Genome Insider S5 Episode 5: Adopt-A-Genome

Menaka: This episode, undergrads are adopting genomes. So what that means is these students, with a professor, take on a genome that the JGI sequenced as part of a larger project. These are genomes that never made it into published papers, and these students get to change that. They do some genome analysis, write a short report, and bring that report through peer review -- so they publish a first author paper as part of this project, which then makes it easier for other researchers to reference and learn from these genomes later on.

Menaka: It's a really unique project, and we'll explain more specifics in this episode -- but first, I want to introduce one team of genome adopters. They're from California State University Fresno.

Menaka: Here's Kalyani Maitra, from CSU Fresno's department of chemistry and biochemistry.

Kalyani Maitra: Currently I'm an associate professor and I have a research group where I have about seven undergraduates and four masters students working on different projects.

Menaka: Recently, two of those undergrads were Angela and Mark Soghomonian.

Kalyani Maitra: They happen to be twins. Extremely efficient.

Menaka: Here's Angela,

Angela Soghomonian: I was a research student with Dr. Kalyani Maitra and I actually started in her lab when I was a freshman at Fresno State.

Kalyani Maitra: And Mark -- he joined Kalyani's lab at the same time as Angela, and that was a few years ago now.

Mark Soghomonian: I recently finished my undergraduate education at California State University of Fresno with Dr. Maitra.

Kalyani Maitra: I like the students to call me by my name, but I never got these two young kids to call me by my name.

Menaka: You were always Dr. Maitra for them.

Kalyani Maitra: Yeah .

Menaka: So Angela and Mark both spent a few years working in Kalyani's lab, where she had them working on a few different projects.

Mark Soghomonian: And then she introduced me to the JGI, the Joint Genome Institute, and the

opportunities they had for undergraduates to work on genome analysis.

Angela Soghomonian: Dr. Maitra told me that this project was kind of in the beginnings of its stages. And she just asked if we wanted to try it out. And we said, sure, why not?

Mark Soghomonian: Which was really good for us because we learned what resources were available and how to use them.

Menaka: And that paid off in a few ways --

Angela Soghomonian: This was our first paper that we got published.

Menaka: F rom an instructor's point of view, Kalyani sees lots of benefits to this kind of project.

Kalyani Maitra: This relates to many areas that they work on nowadays, they can equip themselves to join the workforce or continue to higher education. They will know how to handle data, work through it, application of bioinformatics, where it leads to.

Menaka: And publishing a paper as an undergrad is a pretty nice result -- one that's normally out of reach for many students. Here's Matt Escobar. He's a professor at CSU San Marcos who worked with the JGI to set up Adopt-A-Genome.

Matt Escobar: It opens up a different kind of research, It is maybe more amenable and accessible to a broader array of our students. So a lot of our students, you know, are working 20, 30, 40 hours a week outside school. They can't commit to, you know, 8, 10, 20 hours a week in the research lab, but they can commit to a semester of intensively reading some papers and doing some, some computational analysis. So they're getting research experience “light” and they're getting something out of it in terms of a publication at the end, JGI is getting something out of it.

Menaka: Kind of a win, win, win -- so today we'll get into how Adopt-A-Genome got started, and what it was like for Angela and Mark to take on their genome sequence.

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 a more sustainable future. I’m Menaka Wilhelm.

Menaka: Today, we're diving into a project called Adopt-A-Genome, where students in the California State University system have taken on analyzing and publishing about a genome sequence that JGI has produced.

Menaka: And JGI publishes thousands of genome sequences every year -- all with the goal of understanding how organisms cycle nutrients, create compounds and materials, and break

down waste. Then our users and researchers publish papers on what they find, so that one day all of this data could build new knowledge and help create new technologies that draw on biology's strengths.

Menaka: So the genomes that undergrads are adopting are sequences that JGI has produced, and released, but never published. And to cover how this project got started, let's get into how that happens -- how a genome becomes adoptable.

Menaka: I called up Rekha Seshadri to find out about this. She's a computational biologist at the JGI who got this project off the ground. And she told me, these adoptable genomes come from big, ambitious projects where JGI users set out to create lots and lots of sequences. When these projects get approved, the JGI sets aside resources and time to sequence a set number of genomes, say, 1000, in a sequence allocation. Here's Rekha.

Rekha Seshadri: I'll just give you like a small example that I can think of off the top of my head - we were supposed to do a thousand and then we were waiting, waiting because, you know, things, you know, don't always go according to the timeline, but we published 824, and then we still had the sequencing allocation, and the genomes came in. But they no longer , the umbrella project was over.

Menaka: And this is where these stragglers crop up -- because, JGI still sequences those samples that show up late -- and they still end up in public databases, but they don't get a literature citation.

Rekha Seshadri: So we're not going to publish individual genomes, typically. And most of these are part of the Genomic Encyclopedia of Bacteria and Archaea. That's been sort of a flagship project for the JGI, where the idea was to increase the phylogenetic breadth of sequence, available sequenced genomes.

Menaka: So that Genomic Encyclopedia of Bacteria and Archaea was about sequencing far and wide, to build out wider family trees for microbes. These are high quality sequences that represent new parts of the microbial tree of life. And having them written up in the scientific literature helps other researchers find and learn from them. But for a long time, it's been hard to find a way to publish these straggler genomes. That changed with this Adopt-A-Genome project -- but Rekha couldn't have predicted that.

Menaka: This project has turned out great -- but like many good things, it wasn't totally master-planned.

Menaka: At first, it was just a spark of an idea at a workshop that Rekha helped put on for CSU faculty members.

Rekha Seshadri: I mean, we hadn't conjured up Adopt-A-Genome at that point. Our only objective was to educate the educators and hopefully get them comfortable with the idea of

using genomic science in the classroom.

Menaka Wilhelm: Two of those educators at the workshop -- were Matt Escobar from CSU San Marcos, and Kalyani Maitra from CSU Fresno, who we heard from before. Here they are.

Matt Escobar: So there was about, I don't know, between a dozen and 20 CSU faculty, that attended this workshop that was taught by Rekha Seshadri and Natalia Ivanova from JGI.

Kalyani Maitra: And it was a two day workshop. Rekha walked us through all the different procedures that is needed, how we can teach our students.

Menaka: They covered a data portal that the JGI has created called IMG - Integrated Microbial Genomes. It also includes analysis tools that undergraduates at Cal State Universities might use for coursework or research. So that was very useful, but it's not what sparked Adopt-A-Genome. Here's where that spark happened.

Matt Escobar: And at the end, Rekha was mentioning this problem that y'all have at JGI. Which, a good problem to have, but it's essentially that you've been involved since the early 2000s in so many, uh, microbial genome projects that you have at this point, thousands of sequenced, microbe genomes, which have been deposited in sequence databases, but have never been sort of formally described or reported in the primary literature.

Menaka: These are the stragglers that Rekha mentioned before -- she's been interested in getting more of them published for a long time. It would be a bit of an undertaking, but this seemed like a good audience to pitch.

Matt Escobar: One of the things about the CSU is we have, oh, collectively between 23 campuses, between 400,000 and 500,000 undergraduates.

Menaka: So, that's quite a pool of people who could be interested in creating reports of these unpublished genomes. At the same time, obviously not all of these undergrads are interested in genomics, and writing a peer-reviewed report is a pretty big project.

Matt Escobar: So you know, there was this initial doubt, but as soon as we got to discussing among the folks that were microbial genomics people in the room about these, these mini papers that MRA publishes. That was the link that really allowed it all to come together

Menaka: So these are short papers called Microbiology Resource Announcements published by the American Society for Microbiology. They're length limited -- to around 600 words, which is enough space to describe a single genome, but approachable for an undergrad.

Matt Escobar: And at that point I really advocated strongly for it and we got sort of our core group both from JGI and the CSU that, that put together the, the shared information and we started adopting genomes.

Menaka: So Matt could see this working well for students -- and Rekha too. Because as research students take these genomes on,

Rekha Seshadri: They are basically furnishing all the details that you would need, to sort of, understand, the quality of the data and, you know, where the data came from or all the different processing tools that are employed in generating that genome sequence, the complete genome assembly,

Rekha Seshadri: The advantage of these sort of resource announcements is, is to give that level of detail, right?

Menaka: The idea is to make everything crystal clear, so other researchers can continue building more high quality work from these sequences. With more information, they can later create better comparative papers, or include more breadth in their work.

Menaka: And this framework is working -- when this episode comes out, 3 of these reports have been published. There's another report that's in-press, with nearly 20 more on the way. Faculty at 11 different Cal State Universities have taken on some of these genomes with their students.

Menaka: And Matt has done lots of work to build instruction manuals and resources for these teams, and to recruit people for the project. So -- let's get to how they've set up students to adopt these genomes. The way Matt explained it, adopting a genome is claiming a genome -- saying,

Matt Escobar: I'd like to work on this genome with undergraduates and we'd like, ideally, to be able to publish about it at the end.

Menaka: And naturally, I wanted to know what that claiming looked like.

Matt Escobar: Yeah. So, basically JGI, Rekha, Rekha Seshadri specifically, put together a giant spreadsheet which contains all the genomes that y'all have sequenced that are available for, for this project.

Menaka: So within that spreadsheet, is it like a Google sheet that people are just in with you? Like are you choosing with, like, Anonymous Llama?

Matt Escobar: It's so funny. I mean, it is extremely low tech. So, um,

Menaka: We're laughing, but this is working well -- it's a spreadsheet in GDrive that researchers can access by contacting Rekha or Matt. And the information inside the spreadsheet is definitely not low-tech.

Matt Escobar: There are 1,900 genomes on there and currently, I think about 120 of them have

been adopted, and that's by about 11 faculty. Some faculty who have claimed genomes, it's like, there's one genus that they're super interested in, they claim all the members of that genus and they're working on that in their research lab with their research students.

Menaka: So in those cases, faculty are folding this project into their existing research work, choosing microbes that are related to what students might already have worked with. But they're all working a little bit differently, and there's room for more teams to join.

Matt Escobar: I suppose this podcast is a way to get the word out that, hey, if you're , if you're interested and you're, you're committed to, training undergraduates, you know, this is not an exclusive club. So, you can contact Rekha or I, and, and, uh, we can get you up to speed.

Menaka: We'll have contact info posted in our episode description, and on our website. This program is set up to target Minority Serving Institutions, like Cal State campuses. And next, we'll hear from Angela and Mark Soghomonian about their adopted genome. But first, a quick break.

Allison:

Each year, the JGI accepts proposals for functional genomics research – and there is a deadline coming up. We're currently accepting proposals until January 30th for our CSP Functional Genomics Call.

This call is for researchers who translate genomic information into biological function. To support these projects, the JGI offers a variety of DNA synthesis and state-of-the-art omics capabilities. It's a great opportunity for researchers, but of course you don't have to take our word alone – here's Tae Seok Moon , a professor at the J. Craig Venter Institute in San Diego. He's currently working on a project with support from the JGI's DNA Synthesis program.

Tae Seok Moon: My lab's goal of research is to solve global problems, including pollution, climate crisis and, of course, the waste problem such as plastic waste.

JGI is very, you know, important for my research because I wanted to solve the plastic problem, but to do so I need to find some useful, efficient, you know, enzyme to break down plastic,

So we wanted to find fundamental, generalizable approach to develop or discover new enzyme with the support from the JGI.

I submitted my proposal to DNA Synthesis Program. And then, and that's how I, you know, uh, joined the JGI, you know, network.

So basically the goal of my project is utilizing computational design algorithm to design the enzyme with potentially high activity on all the plastic and then use the DNA synthesis, you

know, capability of JGI to make all the enzyme DNA synthesize and then we want to test all of it.

So I'm so excited.

Allison:

You can find out more about submitting proposals to the JGI on our website. Head to jointgeno.me/proposals. We've also got a link to our website waiting for you, wherever you're listening to this episode — either in the episode description, or the show notes.

Menaka: This is Genome Insider, from the Joint Genome Institute. As we head back to our Adopt-A-Genome project, a quick refresh: Adopt-A-Genome is a project to get genomic stragglers the literature references they deserve -- so other researchers can include them in their work. So professors and research students across the California State University system are taking on unpublished genome sequences, analyzing them, and writing published reports.

Menaka: In this half of our episode, we're rejoining Angela and Mark Soghomonian -- two students who successfully adopted a genome and published their paper earlier this year. They worked with Kalyani Maitra at CSU Fresno. So they'll take us through their process of adopting a genome.

Menaka: As their first step of this project, they had to pick a species.

Menaka: Kalyani's group works on algae and the nitrogen cycle, so she helped them narrow their choices to bacteria related to the work they'd already done in her lab. Here's Angela Soghomonian.

Angela Soghomonian: There was a huge spreadsheet of a bunch of different bacteria,

Menaka: That's the 1900-entry spreadsheet that Rekha Seshadri put together at JGI, of all the adoptable genomes. And here's Mark Soghomonian.

Mark Soghomonian: There was lots of, back and forth conversations. Um, we tried to search for keywords in the genus name or even the species name. And we found that nitrobacters were likely participating in nitrogen cycles.

Menaka: Bingo -- bacteria related to what they're already looked at in lab. A promising group.

Angela Soghomonian: We ended up picking Nitrobacter vulgaris. And this is a nitrate oxidizing bacteria involved in the nitrogen cycle, but specifically Nitrobacter vulgaris is an oxidizer and it oxidizes nitrite to nitrate.

Menaka: So their next step was to gather information about the species that had been sequenced.

Angela Soghomonian: And this specific bacteria was actually isolated from a water facility in Germany, specifically from a sand filter, so that was kind of interesting. But that is where other, that is a similar environment where other species and strains of this bacteria have been found, so not surprising.

Menaka: And they also took a look at the literature to see if this species had been described -- even though its genome would be new in a paper, people have been looking at bacteria for a very long time.

Angela Soghomonian: We were taking a look at the papers that we believed were the first scientists to notify or discover this bacteria, they did the earliest research on this bacteria, and that's still relevant today. That's how we identify certain bacteria and verify its abilities, what it can do, how it survives in the environment, so yeah, we were reading all of those really old papers that describe this bacteria.

Menaka: Matt Escobar from CSU San Marcos recently did a report with a few students as well, and he agreed -- combing through old literature was fun.

Matt Escobar: So we're sitting that down on a Friday afternoon, reading papers, one from 1956, one from 1960. And, you know, it was just tons of fun. We're basically using this material that was generated in their grandparents' time.

Menaka: So with information about how this species had been discussed before, Angela and Mark set to work on their genome analysis. As they started that, Kalyani Maitra encouraged them to broaden their view out to how this species would matter to a wider field. Here's Kalyani.

Kalyani Maitra: The main thing that we had to think about was, how this particular bacterium would be useful.

Menaka: And they kept in mind that some species are useful immediately, and others become useful much later on.

Angela Soghomonian: A lot of these bacteria, especially on the whole list that JGI gave us, most of them have very few, if any, strains that have the entire genome sequence.

Angela Soghomonian: And some only have parts of the genome sequenced. And for many bacteria, we aren't sure of what they're fully capable of, in the sense that they may have some genes that would be significant or of use to something else. And we don't even know that yet. So really just being able to analyze the entire genome and compare it to other related species or other strains and see how they differ. gives us insight to a whole new world of: What genes does this bacteria have that could be for a potential use that we don't even know about?

Menaka: They'd get to highlight a few of those genes or traits in their report -- and they used

JGI's tools to scour their bacteria's genome for interesting tidbits. In particular, they used a data portal that JGI has developed with lots of tools -- Integrated Microbial Genomes, or IMG.

Angela Soghomonian: Dr. Rekha Seshadri, actually, and Dr. Matt Escobar, they introduced us to the IMG website, which is a whole database of genome statistics and anything related to the genome about what genes it has or how to compare it to other species or strains. And we used that website practically for everything.

Menaka: And of course, this was a team effort. Here's Mark.

Mark Soghomonian: The IMG was crucial to our paper. It was really the crux of it. We couldn't have published it without it because you can know about your microbe with the genome and all the information you have about it. But with the IMG database, you can know much more about it comparing it to other known genomes.

Mark Soghomonian: And that was what really made our paper viable. And it was nice to learn how to use it and to be introduced to it.

Menaka: So specifically, when they were doing this analysis, they were curious about what their bacteria might be doing that other similar strains weren't. So they compared the similarity of this bacteria's DNA sequence to other strains using IMG's tools.

Angela Soghomonian: We ended up finding some interesting differences in this particular strain compared to other strains during our analysis.

Mark Soghomonian: And we noticed that there could be a nitrous oxide reductase gene. We had to report that this gene is predicted because within our capacity of analyzing the data and experimental capacity, we did not measure if its transcription and translation actually produces that enzyme, but that was a very interesting thing we noted and we didn't see it in other strains.

Menaka: So that ended up being something they put in their report. And on the whole, Mark and Angela both really liked working on this project.

Mark Soghomonian: I liked doing the genome analysis, and I liked that prospect of getting introduced to it compared to undergraduate classes, they mention it and you learn about it, but it was nice to actually be a part of one. It makes it feel more tangible to you and it gives you more confidence if you want to move forward or dive deeper into it.

Angela Soghomonian: This project overall was a really nice way to get students familiar with how to publish a paper. And how to do a literature search on your own and really how to use online databases, especially for genome analysis. How to interpret what you're seeing, and how to find and search for the correct genes and statistics.

Mark Soghomonian: And it's very translatable. You don't only have to work with microbes. I

know that IMG is big on microbes, but it's a really convenient tool to be familiar with and to even translate it into other fields that you work with genetics with.

Menaka: And, of course, the instructors who are bringing this project to students agree -- this is a cool project to take on as a research student. Here's Matt Escobar.

Matt Escobar: It's one thing to talk about it and read a paper, and it's an entirely another thing to do a literature review of the type you need for something that's going to be peer reviewed.

Matt Escobar: Or to, to start, mastering some of these computational tools. So, you know, students are gaining an ability to in depth query and read the literature. They're gaining some computational skills in genome analysis. And then of course, at the end of the cherry on top is they're actually producing a product that can be submitted to a scientific journal, peer reviewed, and ultimately published.

Menaka: Now that she's worked through one report with Angela and Mark, Kalyani is excited about bringing Adopt-A-Genome to a small course that she teaches.

Kalyani Maitra: Since I'll have 15 students, I would put, divide them into – that's my thought right now – divided into groups of five for this paper. Initiate this module because they need to know how much resource is out there, the Department of Energy, you know, JGI has in their portfolio that can cater to a bigger cause.

Menaka: And that's exactly what Rekha Seshadri had hoped for when she led that workshop for CSU faculty, before Adopt-A-Genome was really even a project.

Rekha Seshadri: My objective really is to bring IMG and JGI resources more front and center.

Menaka: Because as Rekha has worked to build these tools -- those same tools in IMG that Angela and Mark used for their analysis -- she's seen genomics change, and from her vantage point, those changes should help welcome more researchers in.

Rekha Seshadri: I mean, you know, I got into this field when it was a new thing, right? You know, microbial genomes was, a bacterial genome was about 2 million dollars to, to sequence. So it was pretty specialized back then.

Rekha Seshadri: I think it's gotten to a point, the field, the supporting tools, I think it's been democratized. The barrier to entry has been lowered, and, and I'm hoping that, you know, through this CSU Alliance and people like, you know, like Kalyani and Matt are awesome because they're also building these little workflows and work plans and, and making it available to all the other CSU faculty to use.

Menaka: And making a new entry point for more new researchers to contribute could eventually mean lots more new discoveries, since basic science builds on itself over time. So these

adopted genomes could be starting points for all kinds of new knowledge.

So again, that was Kalyani Maitra from California State University Fresno; Matt Escobar from California State University San Marcos. Rekha Seshadri is at the JGI, and these days Angela Soghomonian is studying medicine at UC Davis, while Mark Soghomonian studies medicine at California Northstate University.

We'll link to contact info, and more information about Adopt-A-Genome in our episode description. You can also find 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!
Show Notes

Submit your own proposal to work with the JGI

For more information about Adopt-A-Genome:

Rekha Seshadri: rseshadri@lbl.gov
Matt Escobar: mescobar@csusm.edu

Adopt-A-Genome Papers:

Draft genome sequence of Nitrobacter vulgaris DSM 10236T

Draft genome sequences of Butyrivibrio hungatei DSM 14810 (JK 615T) and Butyrivibrio fibrisolvens DSM 3071 (D1T)

Genome sequences of key bacterial symbionts of entomopathogenic nematodes: Xenorhabdus cabanillasii DSM17905, Xenorhabdus ehlersii DSM16337, Xenorhabdus japonica DSM16522, Xenorhabdus koppenhoeferii DSM18168, and Xenorhabdus mauleonii DSM17908

Our contact info:

X: @JGI