

Menaka: Today -- an interview with two researchers who have studied a very sneaky, but influential part of a microbiome -- the pieces of DNA that can move to-and-fro between different organisms. These are mobile genetic elements or MGEs for short. And we want to understand them, because as they move around, they impact entire communities of microbes. Sometimes they bring an organism a totally new ability, other times they power up or power down their existing genes. And all of that matters if you're trying to get microbes to work on your projects rather than just their own. Which, increasingly, we are.

Here's Sarah Bagby, one of the researchers in this episode.

Sarah Bagby: If you're trying to -- as the DOE is, right -- be able to precisely engineer a microbiome for biofuel feedstock degradation, it's really important to understand whether these critical degradation capabilities are going to be stable in the community. And it looks to us like understanding mobile genetic element activity is gonna be a key part of that.

Menaka: But this is new territory. For a long time, it's been tricky to try to pin down these mobile elements.

Sarah Bagby: We have had so many good reasons for not trying to analyze these elements, right? They have been intractable. And it's still pretty rare for them to be subject to this analysis. But think how much we're missing!

Menaka: The answer is, a lot. So Sarah and a whole crew of other researchers have been working on understanding these mobile genetic elements for a while. They have a paper out this week in Nature Microbiology.

In this project, they combined high quality sequencing from the JGI, and a long term dataset at their study site in the north of Sweden, Stordalen Mire. Together with some new analysis, they've gotten a clearer view of many of these moving pieces.

Sarah Bagby: We have finally got some bounds that we can put on just how active these things are, just what scale of impacts they might have, just what range of functions they might affect. And that really tells us that we need to structure our sampling to go after capturing the mobile element part of the story when we're looking at microbiome response to change.

Menaka: So we're covering what makes mobile genetic elements so interesting, and so elusive, how they managed to see many of the mobile genetic elements in their data more clearly, and what they want to do next.

[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.

And we're talking about the work behind a paper in Nature Microbiology. It's called 'Mobile genetic elements shape microbial diversity and functions in thawing permafrost soils'. This is a project on samples from a sort of research nature preserve way up in the north of Sweden, called Stordalen Mire. It's a permafrost region that's thawing these days, and so the microbial community has changed a lot in the last decade.

We know that, because researchers have studied this site for much longer than 10 years. The JGI, also, has supported this work across multiple project proposals and collaborations. All of that makes for data that could help us engineer microbial communities to do lots of interesting things, from breaking down plants, to creating biotech products. And all of this information, from the JGI – is high quality, and legible for both researchers and AI models.

So in this project, a team took the opportunity to understand mobile genetic elements in this changing environment.

And they were able to get a new view of some of the moving pieces in these microbial genomes, even though these genes have often hidden in plain sight. They used an 8-year dataset, deep sequencing of those samples from the JGI and some clever analysis to see some things that surprised them.

So let's meet our researchers.

Sarah Bagby: My name is Sarah Bagby. I'm in the Department of Biology at Case Western Reserve University in Cleveland, Ohio. So I'm an environmental microbiologist, but I got here via structural biochemistry, and I really like thinking about everything from molecule scale up to big system scale and how the interactions at all of those levels in between really make things work.

Simon: I'm Simon Roux. I am a staff scientist at the Joint Genome Institute, and I am part of the metagenome program. So my work is really focusing on trying to explain and characterize microbiomes from large scale shotgun sequencing.

Menaka: Sarah Bagby and Simon Roux are two of the three senior authors on this project, Matt Sullivan at Ohio State is the third. Next, let's meet their study subject: mobile genetic elements.

Sarah Bagby: The basic thing that makes something a mobile genetic element is that it is a piece of DNA that can move. So it can move potentially from one place to another place within a genome or sometimes from one genome to a different genome.

Menaka: There are a few different kinds of these gene jumpers – some of the genes that move around called cargo genes – they bring entirely new tools into an organism's genome.

Other mobile elements work differently, almost like a switch or a volume knob that affect a microbe's existing genes. Just defining those elements, and the list of them, the mobilome, is a bit of a moving target. Here's Simon.

Simon: Every time we define a box, something straddles a boundary or like – it's, it's endless. So for instance, in this paper, we define the mobilome based on the way we identify these elements, which is – we looked for genes directly responsible for integration into the genome.

The flip side to this is we, we know and we acknowledge this, we miss a whole part of the mobilome that is all these elements that are able to move from cell to cell but never integrate.

In this case, we looked at four categories that are called integron, conjugative elements, transposase, and then the one that are looking like phages or viruses. These four categories have four different ways essentially to integrate into the genome of a host. And so when we look for them, we look for four different types of genes, broadly speaking.

Menaka: Within these limits, you basically still get a galaxy of possibilities – it's part of why Sarah and Simon both have studied mobile genetic elements for years.

Sarah Bagby: One of the things that I think is so cool about these elements is that, you know, we know in all kinds of systems, you know, eukaryotes on down, that the effect of a given genetic change really depends on what we call the background.

What is the other set of things encoded in that organism? That's gonna be sort of the canvas on which, you know, the impact of one of these elements plays out. Microbial populations are so large, that as one of these elements moves from cell to cell, the same element might have different impacts because, you know, it's in different contexts, both its own immediate context in the host genome and also the broader suite of genes that are encoded by that genome.

And so, like, just thinking about the combinatorial potential of these things to just sample different ways of being, right? It's enormous. It's really cool.

Menaka: And even if we don't have a complete map of all the types of genes that can move between organisms, we know they're important.

Sarah Bagby: We know for sure that they drive a huge amount of the gene flow that allows these communities to respond to change.

Menaka: These are a special piece of a microbial genome, different from the genes that a microbial species carts around by default.

Simon: Their pace of evolution is really different from a regular rest of the microbial genome. And so that's really what we were looking after, is like all these things that are more agile and, and maybe more variable and, and we can finally look into them.

Menaka: He says “finally” there, because it's relatively new to be able to look at these. I asked Sarah why MGE's, historically, have been so good at escaping analysis.

Sarah Bagby: Yeah, absolutely. Oh man, it's exactly the reason that they're so interesting, right? So okay, just for a little bit of context, when we go out and sample a complex community for sequencing, it's very, very different from when we try to sequence a pure culture in the lab. Like in the lab, we've got a pure culture.

There's just one genome involved, and you can grow up however big a culture you like and extract as much DNA as you like, and you can sequence every last nucleotide. It's fine. But in the complex community, right, we have to do what's called metagenomics. We extract DNA from whatever mixture of microbes is present.

We don't know what that mixture is. And we break the DNA into small pieces, and then we sequence a random subset of those small pieces. And then when we get the data back from JGI, uh, we have to work out which sequencing reads arose from which genome.

Menaka: This is called binning – sorting sequencing reads into bins for each genome. Like sorting laundry, if instead of clothes, you had bunches of pieces of DNA to match up. And instead of a light and dark laundry basket, you had a Chloroflexota bin and a Verrucomicrobiota bin.

Sarah Bagby: So there's a ton of neat biology here in the signals that we use to allow us to bin these sequences. What it all boils down to is that, that genomes sort of have signatures, right? They look like themselves, and those signatures allow us to say, “Aha, you know, you are thing one, and this other piece is thing two.”

But mobile elements don't have their host signatures. Mobile elements are their own thing. And so if you've got a mobile element that has managed to get assembled into a long stretch of DNA, it kind of breaks the signatures of that piece of DNA. And so those pieces fail to get binned, and so then we don't see them.

And then the last thing that happens, like if you have had something that makes it in, the last thing that happens is that we have to reduce the complexity. So we might have many different versions of a species genome that's represented on the pieces of DNA that we managed to put into a bin.

But we have to build one consensus sequence, like what is the predominant way of being for this species in our dataset? And that means that we have to collapse all of this variation. And so we might collapse it in a way that has the mobile element. We might collapse it in a way

that doesn't have the mobile element, right? That's gonna mean that again, we lose a lot of these really interesting dynamics.

Menaka: Yeah, sure. So just algorithmically handling mobile genetic elements is like a whole other layer of complexity.

Sarah Bagby: Oh my gosh, yeah.

Menaka: Yeah. But then when you have this long time series, you do get to make comparisons where you could follow these mobile genetic elements.

Sarah Bagby: Absolutely. Yeah, exactly. Which is so great, right?

Menaka: Yeah, so, how does a long term time series sort of solve those problems for you?

Simon: Yeah, I actually thought about this a lot in terms of how to explain it best, so we have a few ways to – we look at activity, what we call activity, broadly speaking, of MGE. But the one that is directly linked to a time series is essentially a game of red light, green light.

Menaka: So that's the game where kids are meant to run around, but they can only move when a caller yells "green light", with their back turned. As soon as that caller yells "red light!" and turns to face everyone, the kids have to stop running.

Simon: Which I actually had to go and search a translation. Because it is not at all the same in French. In French, we call it one, two, three, sun.

I never thought about this. It's like you just do, one, two, three, sun. That's just, of course.

<laughs>

Anyway, but the bottom line is this really is a game we are playing with these guys. We are looking at the time series, and we are comparing samples through time, and we are catching movement by seeing them change places. And we are very rarely changing, like we are very, like very rarely are we actually catching them right in the act of moving.

But we can see like, "Oh, a year ago you were here, and now you are all there." So something happened. So that's really the way time series is, is becoming super important and, I would say it's a transformation for the field. This is not the first time series. I mean, I remember you talked to Trina and Robin about the long-term time series in Lake Mendota, which is another really, really great time series.

Menaka: It's true! That's Trina McMahon and Robin Rohwer, who worked with a team on a metagenomic data set from freshwater Lake Mendota in Wisconsin. We did a 3-episode series on it, and we'll link it in the show notes. Back to Simon --

Simon: But what I mean by transformation when I started my work in this field, because of the limitation in terms of technology and, and the cost of sequencing and all these things, we were really taking snapshots. Maybe two time points. Hey, you know, for this year in this lake, I have a fall time point, and I have a spring time point. Great. That doesn't tell you a whole lot about changes and dynamics in your environment.

But right now we are starting to, you know, generate this kind of datasets that really transform these kind of, fixed pictures into almost flip books.

Menaka: Yeah. And incredible that the sort of the ability to process this much data and store it and, I can only imagine how much of a change that must be.

Sarah Bagby: Oh my gosh, absolutely. I mean, I think it, I cannot overstate how important it has been to have like sustained long-term investment from DOE and also from NSF in making it possible for us to even envision the kind of sampling campaign, let alone carry it out, you know, that has let us build this really globally unique data set that is letting us, you know, ask questions that nobody's ever been able to ask before 'cause the data just wasn't there.

Menaka: Yeah, definitely. And I would love to talk about the sample site.

Sarah Bagby: Yeah, so it's, it's a really neat site. The site itself is called Stordalen Mire, and it's located at the Abisko research station in northern Sweden. So it's just like, part national park and part research station.

So Stordalen sits at what is called the discontinuous permafrost margin. So sort of on the southern edge of the permafrost zone where it's, it's kind of patchy. And we can see habitat change on the timescale of a single PhD thesis. We go out and we build boardwalks that will let us not disturb the site when we sample, and then we go back a few years later, and those boardwalks are underwater.

Menaka: So it's an environment that evolves as that permafrost thaws.

Sarah Bagby: So there are sort of three canonical stages along the thaw gradient of the site, and we see them sort of over and over. They seem to be kind of stable states, in this process of thaw.

So the intact landscape is what we call a palsa, P-A-L-S-A. It's not a word most people know. And it's sort of hummocky high ground. The soil is really dry. It's underlain by intact permafrost, and there's this lovely diverse plant community. The microbes in the top layer of soil are seeing an aerobic environment, Life is, is pretty good.

And then as the permafrost underneath starts to thaw, you know, all of a sudden this thing that's been holding the land up turns into a puddle. And so it's kind of slumping down into

what is now a wet soil column. And what that means for the microbes in the soil is that they're going back and forth between aerobic and anaerobic life.

And so like those are big fundamental changes, to a microbe.

And then the last stage in the thaw is a fen, where again, there's a different plant community and a different microbial community.

Menaka: So you're watching this change on kind of a time scale, but also on a pH scale, and also on an anaerobic, aerobic scale, also on a water scale,

Sarah Bagby: Yeah, there are just so many axes, right?

Menaka: Yeah. Yeah, but I mean, that sounds like a great site for that reason, so.

Sarah Bagby: Yeah, absolutely. I mean, It's one of these things where you think, "Oh, it's really unfortunate that this study site exists." During my postdoc, I did some work on oil spills. I'm like, "Oh man, this is not an experiment that I could ever, ever do. But since it's here, let's see what we can learn about it."

Menaka: Yeah, yeah, sure.

Menaka: So with great sequencing and assembly, plus this super dynamic, long term site -- they got to really ask questions about the genetic elements that moved around in these microbial communities.

Simon: And so the question was, okay, we know we have changes, we have stresses. These microbes are not living their best life all the time. That's not possible. It just changes too much.

We know we have these elements in there that can, again, be the one that are agile, the ones that confer some trait and then move them here, move them there, kind of influence, can kind of like really toggle a lot of switches, really kind of push a lot of buttons.

And we wanted to know more about like, how much do they do actually, that. We know it's possible, but can we observe it, even?

Menaka: Let's get to what they did, and what they saw. Here's Sarah again.

Sarah Bagby: We had soil cores that were collected from each of our habitat types, in every year for eight years running, and that were sectioned by depth. And so we had a lot of samples to play with.

Menaka: This team worked with about 600 metagenomes to get this paper together -- so all from samples taken throughout the year in that site in the north of Sweden.

Menaka: Can you outline what's new about the ways you analyze this data?

Simon: So we kind of started back almost from scratch. And we actually don't use too many existing off-the-shelf tools because we want it to be pretty comprehensive and also, in a way, the naive approach was, in our mind, the right way to go here.

So I mentioned it briefly before, and I can explain it a bit better. We decided to go after these mobile genetic elements that are able to integrate into the genome of their host.

And so the good news is, before our work, the lab of Peer Bork, actually in, EMBL, they had done this kind of survey, exactly the same kind of survey on a collection of isolate genomes. So they had taken all these genomes, they had designed these models to go after, what we call again, integrase, broadly speaking, which has – which are genes that let you integrate into a genome.

So we had this starting point, and as far as we could tell, no one had really taken this framework and applied it to a natural community, especially not as complex as soil. And, and that's really where we kind of got a lot of the novelty out in the sense of we were able to kind of gather this very large catalog of MGEs from this extensive time series by combining this method that was developed initially for isolates and adapting it to this pretty exceptional data set that we have from this soil microbiome.

Sarah Bagby: That was great. That was, that was a huge help. That allowed us to identify, you know, how many recombinases we could detect. It didn't let us say, "How many did we miss?"

Menaka: But they did want to look at that!

Sarah Bagby: Because as we were talking about before, we're gonna miss a lot. And so we also did some benchmarking, where in addition to the sequencing that we had done on that, you know, near decade of data, we took a small number of samples and used a different sequencing approach called long-read sequencing, that is not as prone to some of the issues that short-read sequencing, like we typically do, is prone to when it comes to mobile elements.

And so we were able to compare our assembled and binned sequences from the small set of samples where we had both long-read and short-read sequencing to say, "Okay, can we bound how bad the problem is? Can we say how many we've missed?" Right?

Which is really useful information that we haven't had before, right? and so, you know, we think that typical metagenomic approaches miss somewhere between one-third and two-thirds of mobile genetic elements, which is, you know, is a fair chunk. But now we know that. And so that was, that was the first piece.

Menaka: The next piece was asking two big biological questions.

Sarah Bagby: One is, what are the functions that these elements affect? And one is, how active are they? We say that they're mobile, but how much, right? Because it's a very different scope for what their possible impacts could be if they're mostly sitting tight and, you know, every so often they move. So our initial analysis of how many are there, together with our benchmarking, leads us to conclude that in the wild, a typical microbe is going to have somewhere between 38 and 66 mobile genetic elements in its genome, which is a lot!

That is a whole lot. And so we kind of filed that information away. We looked at the functions next, to see like, what genes are being affected? And I think I mentioned before, you know, we were trying to be super, super strict about saying, "Okay, this function was definitely affected by a mobile element."

The reason that that's tricky is that while we're able to recognize the recombinase genes themselves, recognizing the boundaries of mobile elements is really hard. There's not a good way to do that currently.

Menaka: So – they'd found 2.1 million mobile genetic elements – and with the strict criteria Sarah mentioned, they could ID genes affected by about 16,000 of those elements. But then predicting functions narrowed that group again.

Sarah Bagby: And so when we went to look at functions, we were able to look at, oh, a fraction of a percent of the millions of recombinases that we identified. But we thought it was really important to look kind of agnostically. And the reason I say that is that, a lot of interest in mobile elements has come up in the context of antimicrobial resistance, which is a looming crisis for human health. And the reason that it comes up there is that it's very common for mobile elements to move genes that enable microbes to resist antimicrobial drugs.

And, you know, our perspective was, this is an important thing that cells do. So it's really important for us to understand how antimicrobial resistance genes move around, but that is just one slice of cellular activities.

And we don't really have a reason to think that that's the whole story for mobile elements. And so in fact, you know, like we hesitate to make statistical claims because, again, we're looking at such a small slice of these genes, but a little over half of the functions that we found being affected were for what we call sort of 'domestic processes', like getting and using nutrients and energy. That kind of like really fundamental, "Would you like to continue being a cell? You should do these things!" right?

And then, you know, the rest, like there are loads of stress response functions, like antimicrobial resistance and heavy metal tolerance. Absolutely, those genes are there too, but they are definitely not the whole story.

Menaka: The last piece was looking at how active these mobile genetic elements were.

Sarah Bagby: So we had three different ways of doing this that are kind of sensitive to different timescales. When we look at those different measures of activity, we see numbers that, you know, on first glance, yeah, if you say like, “oh, it's, it's 1% of these elements that are active,” that doesn't seem like so much.

Simon: We were, I think, probably a bit surprised, or at least I was surprised. I can say it, I was surprised. Lots of these mobile genetic elements that we highlighted as like, oh, they have the ability to move as far as we can tell, they did not seem to move much.

Sarah Bagby: But then you think back to how common the elements are. So if you've got something like 50 of these elements in a genome, and 1% of them are active at a given time, that's enormous? There's a huge unknown that we really wanna go after, which is, how is the activity that we see distributed – like, Poisson distributed? Or is it like in this cell at one time, absolutely all of the mobile elements are freaking out and expressing their recombinases? Because that would mean, like, huge combinatoric potential, but greater stability in the rest of the community? By contrast with, you know, maybe there's some stochastic firing of these elements, and so at a given time, you know, half of the populations are seeing some genomic reorganization, which is still, like, it's not trivial.

Menaka: Yeah, and is there anything you would say that you had to like set up or, change or that you would share with other people, about setting up these analysis types?

Sarah Bagby: Ooh, that's an interesting question. I think it is really helpful to have a little bit of long-read sequencing. You know, if you can find space in your budget to have a little bit to give you some scaffolding for interpretation, that can really help.

I am never gonna stop beating the drum for reproducible research methods. You know, making sure – like these analyses are so complex, and the code base gets very big very fast. The more you are working openly in your code from early, the more eyes you can have on it, the better.

Menaka: Yeah. Yeah, that makes absolute sense.

Sarah Bagby: But you know, aside from that, it's just having deep enough sequencing at enough time points, that there's going to be a signal that you're able to, to pull out from all of the noise.

Menaka: Deep enough sequencing – as you might guess – is a big way that JGI enabled this work.

Sarah Bagby: Yeah, absolutely. This was huge. So JGI actually did virtually all of the sequencing that was involved in this project. So the collaboration that has been studying Stordalen Mire intensively in these last 15 years or so started off as a DOE-funded consortium. It was the IsoGenie Consortium.

And that actually, was responsible for all of the sampling during the years that are represented in, in the analysis here. So that was sustained investment for collecting and analyzing the samples and JGI, you know, doing the sequencing that we were then able to work with. And one of the things that's really important about having a single facility that could do that sequencing for us that whole time is that, you know, with all of the different methodological hurdles that we had to deal with, we did not have to deal with batch-to-batch variability. That can be a real challenge for long-term projects like this when you say, "Well, I did all the sequencing. It was long-term, it was beautiful, but year one, there's no relationship to year 10."

Menaka: Horrible.

Sarah Bagby: Total nightmare.

Menaka: Yeah.

Sarah Bagby: So, you know, we have worked really hard with the JGI over the years, like, as new sequencing platforms have come online, and JGI has said, "Okay, well, we're gonna switch to doing things this way." They have been great about saying, "Well, we're going to make-- we're gonna do some sort of bridging calibration analysis.

We're gonna take some of your old samples and run them on the new platform. We're gonna take some of your new samples and run them on the old platform." And then we have a way to see, like, to what extent do we trust that what's coming out is representative of the same community and sort of the same filtering processes, that, you know, it was when we started this project.

Menaka: And this work fits in really well with two kinds of projects that the JGI is aiming to do more of. Here's Simon.

Simon: One of these is time series. We already, you know, generated some data for multiple time series, but like I said, I think this is one of the big things coming and one of the major advances and major sources of new biological insight in the next five to 10 years. So we definitely are keen on generating more of this data.

And then we're also thinking of – what kind of analytics do we need specifically? We already have tools to do this kind of comparison and say, "Hey, I found the same microbe here from there," et cetera, et cetera.

But we are very interested in developing more analytics specifically for this recurring sampling of the same place and same microbiome. And, and we want to help people look at the kind of question we are looking at here, even beyond MGE, but how do things change? How can we measure change? How can we distinguish things that change, quickly versus slowly versus oscillating by season, for instance?

All of this is on the table, but right now it's really done by, it's really the beginning, so it's done by ad-hoc scripts and analysis. There is no automated tool. There is no online platform that lets you do it. And so I think that's one of the things we really are interested in at JGI, is generating the data and establishing some framework or helping to establish some framework to analyze this kind of data.

And at the same time, we are also very interested in complementing the description of diversity with a description of activity, and that's my metatranscriptomes. That's my, "Great, I have five thousand microbes. Tell me which one is doing something and what." And so we have had a lot of, you know, a lot of great successes over the last five-ish years at having some of our users, being able to, send us enough samples so that we could generate both metagenomes and metatranscriptomes.

But I feel there is still a lot more we can do, and that's not just generating more of this data, but the same way as for the time series, it's also generating more of the analytic framework.

How can we give the tools to everyone so that they can look into these questions more?

Menaka: Yeah. Yeah, yeah. And I imagine it's so much information that AI and machine learning are basically just baked into any analytical tool you would use.

Simon: I think at this point, that's what we are looking at. The key part will be, how do we, how do we find the right training data for something like AI to really take full advantage of these frameworks? But they are-- this kind of technique, and specifically the machine learning technique and some of the AI techniques are really good at finding the needle in the haystack, and that's a lot of what this is about.

Menaka: And beyond just sort of like, getting an even more detailed picture of what's going on, are there specific questions that you would want to address, within this dataset, or this environment, in kind of the next work?

Sarah Bagby: Oh, we're so excited to do this.

So we talk about it as trying to span scales of time and of space and of complexity, and working across scales, you know, with a combination of observational and manipulative work, and also modeling work to try to fill in the gaps that we're not able to assess with, you know, direct observation of molecules.

We ran a microcosm experiment in my lab a few years ago, where we took peat from the bog and peat from the fen, and then subjected it to temperature perturbations over a period of a couple of months. And we sampled all along the way for sort of everything that we could possibly sample for. You know, DNA and RNA and proteins and metabolites, just all the things, right?

And so we're gonna get a more full picture of activity, and we're gonna be able to tie that activity of mobile elements to, well, what's actually changing chemically, as the microbiomes keep up their activities, right? And is the activity of these mobile elements increasing when we turn up the temperature? Is it helpful? Like, is it evolutionarily advantageous to a strain to have lots of mobile elements, or do they tend to get rid of them? Who's winning and who's losing, and how are mobile elements enabling those changes?

So I think we're gonna be able to use the analysis framework that we built here and then push these questions farther, in ways that'll tell us what to look for when we go back to the field.

Simon: My personal two questions with this kind of data, including in Stordalen Mire are revolving around this specific type of MGEs, which are these phages, these viruses – what drives their host range. And so now that takes us to the next question, at least to me.

One of them is, what are the constraints for a virus in a microbiome in terms of host range? When can they go into this cell and this cell and not this next one? What makes a virus able to maybe enter many microbes, or why would they even do this versus what is the type of selection that pushes a virus to become extremely specialized and infecting only one type of microbe?

And this is not – these are not new questions. These are fundamental questions in biology and virology. But the ability to look at them in this kind of soil microbiome, as we just described, complex microbiome, lots of things happening on a very microscale, very much not a mixed, homogenized environment.

And it ultimately, that takes you to identify these key MGEs or viruses that may be doing or may be having a, a disproportionate impact. And because, I mean, I have my curiosity, I love my viruses, I could do this forever just for the sake of it, but that's, that's just me.

To bring that into a bit more of an applied question, one of the reason why we want to be able to, look at these MGEs and look at this question of adaptation, or in my case, host range and, and activity, is because by better understanding this kind of activity and dynamics of microbiomes, we will be able to predict the microbiome processes better, so having better forecast of what a microbiomes will do if you, you know, change the condition or something like this, and then engineer them.

Because once you understand that, oh, the key parts that really drive this key metabolism are these three MGEs and these four microbes, okay, now I have something to go from when I started from my billions of cells and thousands of genomes. Now I know where I can focus. And so I think that's a lot of what I hope happens downstream, and that's what we want to enable ultimately.

Menaka: In the end, MGEs are one part – a key part – of understanding more about microbial communities more broadly, both in this environment, and others. These years of sampling, sequencing, and analysis show us more about how microbes break things down, produce

other compounds, and cope with change. All things we'd want to engineer them to do, which opens up even more possibilities for the future.

[THEME]

Menaka: So again, that was Sarah Bagby from Case Western Reserve University and Simon Roux from the JGI. We'll link to their recent paper, and an episode transcript in our episode description.

The JGI supported this work through a few different project proposals, including proposals in partnership with the Environmental Molecular Sciences Lab, through the FICUS program. You can learn more about that program, and how to work with us, at jointgeno.me/proposals.

In line with the JGI's work as an AI-centric user facility, as Simon mentioned, we'll also continue supporting users with data, software and analysis that power and leverage AI in service of biotech and bioproduction.

This episode was written, produced and hosted by me – Menaka Wilhelm. I had production help from Allison Joy, Massie Ballon, and Ashleigh Papp.

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!