](#).

tags: [publication](#) [software](#) [citation](#) [sustainability](#)

The question du jour: how should authorship on software papers be decided?

### Some background - our previous take on authorship

Those of you with long memories may recall a [hullabaloo in 2015](#) over this occasioned by [the khmer v2.0 paper submission to F1000 Research](#). Briefly, some took exception to our offer of authorship to **all contributors to the GitHub repository** in the publication, while others thought it was just fine. The reviewers had some interesting things to say about our authorship considerations (see the "Open Peer Review" section of [the paper](#)) but despite some reservations ultimately the paper was approved for publication.

One of the strongest outcomes for me of all of this was that I realized how inane any cut & dry principle of scientific authorship was - or, to rephrase, authorship decisions **often contain a strong subjective component**. Arguments for and against authorship can easily be made in many situations, but there are always corner cases that break each argument.

And this is true even when you have full version control history, as we do on khmer and sourmash.

For one example, it's always possible to argue that even significant code contributions are not major intellectual contributions. "It's just a bug fix", or "that's just engineering", - those or similar arguments can always be made. This makes it difficult to just look at the git commit log!

The same goes in reverse: someone who didn't contribute to the code base at all (and for which there are no tracked contributions) might have made significant intellectual contributions. Lots of action on scientific and software projects happens in other forums or in meatspace, and those should be rewarded too! (In the khmer foofarah, I note that I could have easily said, "hey, look, en zyme and I took long moonlit walks along the banks of the Willamette, and discussed khmer deeply and thoroughly; on that basis I consider their intellectual contribution to be significant and deserving of authorship." Who could have gainsaid me, other than en zyme?)

The bottom line is that I think might not have been entirely right to offer authorship to all committers (in the strict scientific sense of authorship), but at the same time people arguing against it were also off base. In fact, by focusing on git commit records we probably spiked the conversation and centered it around something reasonably verifiable rather than the deeper and more interesting questions. So it goes.

My real conclusion is this:

Fundamentally, in order to nurture a diverse array of valuable scientific contributions, we need new models of

publication with new models of authorship.

## Authorship v3.0, for sourmash v2.0?

Authorship for khmer was confused by the lack of a good publication venue - F1000 Research is too close to a traditional journal, basically. That confused things.

Conveniently, in the intervening years, a wonderful group of people created the [Journal of Open Source Software \(JOSS\)](#)! And, in fact, this is where sourmash v1.0 [was published](#). But authorship for sourmash 1.0 was easy: it was just me and [Luiz Irber](#), because we were the only contributors!

It is now time to release sourmash 2.0, a much revised and expanded piece of software that a number of people have been using (in my lab and outside it) for many many things. And of course we want to start by publishing an update in JOSS!

Since the publication of sourmash 1.0, however, we have had some amazing contributions to sourmash. Some really significant code and doc contributions have made it in, because we do our best to be friendly about accepting code. But there's more!

Many different members of the Lab for Data Intensive Biology have contributed extensively to the sourmash software by using sourmash in their own projects. A number of people in my lab engaged intellectually with the MinHash and modulo hash methodology, found and reported bugs, asked questions that led to new docs and tests, and motivated new functionality.

Outside of the lab, there were also a number of contributions. We had one user who picked up sourmash for their own project, and over a period of almost two years has reported many bugs and UX problems. A colleague at another institution pointed out that some of our calculations were incorrect, and provided utility code to validate our calculations. Another colleague built a pipeline on top of sourmash and reported several different bugs. Some collaborators really dug into sourmash functionality for another pipeline implementation, and suggested a lot of features. And yet another wonderful person created a bioconda recipe for sourmash, which led to many more users.

From my perspective, all of these people contributed to sourmash, and made the software better. I think they deserve to be offered authorship on the software software paper in JOSS.

(*Tracking* who receives consideration for authorship is hard on a multi-year project, especially when many contributions come in informally via e-mail or in-person interactions. Thoughts there welcome...)

## A rubric for authorship on software publications

I think the right rubric for authorship on software publications is [engaged effort in the project](#). (This comes out of a broader effort to think about [sustainability using the Common Pool Resource framework](#)).

What this means in practice is this:

If I or anyone else involved in a project can pinpoint the contribution made by an individual, **and it was [a positive contribution, as opposed to an extractive one](#)**, that person is a contributor. This excludes users who just used the software, but could include users who filed bug reports, asked good questions that led to new documentation, and otherwise engaged with the software.

And contributors deserve to be offered authorship.

## Who decides?

I've been involved in several authorship disputes over the years, and, in my opinion and my experience, the senior author decides authorship. After that point other authors can decide if they want their contributions to be part of that scholarly work.

So, basically, in the case of sourmash I get to decide who to invite; at least in this case, there is no question who is the senior author on sourmash 2.0.

In this case, I believe that having submitted code via a PR is definitely a contribution. I'm going to have to think about how to evaluate issue contributions; right now I don't have a good way to summarize them...

## Other questions

Authorship discussions are an excellent opportunity for involved commentary and scholarly nitpicking! Here are some of my nits --

When people (inside or outside the lab) use unpublished features of sourmash in their own research, should they include some or all of the sourmash authors on their work? I've generally defaulted to "no" - they should certainly cite sourmash, but even if sourmash enabled otherwise impossible analyses, we didn't contribute specifically to their research, and shouldn't be co-authors.

---

Is publishing the software in JOSS and then writing papers on various bits of sourmash for other journals so-called "[salami slicing](#)", or breaking up publications into their least publishable units?

For example, here we're expecting to have a version-specific JOSS paper, an F1000 "sourmash use cases" paper, and at least two (maybe three) different research papers investigating and using sourmash for different projects. These papers will have different messages, different workflows, and different (but overlapping) sets of authors, so my judgement is that it's not salami slicing. But even if you agree, so many papers will make citing sourmash challenging and complicated, and probably "dilute" the citations to any one paper... a conundrum for those who count citations-per-person!

I think it's up to us to make the message of each paper clear, and provide clear citation guidance for different features (as we [have done for khmer](#), note).

---

Another question: does authorship keep accruing over versions? Should all the authors on sourmash 2.0 be authors on sourmash 3.0? I'm not sure how to deal with this. Wait a few more years for the khmer 3.0 and 4.0 releases and we'll see ;).

## The bottom line

Citations are currency in academia, for better or for worse, so discussions of authorship are always fraught with tension around dilution of this currency. I think it's better to acknowledge people's engaged effort than not, and leave it up to co-authors to explain their own effort when asked. This is more or less how our merit and promotion works here at UC Davis - we have to explain clearly how we have contributed to each paper.

From the perspective of sustainability, I would argue that thinking broadly about contribution is a way to recruit, reward and retain maintainers and engaged users, which could lead to increased sustainability.

(This is all quite different from legal authorship on the software itself, which is a different and differently challenging question that I'm not remotely qualified to discuss. :)

--titus

p.s. Thanks (as always) to Michael Crusoe and Luiz Irber for their major impact on my thinking here! I would also like to especially thank Matthew Turk and Daniel Katz for their engagement with me on this topic over the years.

p.p.s. I would be remiss in not mentioning Casey Greene's tremendously-inspiring-and-as-usual-oddball [approach to authorship](#) on the Deep Learning in Biology review paper. So! It is now mentioned!

## Comments!

([Please check out the comments policy before commenting.](#))

3 Comments Living in an Ivory Basement - blog

1 Login ▾

♥ Recommend 2     Tweet     Share

Sort by Best ▾



Join the discussion...

LOG IN WITH

OR SIGN UP WITH DISQUS 

Name



**Luis Pedro Coelho** • 4 days ago

I started writing an answer here, but it grew so big, I just posted it on my own blog like in the olden days of blogging: <https://metarabbit.wordpress...>

tl;dr I am in favor of the significance criterion even though it requires judgement calls. Also, I wish the community (and journals) would converge on some expectation of maintenance going forward, so that being an author would imply taking over that responsibility too.

^ | ▾ • Reply • Share ▸



**Daniel S. Katz** • 6 days ago

Thanks for this nice post.

A couple of comments:

1 - I agree we need a discussion about authorship, but I don't know that software is so different than anything else today in terms of authorship that we need a different rubric. I think what you state for software is probably sufficiently general to work for most papers, as well (though publishers and editors often think they should be the ones to decide which contributions merit authorship): In brief, if the positive contribution made by an individual can be clearly identified, that person is a contributor and should be an author.

2 - Regarding accruing authorship over versions, I would go back to the same rubric: if the person's contribution is still part of the product, they should still be an author.

^ | ▾ • Reply • Share ▸



**Titus Brown** Mod → **Daniel S. Katz** • 6 days ago

Thanks, Dan!

1 ^ | ▾ • Reply • Share ▸

## How open is too open?

14 comments • 7 months ago



**benmarks** — This post speaks to much of what we've implemented and discovered in the last couple of years at Magento. Our 1.5-year-old ...

## For want of \$2.41... some background on reimbursements.

4 comments • a year ago



**Erich Schwarz** — Caltech used to be quite generous about providing advances for travel, etc., to its graduate students in biology. I know ...

## My approach to community building and coordination

2 comments • 10 months ago



**Rayna** — Thanks for the blogs Titus and Stephanie. Stephanie, I took your advice and started a #welcome channel for the Carpentries :)

## Pydoit, snakemake, and workflows-as-applications

5 comments • 10 months ago



**Titus Brown** — :)

---

 [Subscribe](#)  [Add Disqus to your site](#)[Add Disqus](#)  [Disqus' Privacy Policy](#)[Privacy Policy](#)[Privacy Policy](#)

Proudly powered by [pelican](#), which uses [python](#).

The theme is subtly modified from one by [Smashing Magazine](#), thar

For more about this blog's author, see [the main site](#) or [the lab](#) :

While the author is employed by the University of California, Davis, his opinions are his own and almost certainly bear resemblance to what UC Davis's official opinion would be, had they :