SIPs with Standards

Menaka: This episode, we're zooming in on a specific technique -- it's called Stable Isotope Probing, or SIP, for short. And first, I want to set up what this kind of experiment offers us.

SIP is a way of studying microbial communities. It's a view into the organisms that exist in environments like soil, and freshwater. And these communities are tiny, but vast.

Researchers often talk about a teaspoon of soil containing about a billion microbial cells -- a huge population.

So, imagine a microbial community like that as a place like Tokyo, Japan <city sounds> -- the biggest city on Earth, where 37.4 million people go about their lives. In such a big place, there's a lot going on. There are imports coming in, people making food and goods, others delivering stuff around, and lots of these activities are interconnected. They depend on each other.

In a microbial environment, things are also bustling this way. Bacteria, archaea, viruses, and fungi are operating their own metropolis. Some organisms are bringing in nutrients like carbon and nitrogen, some are processing those nutrients into different forms, and passing them along to other microbes, others are releasing chemical compounds that we're interested in.

And we want to understand all of these microbial activities, because mapping them out could help us answer all kinds of environmental questions. Questions like, how we can keep more nitrogen around in crop soil, to use less fertilizer? Or, how can we lower the amount of methane coming out of a freshwater lake? These are big questions that start with tiny microbial activities.

But -- the same way that it would be super complicated to follow how all kinds of goods travel around Tokyo's networks, getting a clear read on a microbial community is very tricky. All of the nutrients that researchers want to trace go through many organisms, multiple steps, and many chemical forms.

This is where SIP comes in. As a technique, it essentially lets researchers trace different kinds of nutrient deliveries as they go around a microbial city. <city sounds out>

To do that, researchers pick a nutrient to trace, then feed a microbial community a version of that nutrient that's just slightly heavier than normal -- a version of the nutrient made from a stable isotope like Carbon-13 instead of Carbon-12. Then, when microbes use that heavier nutrient, it's traceable -- so SIP experiments can show which microbes are handling specific nutrients, or what they're doing with those nutrients, and even how quickly. Which is amazing -- so, of course, SIP is a pretty intensive lab technique.

Here's Roli Wilhelm, a JGI user who is an assistant professor at Purdue University.

Roli Wilhelm: Yeah, I spent my whole PhD SIPping. I think I tallied the number of hours just that the ultracentrifuges were spinning and it was something like three months.

So it's a lot of energy, it's a lot of time. And then for each one of the samples you generate in a SIP experiment, you typically have more than one downstream.

Menaka: In the last few years, the JGI has set up Stable Isotope Probing pipelines, first for regular DNA, and now quantitative versions of these analyses. Rex Malmstrom runs a group that does SIP analysis at the JGI.

Rex Malmstrom: By offering it as a service at JGI, we're trying to bring in a lot more people who want to do it but just otherwise couldn't or wouldn't.

Menaka: Because the analysis is tricky, but the data the SIP offers is so valuable. And as SIP becomes more and more available both at the JGI and other labs, Rex, Roli and their collaborators want to create great results -- in other words, data that can be useful to multiple teams and analyses.

They want to make the most of every SIP situation. Partly, because the experiments are tough, but also, because the more we can connect the dots of these experiments, the more global information we'll have about these important, interconnected microbial processes.

So today -- we've got a chat with Rex Malmstrom, and Roli Wilhelm, about a few different ways they're working to make this technique, SIP, more standardized -- more reproducible, more reusable, and ultimately, more insightful, for the future of studying microbial communities.

<THEME>

Menaka: This is Genome Insider from the US Department of Energy Joint Genome Institute. Where researchers discover the expertise encoded in our environment — in the genomes of plants, fungi, bacteria, archaea, and environmental viruses — to power our future. I'm Menaka Wilhelm.

This episode is a chat about standardizing SIP, with Rex Malmstrom, head of the Microscale Applications Group at the JGI, and Roli Wilhelm, who's a JGI user and an assistant professor at Purdue University. I talked with them, but they're working on this with lots of other people. You'll hear in their interviews — collaboration is a big priority in the way they're working to improve data reporting and analysis.

And we'll start with Roli's mission to help standardize the data that researchers report from their SIP experiments. His goal is for these intensive experiments to be useful to multiple teams once they're done -- and that became a goal when he first started as an assistant professor at Purdue, 3 years ago. One of the first projects he wanted to do was a comparison across SIP studies -- but that turned out to be less possible than he would've hoped.

Roli Wilhelm: As soon as we started out, my student, Abigayle Simpson, or actually she was a technician, She found that most of the data that was publicly available, published, archive data just was lacking the basic fundamental information to be able to use it in a SIP.

So you couldn't actually, you couldn't replicate some of these studies and that was quite bad. And so we were motivated to make some lemonade from those lemons. So we thought, let's go on a pilgrimage. Let's, let's make a difference, from a different direction and that's helped people do a better job curating the data.

Menaka: And that pilgrimage took the form of a project to create a standard for the Minimum Information for any Stable Isotope Probing Sequence. That acronym-izes to MISIP, and Roli, Abigayle, and their team published this in a paper last fall.

Menaka: Yeah. It seems like you assembled a team basically to work on this paper. So who else did you know who would be interested in this kind of standard?

Roli Wilhelm: Well, it was interesting. I started with reaching out to a contact of mine who I knew worked on KBase at Argonne. That was Pamela Weisenhorn. we'd worked on some KBase related things.

Menaka: Kbase is the Department of Energy Systems Biology Knowledgebase -- it's an open source web platform where researchers can collaborate to analyze data and build models. KBase is a place where lots of people are interested in SIP experiments, and also a place that's enthusiastic about data standards. So Roli knew they'd be helpful in terms of setting up parameters and vocabulary and standardization. But his initial contact, Pamela Weisenhorn, was busy. She sent Roli to Elisha Wood-Charlson, who's User Engagement Lead at Kbase.

Roli Wilhelm: Who was polite enough to meet with me and entertain my bizarre ideas and then put me in touch with some other folks at the JGI and Lawrence Berkeley and Lawrence Livermore, as well as Northern Arizona University.

So Bruce Hungate, Jennifer Pett-Ridge, Rex Malmstrom. And we were all then ushered into a zoom room, and we had sort of a discussion about what we were wanting to do, and what we're thinking about in that space.

Menaka: The space being -- how to help SIP studies include all the information that someone would need to replicate them or reuse their data. They started out by basically collecting information about what existing SIP studies were reporting. In some ways, deciding what to include in a standard like this is a little like deciding what gets included in a dictionary – and there are different ways to go about that. Here's how Roli thought about it.

Roli Wilhelm: the Oxford English Dictionary is a good example of the difference between a sort of a collection of ideas versus a prescription of ideas, right?

So the French famously have dictionaries that say, this is how you, this word should be used, any other way, it's wrong. Whereas the OED came along and said, you know, if the word is used three times, In print, back in the 1800s when printed word was pretty sparse, they said, send us that, you know, those three prints and we'll put it in the dictionary.

It doesn't have to be proper or aristocratic. If it appears in print three times, we'll put it in there. And that's kind of the approach we took, is we went out and we, combed through the literature to see what were people doing and using and how are they doing SIP. That gave us a bit of a framework to go on.

Menaka: More of a democratic, OED framework than an extremely rule-bound, strict set-up. And that framework was important because SIP experiments vary. Some are broad, and very quantitative, so they collect all kinds of information, while others are

much smaller scale. **Roli**, Abby and this team had to navigate trying to set up a standard that took all of this into account.

Roli Wilhelm: So what are the minimum bits of information? I actually think that that's sort of the, foresight and wisdom of the, Genomic Standards Consortium. Who've published a lot of these minimum information for X sequences standards, right? Ours, MySIP is a one in a long chain. Ours is the minimal information for a stable isotope probing experiment.

Menaka: Similar to assembling a team at the start of this project, they cast a wide net for feedback as they got to the end.

Roli Wilhelm: So this is part of the process, putting out a pre-print and then blasting it across social media to say, hey, what are your thoughts on this?

We got some feedback. Not enough, so we then went through the literature and found all of the SIP studies that have been published in the last five years, grabbed the corresponding authors and emailed them saying, Hey, what do you think of this? And we got some much better feedback there, with a survey, that was kind of cool. We got a lot of feedback from researchers in China, other parts of the world.

Menaka: Including, the JGI's Rex Malmstrom -- who was happy to see a start to standardization for data and metadata reporting.

Rex Malmstrom: Those metadata are critical for — they are critical for doing things like taking different studies and comparing them and trying to look for those bigger patterns. And it's kind of making people aware of like, hey, it doesn't need to be just this one off study. If you collect and provide this metadata, you are now enabling these much larger comparisons down the line.

Menaka: Yeah, like ideally you have a really comparable metadata and people can see it

Rex Malmstrom: and can see it and they can find it and all those things.

Menaka: For **Roli**, creating more awareness is a key way of setting this up -- and he's seen researchers show interest in improving this process, too.

Roli Wilhelm: Well, I would say within a few weeks of publishing MySip, uh, officially, uh, I had a student from Germany reach out asking me, is this, is this what I need to do?

And we worked through some of the things and they had kind of nailed it. They had really done a fantastic job and it just gave me honestly, the sense of hope, you know? And so what, what I like to take away from this is that we can do it right. And we can feed this future need of having high quality, useful data that anyone around the world can then tap into, right? So that's what's so beautiful about standardizing data is it's actually like a democrat democratizing process.

Menaka: Ultimately, better data reporting speaks to the goal of making more studies more useful -- basically, the place where Roli started with this project, when he and Abby realized their meta-analysis would need way more information.

Roli Wilhelm: And we felt like we're sort of stemming the tide of, of waste. We're like, if we could only, if we could only do this better.

Menaka: Yeah. Yeah. No, it's sort of infrastructural, but like lots of infrastructure, it lets lots of things happen later on. Right. That's the goal.

Roli Wilhelm: Yeah. And also, we started this before the first large language model became public utility.

Right. And so this idea of like data being gold, it's, you know, it's been kind of around, but this is one key limitation to all the data we've generated thus far, like, where would we be now had we been, had MISIP a decade ago, right? Like we'd have so much more useful data and that really has this multiplying effect when you think of these, um, you know, language processing, data processing, data hungry types of algorithms.

Menaka: Yeah, yeah, that's interesting. That's true.

Roli Wilhelm: The truth is what I'm also hoping for is, you know, That this information can be then used in language models to train a more interactive sort of dialogue to say, Hey, I've done this kind of experiment of collected this kind of data. Can you guide me on what meta information and what labels to use? Or just that simple consistency of even how you upload the data right — we're learning how to do a better job of feeding that future creativity, motivation and interest of that next generation of, students, maybe AI bots, maybe the friendly, cool ones, the granola crunching AI who want to learn more about the natural world and the history of our planet.

Menaka: Because the better the information, and the better the labeling and reporting, the better the studies that build on initial work -- something we could all benefit from down the line. Next, we'll hear from Rex about work on standardizing SIP analysis -- that's after a quick break.

BREAK

This is Genome Insider, from the Joint Genome Institute. As we head back to our SIP standards, a quick refresh -- Rex Malmstrom runs the microscale applications group at the JGI, where users can request qSIP analysis as part of their projects. Roli Wilhelm is an assistant professor at Purdue University. And they're both interested in making SIP experiments more standardized, so that researchers' data can translate into more discoveries down the line.

We spent the first half of this episode with Roli Wilhelm, and his quest for data standardization -- and now, we'll jump to some ways of standardizing the analysis that creates that data. Rex Malmstrom got interested in this when he started working to make SIP and qSIP more available to more researchers, through JGI's proposal calls. Here's Rex.

Rex Malmstrom: There's a lot of people interested in doing SIP metagenomics,

Menaka: Because remember, the data is very valuable, but the barrier to entry is also very high.

Rex Malmstrom: The trickiest part of the whole thing comes down to this one step where you're collecting these little drops and you're like one, two, three, next tube, one, two, three.

Menaka: <laughs>

Rex Malmstrom: And it's just a real beast to do.

Menaka: Which is why the JGI started offering qSIP as a service to users who have submitted proposals.

Rex Malmstrom: qSIP came to JGI through this program called the Emerging Technologies and Opportunities. So we'd worked with Jennifer Pett-Ridge's group out of Livermore to develop this process .

Menaka: And that meant setting up hardware to automate lab work, plus software to handle the data coming out. That sparked Rex's interest in standardization.

Rex Malmstrom: And as we did that, it became clear to me that when you have quantitative measurements, you need analytical standards. You need ways to know, how good are the measurements that I'm making, how accurate, how precise? And it needs those things to evolve and get better.

Menaka: So they collaborated with researchers at the University of British Columbia, as well as Jennifer Pett-Ridge at Lawrence Livermore National Lab. Together, they all set up some analytical standards for this kind of qSIP analysis.

Rex Malmstrom: And it's now just part of what we do at JGI. It's great that then for JGI projects that they all have this internal standardization.

Menaka: But what if that kind of standardization got bigger!

Rex Malmstrom: Yeah, it really improves the results, we should all be using the same type of standards.

Menaka: To bring these kinds of standards to an even wider community than those direct collaborators, Rex set up a workshop -- a few days where researchers in this community could gather, and basically talk SIP nonstop, to share experiences and come up with new ideas for improving this technique. And he organized that workshop with a team of people, of course.

Rex Malmstrom: So the committee was Bruce Hungate, Jennifer Pett-Ridge out of Livermore Lab, Egbert Schwartz, Emily from the JGI and me.

Menaka: That’s Emiley Eloe-Fadrosh, head of the metagenome program at the JGI. And Rex says Bruce Hungate was especially important to organizing this workshop. As they organized, they sort of realized — no one had done this before.

Rex Malmstrom: This was the first workshop of its kind.

Menaka: A first time for everything! And they chatted about data standards, similar to what **Roli** was interested in from the first half of this episode,

Rex Malmstrom: Those were key parts of this workshop, to say, let's bring together people who are doing this process, and let's talk about it, and see what works, and see if we can come up with some ways to just make it better and more uniform.

Menaka: Along with the analytical standards that Rex has focused on in the last few years.

Rex Malmstrom: That was one important pillar of the workshop, which was, let's talk about standardization rather than everyone kind of — off in their own labs and their own groups doing their own process.

Menaka: And Roli Wilhelm, from the first half of this episode, was one of the few dozen people who made it to this workshop -- he thought all of that reasoning together was very useful.

Roli Wilhelm: This workshop was really nice, people were given structure, but it was all about sharing ideas. All breakout sessions, you know, just enough context to get people thinking and tackling some of the issues and it was really productive in that sense.

Menaka: One idea that the workshop helped propel was an experiment to compare qSIP analysis across the field. Rex Malmstrom and Egbert Schwartz from Northern Arizona University are spearheading that project.

Rex Malmstrom: The notion there is let's do an experiment where we're incorporating different analytical standards,

Menaka: Standards that different teams from around the world have developed. But, in this survey, these teams will all use the same uniform starting material.

Rex Malmstrom: Let them run that material through their own process, and then return it back to us so that we can sequence it, and then we can all look at it together.

Menaka: And then, they can see -- how comparable are the results that these different teams and different standards produce? And that'll be an exciting thing to look at, once it's all done.

Roli Wilhelm: That kind of overarching, you know, community driven knowledge is kind of crucial. I wish we could do more of that in general, and certainly it does matter and reflects on what the data looks like at the other end.

Menaka: Altogether, these pieces of SIP standards that Rex and Roli have worked on tie into each other. In order to get better results out of all of this work, they each want more context for comparison, and a better understanding of how comparable different analyses are. And to get there, they're working across their research

community, to set up approaches that work for a wide net of collaborators, so more people can make the most of SIP as a data source.

Menaka: So again, that was Roli Wilhelm from Purdue University and Rex Malmstrom from the JGI.

We'll link to more information about SIP and qSIP at the JGI in our episode description. You can also find a link to a transcript of the episode there!

As always, you can learn more about how to work with us at jointgeno.me/proposals. This episode was written, produced and hosted by me, Menaka Wilhelm.

I had production help from Allison Joy, Massie Ballon, and Graham Rutherford.
You heard music in the middle of this episode by Cliff Bueno de Mesquita, a former postdoc at the JGI.

If you liked this episode, subscribe or follow wherever you're listening, and help someone else find it! Tell them about it, email them a link, or leave us a review wherever you're listening to the show.

Genome Insider is a production of the Joint Genome Institute, a user facility of the US Department of Energy Office of Science located at Lawrence Berkeley National Lab in Berkeley, California.

Thanks for tuning in – until next time!