

Genome Insider S4 Episode 8: The Megadata of Lake Mendota - Part 3: Boating Out to David Buoy

Menaka: Today, our third and final episode on how a mega dataset from Lake Mendota came together. This project took 20 years of microbial samples and sequenced them into 500 metagenomes so researchers could watch long-term change in that freshwater ecosystem. And today, we're checking out the sampling part of this project — so we've got Allison Joy here, because she saw that for herself! Allison, hi!

Allison Joy: Hey!

Menaka: Start us off with an intro to that trip!

Allison Joy: Sure. So we're headed to Madison, Wisconsin, and Lake Mendota.

Krys Kibler: We're heading out to Deep Hole Lake Mendota, the deepest point of Lake Mendota where we do all of our sampling at.

Allison Joy: That's Krys Kibler. They're driving the boat.

Krys Kibler: Let's head on out.

Krys Kibler collects a lake water sample. (Alex Holloway for UW-Madison College of Engineering)

Krys Kibler: I'm currently a PhD student at University of Wisconsin, Madison. I'm in the department of freshwater and marine science, which is awesome. I get to study lake Mendota and I focus on the diversity of cyanobacteria throughout their growing season in the summer and spring months.

Allison Joy: So, I tagged along with Krys in October, 2022. We're catching the lake at the end of that summer growing season. It's a clear autumn day, crisp and sunny. Fall colors are exploding around campus. Folks are out enjoying the weather — and the view of Lake Mendota. It sits near the center of campus, right off the student union. And we're leaving from the Hasler Laboratory of Limnology. It's a research building totally focused on studying lakes, and other freshwater environments.

Krys Kibler: The university is very lucky where the center for limnology is positioned right at the lake, like right on the shoreline. And so we have a couple of boats in a garage that can literally just be driven right out onto the lake surface.

Allison Joy: This boat is a Badger-red pontoon boat, maybe 15-feet long. Not giant, but big enough to get us where we need to go.

Krys Kibler: All right, so we are uh, coming up to the buoy now,

Allison Joy: A buoy marks deep hole and the buoy itself is special enough to have a name

Lake Mendota's iconic David Buoy. (Allison Joy)

Krys Kibler: And it's actually called David's Bowie. We love puns at the CFL and at the university.

Krys Kibler: I can never say like David's buoy, it's always, you gotta say David's Bowie

Allison Joy: Love David Bowie. But remember, this buoy is not just named after a great talent. It marks the deepest part of the lake.

Krys Kibler: So it's about 23 and a bit meters. We are dropping anchor,

Allison Joy: But as you can imagine, there's much more to this sampling. Go ahead and roll the intro, for now.

Menaka: Great – we'll get back to what happens after Krys drops the anchor soon.

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.

And before we get back to Allison's trip out to Lake Mendota, let's paddle through just a bit of background.

Lake Mendota is the site of a very big project for Krys, their labmates, and the JGI. This is a project that's included 20 years of samples from the same lake. At the JGI, those samples became 500 metagenomes. A 25 terabyte dataset — in 2021, it was the biggest dataset the JGI had ever assembled into metagenomes. This turns out to be such a big project, that we've taken three episodes to cover it.

To go behind the scenes of that project, we've taken a look at a few different things that made this work possible. Episodes 1 and 2 walk through the software, and the supercomputing that handled all of those metagenome sequences. But of course, to get a lot of sequences, you also need — a bunch of samples. So this episode is all about collecting, extracting, and analyzing the lake water that became those hundreds of metagenomes.

All of these samples came from the same research group. It's Trina McMahon's lab at the

University of Wisconsin, Madison.

Trina McMahon: I'm kind of an engineer, limnologist, microbiologist, ecologist, genomics-ist.

Menaka: Trina and her students and postdocs study the microbes in freshwater systems, mainly sampling around Wisconsin. And of course, that includes Lake Mendota.

Trina McMahon: We say it's the best studied lake in the world, and,

Menaka: We heard that a lot,

Patricia Tran: Mendota is the most researched lake in the world.

Krys Kibler: That's true. Like Mendota is the most researched lake in the world.

Menaka: Part of that is how close Lake Mendota is to the university – but it's certainly interesting in its own right, too. As an environment, Lake Mendota is buzzing with life — there are expected species,

Krys Kibler: So you have like Actinobacteria, Bacteroidetes and then Alpha and Beta Proteobacteria.

Menaka: And the lake is home to some harmful organisms, too.

Krys Kibler: So it gets a lot of cyanobacteria or harmful algal blooms or blue green algae, in the summer and spring, which is why I like to study it.

Menaka: Those harmful algal blooms are the result of lots of ecosystems colliding at once. Runoff from agriculture, so — fertilizer, often feeds those blooms, and algae affect other species in the lake quite a bit. The big idea of studying Lake Mendota is to understand how all of these organisms — humans included — affect each other, over time.

Krys Kibler: Really, the long term goal is to continue monitoring and really narrow down on why specific bacteria are increasing like Cyanobacteria or decreasing with climate change and human impacts. And really just trying to untangle how our lakes and the microbes within our lakes are changing with how the world is changing.

Menaka: This is a project that takes a lot of different factors into account.

Trina McMahon: We affectionately call it the Mendota mess because it's just got so many different components to try to, you know, connect and pick apart.

Menaka: And there are other collaborators working to pull apart different threads of that Mendota Mess.

Trina McMahon: Movement of carbon, nitrogen and phosphorus through the food web is something that's really core to people, you know, what people study in limnology. So, you know, there's people interested beyond just the microbial ecologist working on it, which is always nice. So, we can collaborate with them and have a better handle on, like, how that's altering the movement of that matter through the ecosystem.

Menaka: Trina's group leads the charge from a microbial angle.

Krys Kibler: Students from Trina's lab sort of use this dataset as well, to look at specific, bacteria groups. So if someone's interested in Actinobacteria, they can really narrow down and like how the Actinobacteria responds to different nutrients or, changes in something else that they're interested in.

Menaka: And with over 20 years of data collected, there are lots of questions to ask and answer.

Krys Kibler: It's one of the biggest, like microbial museums that you could find. And that's what JGI is helping us with is both sort of processing this microbial data and helping us like curate it and really figure out what's in there.

Menaka: So JGI helped out in a few ways — with sequencing, and then assembling this data with specialized software and supercomputers. But — like I said, none of that could have happened without samples.

Let's head back to the sampling trip that Allison Joy tagged along for.

Allison Joy: So to pick up where we left off, we're boating out to collect samples and take measurements. Krys Kibler is our guide, and there are a few other graduate students and interns tagging along to collect samples of their own. Actually, as it turned out, one of them was collecting samples for a different project with Trina's lab.

Patricia Tran: Hi, my name is Patricia Tran and I'm a PhD student in freshwater and marine Sciences at the University of Wisconsin Madison. And I work in the lab of Dr. Trina McMahon and Dr. Karthik Anantharaman. In 2020, we got a JGI New Investigator CSP to investigate the viral and bacterial communities in Mendota. I study bacterial and viral interactions and the roles in biogeochemical cycling, in freshwater lakes, specifically in anoxic lakes, such as Lake Mendota.

Allison Joy: So this is quite a productive boat ride. We're all headed for that deepest part of Lake Mendota. Remember, right near David Buoy. Here's Krys.

Krys Kibler: So yeah, we sample right by this buoy for, you know, 20 so on years, and basically take a microbial snapshot of the lake.

Allison Joy: That microbial snapshot comes from lake water samples, and also measurements of temperature, pH, and dissolved oxygen.

Krys Kibler: We take a bunch of depth discrete measurements from zero to twenty meters deep, and measure at every one meter interval.

Allison Joy: The deeper you go, as that temperature and pH changes, so do the microbes that you find. So the way they sample lakewater preserves that variety – it keeps microbial communities stratified the same way that they live in the lake. Think of the lake a bit like a high-rise apartment, some microbes live up in the penthouse, other microbes hang out in the lobby, and of course everyone in between. To get samples with everyone in their respective spots, the researchers use — a really long tube.

Krys Kibler: So basically this tube is zero to 12 meters long. So basically we just lower it into the water, collect that chunk of the water column and bring it back up, basically.

Allison Joy: This kind of sounds like it wouldn't work, but it does – the tube drops down, and as it goes deeper into the lake, it grabs a floor-by-floor slice of the water. It operates the same way that you can hold soda in a straw by putting your finger over the top of it.

Krys Kibler: We put a stopper at the end of the tube just to make sure the pressure holds that water in place and we bring it up back onto the boat, take out that stopper and just pour it into one of our sample bottles.

Allison Joy: In the lake, the tube gets a little heavy. But it's a method that's stood the test of time. They've used the same exact way of sampling for over 20 years, for consistency.

Krys Kibler: Yeah, that's the 20 year old tube that people ages ago — ages ago. I'm just kidding.

Patricia Tran: Ages, ages ago.

They've been using it 20 whole years.

Yeah, it's longer than I've been alive...

Allison Joy: Oh my god. I just realized that means they started using it when I was in school here.

Krys Kibler: Ah,

Allison Joy: Oh my gosh.

Allison Joy: It's true! In the ancient age of this mega tube's inception, I was an undergrad on the University of Wisconsin campus. Seriously though, that long history of scientists gathering data from the same place with the same tool is something Krys told me they've thought a lot about.

Krys Kibler: So I think that's what's amazing to me. So when I think about, I potentially could drop this 12 meter long tube into the bottom of the lake and never see it again. I think about how some person, 15 years ago was probably thinking the exact same thing.

Allison: I mean yeah, probably! And getting such a long-term view of this ecosystem is a really unique strength of this sample collection. It's led to some interesting results.

Menaka: Definitely. Thanks Allison!

Allison Joy: Of course.

Menaka: And soon, we'll get to what all of these decades of samples have become. First, a quick break.

Dan Udvary: The JGI supported this project via the Community Science Program. This program provides genomics resources for projects with Department of Energy relevance. And we accept proposals from scientists at all career stages. But you don't have to take it from us. Here's Trina:

Trina McMahon: I just love JGI, like, I just think that what they do is so amazing in terms of helping junior faculty or junior scientists get off the ground and, and yeah, just generating these data sets that support so much student training and of course science and, making an example of how to do it well. Like I, I love IMG, so all those things. I just am a huge fan.

Trina McMahon: I should say that one reason I love doing these giant projects at JGI is because, you know, each data set that we generate has like 10 PhD projects embedded in it, right? Like, you can just look at so many different angles, that it really keeps the lab going for, you know, the next 10 years. I joke about how, maybe this will be the data set that I retire on, but that's still a while to go.

Dan Udvary: You can find out more about submitting proposals to the JGI on our website. Head to [joint geno-dot-me slash 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. To recap where we've been so far, we've tagged along with Trina McMahon's students, to see what kinds of sampling have made the 20-year Lake Mendota dataset possible. Each time they go out and collect lakewater — and the rest of their

measurements, they get, essentially, a picture of what's happening in the water at that moment.

And over the years, with sequencing, and assembling, all of those snapshots have come together into movies — about Lake Mendota. First, they started out knitting together smaller strings of snapshots, and eventually worked their way up to that 20 year data series.

This is a major thing the McMahon lab has pushed forward — how to watch a movie like this, and discover things from this kind of metagenomic data.

Because around the world, there are only a few time-series datasets of this size or strength. Lake Mendota is one of them, because Trina McMahon and her students have been out there in boats for decades, collecting samples consistently. And they've been sending those samples to the JGI for almost the same length of time.

Trina McMahon: At each of those data sets, we learned something more about the lake and what we expected to find when we did the next one.

Menaka: To see how that's happened, let's head back in time, to 2010. Trina's group wrapped up a previous dataset around then. They sent the JGI samples from Lake Mendota and another lake in Wisconsin.

Trina McMahon: And at the time we had a hundred samples from each one, give or take.

Menaka: So, not as big as the dataset that the JGI assembled in 2020, but a heavyweight for 2010.

Trina McMahon: At that time it was sort of the biggest data set that, that JGI had done metagenomics on. And we actually kind of broke the sample submission system because back then it was all like, web field, typing in all the sample information, all the metadata for like 200 samples was just insane.

Menaka: And it wasn't just submission systems they were testing – they were also testing the actual metagenome assembly capacity for the JGI.

Trina McMahon: At that time also, the combined assembly that we did on those data sets was going to break the computers. Right. And I just thought this was such an awesome thing, that we would like — break JGI's computers. And so, when we proposed to do the Mendota time series, you know we knew that it was going to be more samples and more data per sample, and it was kind of the same concept of like — let's write a proposal to sequence samples that will break JGI's computers . So that is always in my mind, actually.

Menaka: And testing those limits isn't just arbitrary. It's about stretching the kind of science that's possible. That led to new assembler software that wouldn't break at this scale, including a powerful program called MetaHipMer. Someone's gotta push the limits of the software and

computing around here!

Trina: My life goal...

Menaka: But it's great — that basically, you know, as you're pushing the bounds of what the programs can do, you're also kind of paving the way for more people to collect these big data sets and, and benefit basically from MetaHipMer and all the infrastructure.

Trina McMahon: Yeah, absolutely.

Menaka: Check out our last two episodes if you want to hear more about how software and computing have risen to meet this kind of project.

Menaka: So, after spending roughly a decade pushing the limits of what they could sample, submit, sequence, and assemble, Trina's group gets the idea to pitch a big project. This is in 2017.

Trina McMahon: So we'd collected, you know, 20 years worth of samples and kept them in the minus 80 freezer. And we'd done some analyses along the way, using 16s ribosomal RNA. But when, the sequencing technology got advanced enough, we decided that that was, you know, the perfect time to go back to for the full 20 years and do the metagenomes.

Menaka: So when you think about each sequenced sample as a snapshot of Lake Mendota, that sample archive in the minus-80 freezer is kind of like a giant collection of film negatives. Trina's deciding to basically go back to all of those negatives, redevelop them for even better snapshots, then knit them all together into the biggest movie of Lake Mendota yet.

If that sounds like a big job, it is. Just getting all of the frozen samples wrangled and processed was quite a task. Trina has one former student who did the bulk of this work. Her name is Robin Rohwer, and she's currently a postdoc at UT Austin. She organized all kinds of historic data, took on months of DNA extractions, and sent in the dozens of samples that the JGI sequenced and assembled.

And after all of that — Robin is actually still working on the Lake Mendota Data.

Robin Rohwer: So the Lake Mendota Microbial Observatory has been my baby. I poured a lot of sweat and tears into it. And it's just this incredible data set that I'm obsessed with.

Menaka: And mostly, that's because she really sees the power of a long-term dataset. It's especially useful for an ecosystem like Lake Mendota.

Robin Rohwer: Lake Mendota has really strong seasonal patterns. It freezes in the winter, it

thaws in the summer, it stratifies, so the warm water's on the top and the cold water's on the bottom. And different organisms grow in the lake at different times.

Sample collection and preparation done for this project by Trina McMahon's lab at the University of Wisconsin, Madison.

Menaka: So there are changes constantly happening – and it takes a big zoomed out view to understand how to interpret those shifts.

Robin Rohwer: Oh, yeah. We talk about the invisible present, which was a term coined by Jon Magnuson. And the idea is that without the lens of long-term change, you only see the short term fluctuations and you, and you're stuck in this invisible present where you can't tell if this is part of a longer trend.

Menaka: In other words, with just a few snapshots of this environment, what you're seeing could be seasonal. Or, it could be noise, not signal. But with years of snapshots, or, even, a two-decade movie, you can get more clarity about what's going on.

Robin Rohwer: And so having a long term time series lets you see past the invisible present. 20 years, in terms of climate change, is not that long. But it's still pretty amazing for microbe data. And, we do see long-term changes in the 20 years.

Menaka: Once Robin got sequences and metagenomes back from the JGI, she got to take a look at how that movie of Lake Mendota plays out. As she traced long-term changes in Mendota's microbes, she could look backward, to the start of those changes — and many times, these shifts started with invasive species.

Robin Rohwer: Halfway through the time series, spiny water flea invaded, they're like a zooplankton. They're a big zooplankton that eats little zooplankton and they screw up the food web. They fit in where small fish fit into the food web because they eat zooplankton.

Menaka: Whenever an invasive species comes in, there's a whole cascade of impacts. In this case, there were fewer zooplankton around — and so there were less organisms doing zooplankton's main job: eating algae. As a result, the water got cloudier, and that changed the microbial community, too.

Robin Rohwer: And then we also saw changes when zebra mussels invaded towards the end of

the time series. And so, well this is just an example of being able to directly observe these long-term shifts in a natural system.

Menaka: Within this movie of Lake Mendota, there are also shifts happening that are separate from invasive species. Here's Trina McMahon again.

Trina McMahon: Watching evolution happen, you know, in real time across the 20 years is really amazing because with microbes, they evolve so quickly that their evolution is happening on the same timescale as their ecology. So they're interacting with other organisms and there can be evolutionary processes that play out while they're having those interactions. And, you know, you can't study that very well in plants or animals because they evolve too slowly comparatively.

Menaka: Even after working with this environment for decades, there's still a lot more to learn from Lake Mendota, and this dataset.

Menaka: So next I'm curious, what's next in terms of Lake Mendota?

Trina McMahon: Ah, yes. Well, we're working on new research proposals mostly to the National Science Foundation, to analyze more of the data set that Robin generated. We'd like to use some of the, what we find in that data set to try to culture some of those most interesting bacteria.

Menaka: And because this project is so long-term, each finding is connected to lots and lots of other researchers. Lately, Robin has published some of her results from the 20-year time series.

Robin Rohwer: And I emailed all like 10, 12, 15 people, just the grad students who had led sampling to tell them you know, look what came of all of your work.

Menaka: It's neat – researchers who were sampling off a similar boat to the one that Allison went out on with Krys, trying not to drop the sampling tube — they were all working toward the same goal. Understanding this ecosystem to better inform how we take care of it and handle environments as they continue to change.

Robin Rohwer: And it's kind of a selfless thing because you're collecting samples for somebody else in the future.

Menaka: All of this work — sampling, processing those samples, and then analyzing them, is much bigger than an individual project.

Really, even beyond sampling, a lot goes into this work. There's also the software development, testing, and supercomputing that turn these sample snapshots into larger movies of microbes.

But after all of that, a lot of information comes out, too — beyond neat individual results.

Robin Rohwer: These observational data sets and these time series have so many possibilities in them. So much more than one person's project. So they're a really valuable community resource too.

Menaka: With a dataset like this, other freshwater researchers can dream up their own analysis, or use the wealth of information to contextualize and compare the systems they're working on. Plus, these protocols and programs lend a hand to researchers working on other environments — lately, researchers have used the assembly software that knitted this dataset together, MetaHipMer2, to assemble large datasets from soil.

Overall, this Lake Mendota project is very unique, but it's also a good representation of work at the JGI. The idea is to help lots of researchers build foundational knowledge and approaches — so they can understand new ways of creating the fuels and products we use, while still protecting ecosystems around the world.

Menaka: So again, that was Kryz Kibler and Trina McMahon from the University of Wisconsin, Madison, and Robin Rohwer from the University of Texas, Austin. We'll link to their work in our episode description. You'll also find a transcript of the episode there!

To close out our last episode of the year – thanks for tagging along with us! We've headed to hydrothermal vents deep in the ocean, the greenhouses of University of Nebraska, even the turfgrass species of the 2022 World Cup in Qatar — because we want to go wherever the science is happening. That means helping researchers crack tough environmental questions, and also enabling broadly relevant datasets. 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, Dan Udway and Graham Rutherford.

You heard music in the middle of this episode by Cliff Bueno de Mesquita, who's a multitalented 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

The Megadata of Lake Mendota – Part 1: Many, Many Mers

The Megadata of Lake Mendota – Part 2: Souped Up Computing

Related papers:

Rohwer RR et al. Species invasions shift microbial phenology in a two-decade freshwater time series. *Proc Natl Acad Sci U S A*. 2023 Mar 14;120(11):e2211796120. doi: 10.1073/pnas.2211796120.

Riley R et al. Terabase-Scale Coassembly of a Tropical Soil Microbiome. *Microbiol Spectr*. 2023 Aug 17;11(4):e0020023. doi: 10.1128/spectrum.00200-23.

Our contact info:

Twitter: @JGI