# In Dialogue with Dr. Richard Barker

iGEM CONCORDIA

ASTROYEAST 2020

# Dr. Richard Barker
# UW-Madison Researcher and Co-Investigator for NASA grants
# August 2020

Interviewed by iGEM Concordia Montreal

## *SUMMARY KEYWORDS*

*plants, studies, microgravity, genes, research, yeast, data, gravity, human, researchers, genetic, NASA, blue skies, spaceflight, hypoxia, studies, p-value, threshold, data point, heat maps*

**Bio**

Dr. Richard Barker is a UW-Madison researcher investigating plant responses to spaceflight using custom imaging platforms and cloud-based data analysis. He currently is a Co-Investigator on three NASA grants and has been involved in the planning and launch of multiple AstroBotany experiments to the International Space Station (ISS). As part of the GeneLab project with NASA, he built the TOAST (Test of Arabidopsis Space Transcriptome) database to support the goal of democratizing space life science research. He is currently co-chair of the NASA Advanced Plant Working group.

### PART 1

***[Introductions]***

**Dr. Richard Barker 00:00**

You might have seen some of our 3D Clinostats online. Sometimes I'm a researcher, sometimes I'm trying to get my hardware to people. And ultimately, I'm an environmentalist and a space scientist at my heart. We do have a 3D clinostat as a product. That was actually made from our research. So it was awesome. There's one this project best project based learning, like lesson plans that we were kind of free styling in the college, and just a couple of engineering students who just, we kept sending them electrical and mechanical challenges to allow us to photograph plants growing. Some of them I thought were impossible and they just kept succeeding and doing them. And they said, Well, can you spin it? And then they did.

There's this beautiful video from the students. It was like one o'clock in the morning and they got all to work. You can see the clock in the background and you hear them like declaring their love for another. It was awesome. That's entirely the roots of the projects and how we created this device. It comes from passion for plants. We saw that we needed to get lights in the sensor, because the early pioneering studies or plants in spaceflight were done in bricks, biological research and canisters just grown in darkness and you can learn a lot from that. The reality when your goal like NASA is to feed astronauts, you need light.

**Dr. Richard Barker....**
All the microgravity simulators on Earth when you're spinning the plant to randomize the gravity vector, you're also randomizing the light vector. Anyone that has watched a plant seed germinate knows they use both the light and gravity gradients to orientate their shoots and roots. And in the absence of one, the other takes over. So it's one or the other, right? We needed a system, a clinostat system that could spin and provide a light vector in the center. And that's why we ended up making this kind of funky one which has a lot of cameras and lights in the middle. We realized that kind of extra capacity is over engineered for most people. And we learned this from speaking to Sylvain Costes, the head of the NASA GeneLab program. I'm not sure if you are in collaboration with them, but they're lovely folk.

Essentially, he wanted to do with human stem cells in his radiation chamber, because he's always wanted radiation research. He's the head of GeneLabs on NASA. He's meant to coordinate all the research to try to make information to help inform human life science and space, right. He really wanted to put human stem cells in this radiation chamber in a 3D clinostat. We were quite lucky because we had the final part of the system that he needed. And that was the first time when we actually sold the product. Until then it's always just been a bunch of researchers hacking stuff together. One of the things that for me has meant that has actually inhibited the research side of it. Because every iteration the device was getting better.

We were going through different hardware every time. One of the things I've learned through doing a meta analysis of plants in space, like recently at GeneLab, we looked at 16, the first 16 studies applied in space. And basically what we're seeing is that the hardware is the major determining feature between them, because the hardware that you use determines the organs, the tissues, you get, and other like abiotic factors, right. This is why when we got this functioning design, we realized, we wanted to make it as accessible to as many people as possible.

With the earlier prototypes, the ones that didn't spin, we put the designs on the internet and provided them to people as part of a national biology course, tested at Carthage College. And again, if you're doing that, what we learned was that in the production of them, the biggest variable is the engineer, the person that's doing it, no disrespect to my mate.

But  the temperature of the room makes a difference on the 3D printing of the shapes and so on. And this is why again, honestly, from a research perspective, it makes more sense to have them produce centrally, and then distributed. So everyone has the exact same piece of hardware to make their comparisons comparable. That's why I came to this point when we were looking at people that sell 3D clinostats, they’re all aerospace companies. And they're all over $40,000, which is just impossible. That's why we've

**Dr Richard Barker...**
come down and we say, well, we'll do it for five because, that's where the market is, and we're just a bunch of researchers who have worked out how to do it. But we want most of people's have the same version as ours, so that our results are comparable. And that's kind of the ethos of the cognitive science environment. That's kind of how we came about.

Along with that our long term goal is to then have other non spinning versions, like the 1*g* control that go alongside it, and help people implement citizen science projects for schools nearby. The places they get 3D clinostats kind of have a gentlemens or gentlewoman's agreement with us that they're going to try and do some educational impact with it, you know, because, when kids see these things spin, even big kids see these things. It's so cool. They're like a moth to the flame, and then you have the opportunity to tell them your science and I'm forgetting that again, that's a really important thing for me, just sharing my passion for plants, because that's what manifested all these tools. That's me in a nutshell, that's based on the coming from the 3D clinostat perspective. And again, I could go off on one about transcriptional data, transcriptomics, I'm from a plant perspective, but I do have some yeast datasets. And on that note, I'll stop introducing myself. Now with that context is where I'm coming from. Please introduce yourselves and your context and let's freestyle it from there.

**iGEM Concordia 06:11**
Awesome. Actually, there's one, two people who've joined that haven't introduced themselves. Ben, you want to introduce yourself?

**iGEM Concordia Ben 06:19**
Yeah. Hi. My name is Ben. I'm working on the database for transcriptomics.

**Dr. Richard Barker 06:29**
That's one of my passions. So this classical formula of publish or perish. And the problem is, that means that we got all this really expensive data that's come down from the space station, not just me, everybody, all these different silos. They've done their experiments. But the space station has a limited time. It's not going to be there forever. Who knows what other space stations will come next. But what we would like to do is to share the data as quickly as possible and that's what GeneLab is really all about and make that accessible so people can interpret it from a biological perspective to inform the next generation of studies.

This is why I've basically just been pillaging the GeneLab repository, and then making data visualizations for that. Just fun and dashboards, I'm not getting bogged down in the statistics, because that's a completely different ballgame. The moral of the story

is, I've had access to a lot of these different transcriptional studies, but there's too many for me to actually do anything useful with and I'm always trying to collaborate with people. Because when you process some data, often you don't know if you've done it right, or reinterpreting it right. And really the next thing then comes having someone to share it with who you can trust and have a conversation with and try and work out what is this narrative within the data.

This is the wonderful and terrible thing about transcriptomics. There's so many narratives. And when you look back at all the papers that have been done in spaceflight biology so far, they just scratched the surface. They really do. They put out like the top 10 genes and then little bit about that. There's so much more wealth within these data, particularly the RNAseq. And the microarrays are really a great resource, but again was now transitioning over. So just a quick last question for you. What type of transcriptomics are you interested in and in what context?

**iGEM Concordia 08:18**
We're using, right now we're sticking with microarrays and doing comparisons with space flown or simulated microgravity.

**Dr. Richard Barker 08:26**
Very cool. Are you familiar with the Flo mutant that was flown?

**iGEM Concordia 08:31**
Yeah, we've looked at that one. Yeah.

**Dr. Richard Barker 08:33**
That's the only yeast one I really know. That one, I built a dashboard for using the TOAST infrastructure. Here's the trick. What I did was I linked its genes by orthology to mice and humans and all that kind of stuff. So when we select the genes that change in the yeast, we can then go over and look into the mouse studies and look at the human studies and see what their orthologs are doing. Just going to do that with the GeneLab thing. I know there's all these overlaps that can be really quickly identified, but I just haven't had the time to go in and tell the stories. I've just built the visuals that I've had to move on because of the whole nature of this GeneLab, and the scale of the number of data sets I've got is ridiculous. I'm a plant person too.

When I look at the yeast patterns, I think of some of the genes bubbling up if they have a name, which is similar to the plant ones, like some Heat Shock stuff, I go, Okay, yeah, that one, okay good. The analysis has worked well, Heat Shock proteins going up in space. Okay, there's something that's conserved. And I have to move on because of the amount of mouse data though. What I'm trying to say is don't see me as competition,

see me as a fan that is here to bounce ideas back and forth with because I clearly want to explore this yeast data further. Because it's a really interesting model. We are trying to do this cross species thing trying to work out. I've got bored of looking at plants. I haven't got bored of it, but the boss says I have to try and see what we can take from *Arabidopsis* and see what applies, of course. Because in a couple of years, we're going to have a bunch of crops that have gone up. And then we can do comparative genomics in actual plants. And then my academic ADHD will take me back to my real passion.

I've been in self isolation for so long, it's nice to have a bunch of space nerds to kind of share my passion with. I'm assuming that's your niche based on your website.

**iGEM Concordia 10:30**
Yes, we do love space. Definitely. I just, to get started, I just wanted to let you know that. Are you familiar at all with iGEM?

**Dr. Richard Barker 10:38**
I've been to the website a few times, but do tell me more, imagine I don't, please.

**iGEM Concordia 10:44**
iGEM is the International Genetically Engineered Machine competition. teams from all over the world compete. We represent Concordia University. We designed the two year project where our goal is first to make a sort of bioinformatics tools available to microgravity researchers. So we're building a database, and not only in yeast, but also other model organisms. Right now we have human and bacteria as well. And then using the database we're going to select some genes that we're going to design our reporter strains and then eventually, we're going to make resistant yeast strains. *S. cerevisiae* strains that are resistant to microgravity induced stress. And the goal is to, for space by scientists to be able to use these strains to predictably biomanufacture things like nutrients in space. It  falls under the Bionutrient program at NASA. So we're kind of trying to build that reliable microorganism.

We interview different stakeholders. These could be microgravity researchers, or space scientists, or bioinformatics experts in order to talk to them about our project, see what their concerns are, get their feedback, and we integrate it back into, implement it into our project. And we document this process. And first I just wanted to know if it's okay, if we're recording you, actually started recording from the beginning. But the recording really what we do is we transcribe it, we kind of say on our report to iGEM that we interacted with you, we give a brief description and then we say, this is the feedback that we got, and this is how we integrated it in the project. Is that okay?

**Dr. Richard Barker 12:30**
Most definitely, Yep.

**iGEM Concordia 12:31**
Okay, cool. And we actually have some questions that we prepared. The first one I wanted to ask you is how did you come to be a researcher, a space scientist?

**Dr. Richard Barker 12:43**
Oh, cool. Yeah, just by being mad about plants. I mentioned I'm sorta an environmentalist. At the end of my undergraduate I looked at my modules and everything said plants. So I kind of changed over to plant science. I realized I didn't need to save the planet, we needed to save humanity from themselves, and make sure that we don't bring any new virgin ecosystems into cultivation. The best way to preserve as much biodiversity as possible is to increase the intensification on modern agricultural land, as opposed to bringing in more land. And that's why I did my dissertation, my thesis on the Green Revolution and the genetics of the Green Revolution. The intensification, mechanization, industrialization, modern agriculture post World War II, as a result of the perfect storm like extra nitrogen fertilizer from the scaling up of the Haber process to make bombs, when you don't make the bombs and you have all this nitrogen fertilizer.

If you give that to your normal crops, they get really big and they fall over and your combine harvester doesn't work anymore. So the dwarf varieties identified by Norman Borlaug allowed the system to function, hence the amount of wheat yield has gone exponential along with human population growth. So we haven't had mass starvation, which we would have had if we'd had old fashioned pre-evolution agro ecosystem. There's a lot of negative consequences with that too, but I'm not going to go off on the agro ecology of the monocultures that now prevail across our planet.

That was my research, so just being passionate about plants and I was doing that and trying to find out about the next stage of the Green Revolution, how we're going to continue to intensify modern agriculture, to continue to have the population growing without having the negative consequences of not feeding them. So we're researching the next generation Green Revolution alleles. So new dwarfing plants, so the old dwarfs, they just collectively dwarf. What if you just select which parts of the plant are big and small and do that using genetic engineering techniques. And that's what I did demonstrate, how we can make any organ of a plan bigger or smaller by about 10 to 20%

So after making this framework modification, I came to America thinking well, that's all good and well, but what about the environmental component because the environment is changing. We need to understand that. I couldn't get into the areas I was looking into because salt is terrifying, the salinization of agro ecosystem, a bit like what happens on Mars on a global scale is probably not good for life in general.

**Dr. Richard Barker...**
But as I didn't want to be terrified I thought I'd go and get lost in the clouds and think about something really theoretical, which shouldn't be valuable, which is how the plants perceive gravity and respond to gravity. That's what brought me to America, my fascination with plants and gravity sensitivity. And so we're just turning plants on their side, and then cutting off the roots and then doing RNA seq and proteomics to look at the genes and proteins. The changing and abundance to try and untangle some of the genetic and molecular mechanisms that allow plants to perceive the orientation relative to a gravity vector and then coordinate it, if it changes. I thought this wouldn't be really valuable information for anybody.

We have theoretical blue skies research. You know, that kind of blue skies term doesn't matter why the sky is blue doesn't make you rich, but it's just fascinating physics to think about. And all that time ago when it was discovered, blue skies research wasn't really valuable then, but now we have lasers, thanks to the optics that we learned from understanding that concept.based on where your insertion goes.

And so that was what the gravity research was meant to be just something like blue sky non-applied. And then I realized you can apply it in microgravity environments. So that's just one of these really separate serendipitous moments. My visa was running out. I was about to fly back to the UK as I thought, well, my US adventure was over. It was fun while it lasted. Let's go back to juggling five music festivals. I don't know, this was this opportunity with NASA. I put in the proposal and we found some varieties that were more sensitive to gravity and some that were less sensitive. And they thought therefore yeah, that's that sounds like a good reason to fly them on to the ISS. So there's just this random serendipity in this last moment, like, I never really plan to do anything with NASA and all that. It was just that yeah, chaos is always chaos. Right? So the moral of the story is just be really passionate. That's kind of how you're going to meet that random person. If your passion takes you towards the stars, yeah, it'll happen.

**iGEM Concordia 17:29**
That's awesome. I'm curious. The ones that you mentioned are sensitive, do you see that they have a stress response to microgravity? And does that affect their growth?

**Dr. Richard Barker 17:42**
The prevailing pattern that you'll see in all transcriptomics for life in orbit is hypoxia, basically, space stations but a little bit of extra CO2, a little bit less oxygen, and it's more pronounced for plants and organisms that are like static. Humans move around and where they are, there's fans so that hypoxic suffering isn't so bad. It was worse in the early space flights when they were on the shuttle. But then they went to sleep.

They put fans next to their heads and made it force selective mixing. So the bubble of CO2 was blown away. With plants we can do that now we have fans on the leaves, we still see patterns associated with hypoxia. And that's coming from the roots.

You can put a fan down there. That's the state of the art now for spaceflight research is what's the best way to water a plant without drowning them. So hypoxia is the biggie and I haven't really found that in yeast, but I bet it's the same. I haven't looked at that stuff. So yes, then you look at what is the response to hypoxia, you look at the primary carbon metabolism when I've got a bunch of papers that are going to come out with pictures of that, the TCA cycle, that kind of thing. So what happens when the TCA cycle doesn't flow properly? That's like the next level stuff that kind of molecular stuff around that. Which one day we'll understand. But the moral of the story is... and here's the funniest thing, Mary Muscari. She was a researcher from the 90s. She just recently passed away in this space shuttle era. And all her research showed that plants were experiencing hypoxia, she was pointing at all the things in her studies. There were some opposing pieces of data that were confounding things. But she was totally right. She didn't have the omics data that we have now, she was just like measuring the things that they could see. So it was way more impressive because of the detective work associated with it, essentially. So now we have these like much more advanced technologies, we can look at like thousands of data points, the ones the big stats, and basically all we can really say, Mary was right.

**iGEM Concordia 19:58**
That's awesome. When it comes to yeast, because you said you have a little bit of knowledge and we also looked at the mutant study, can you share with us a little bit what you know, but this?

**Dr. Richard Barker 20:11**
Have you seen my astrobiology website? Do you know where the yeast study is hidden within that?

**iGEM Concordia 20:19**
I remember it was not easy to find. But I did find it. Actually we had already selected it before I spoke to you. Yeah, it was already in our database.

**Dr. Richard Barker 20:28**
Cool cool. Because I'm looking for more yeast studies because, to be honest, that is fascinating because of this Cross Kingdom analysis looking in a bunch of other random studies selected by other random collaborators. The cross kingdom project has not gone as I wanted to test based on too many random events. But what we're basically from looking at all the different organisms and different distresses were essentially found a few transcriptional nodes that we think might be conserved across all the different species.

**Dr. Richard Barker...**
Like conserved responses, this genetic node is responding in mice, humans, and *Arabidopsis* and yeast.

These are the transcription factors that regulate that node. And this is a fundamental conserved response to something about spaceflight. You can't genetically modify humans, and it takes a lot of time in plants, I've got other like, models targets. So one of the things we figured was actually yeast could be a really interesting model organism to take out these genetic nodes that are part of this conserved response to just spaceflight, because that's how we do genetics in *Arabidopsis*. Not sure if you guys are familiar with classical school genetics, but as soon as you just knock out a gene and see if you can measure anything that's different. If you can't measure anything that's different, you then torture the organism in a range of different ways to see if we find a conditional phenotype. It's simple right?

I don't know where you guys come from, I don't know if you guys are familiar with that type of research, I don't mean to patronize, it's for sure. Yeast would be the premier way to do this, in my opinion, because the genetic tools, I think, are already there. I'm not familiar with them because again, it's not *Arabidopsis*, which is my model. Essentially, I think with *Arabidopsis*, we could just order knockouts and do the genotyping, but from what I hear the yeast libraries are better. Then one would look at it and go yeah there's nothing different probably.

The key would then be the phenotyping in the spaceflight related stress and this is why we've been working on developing 3D clinostats that's to be the equivalent as well. As far as I'm aware, most of the microbial microgravity research so far has been done on slow rotating vessels as the terrestrial analog. Correct me if I'm wrong. Again, I'm just trying to remember the metadata from the studies. So we have put a design for that out, again I mentioned this in an astrobiology class, in that we the early prototype to the size spinner. That 3D clinostat was just a 2D rotating vessel. And again we just put a design for the 3D printed base that we used in the Arduino and stepper motor and the code to just have that in a slow rotating vessel mode. Actually, there's a friend of ours, a Japanese teacher in Tokyo, he's doing it now over there. So he's gotten the design off the side of some GitHub somewhere and 3D printed it and now he's got it rotating mizuna plants. He's really confusing those poor poor brassica.

*[Screen shares]*

**Dr. Richard Barker 23:49**
Yeah, so this was the transcriptional viewer that I've kind of created for the GeneLab program. This one is currently called Kingdom and it's not finished. I've got these bloody GeneLab people with the obsession of dissecting mice into tiny little pieces.

**Dr. Richard Barker...**
There was a yeast study. And again, it's quite limited because I'm only really going for the files that have been processed by NASA GeneLab and just like pulling in. So this top part here, this is just an interactive system for what I call iterative filtering. Stats are always the best way forward. Yes, I know. I know. But when you speak to all professors, they have their prejudices. And so they have a particular like I don' t know organelle or place or function; biological process that they're interested in. So to be able to filter for that is kind of useful. And this thing here is just based on ensemble, there's loads of different ways of getting ontology, but I like the ensemble system. It's simple. So this is the study.

Okay, so what did they do? So this is the way GeneLab analyzed it. And so they put out a p-value, standard deviation, and the mean expression value for every probe that they had on that microarray that can be viewed like this. Some people like these sliders and stuff for dashboards. I don't always do it through here. But I was forced to put this on and it took me ages to do and I hate it. It doesn't work. Anyway. But so this is a super oversimplification of this complicated study. If we scroll down, they had three different varieties. So they've got baker's yeast, the wild type they've got FloA and the Flo1 mutant. This overall summary that GeneLab had produced is a gross oversimplification, because it's pulled together all three of these.

It's worked out the standard deviation of the value for these data combined. So it's actually losing the real value of the original experimental design, which is why I had to get back to the second stage behind the summary, then separate out to these three levels. Having said that, this is the GeneLab process data. So from a bioinformatician's perspective, you might want to introduce other ways of processing this, there's different statistical tools. So an alternative way to present this would be through a Jupyter Lab Book, or an Arc Notepad, where you could have the users select their seq or HR.And within a week or two, we should have a version of that for the 15 *Arabidopsis* studies.

In GeneLab, the goal was, in the near future use that code to do the full 90 process studies. So we'd have an example code set like that for this yeast study in the near future for you to look at. The reason I say this is, I'm not going to know it's right, because the way we've set the code up is that it's going to be doing this duplication… We've got it to work with one plant study in the 50. And then 91, including all the different species. So this is why I'm putting all my cards on the table because I need help from specialists in the area to look at their bit and go, 'Yeah, you did a right and this is how our stuff was different. Isn't it amazing how they're similar here or not.' And this is kind of like what I've been doing with GeneLab because as they've been processing through this stuff, they've been overwhelmed by it too. We had a mouse research group originally. I actually found a mistake in that code. They had to redo it, like 10 to 20% of their studies.

**Dr. Richard Barker...**
The moral of the story is that we've all benefited by collaborating with each other. And so, as I show you this, try not to see me as a competitor because yeast isn't my area. So this is all fun. We could select the most significant study genes that responded based on the overall summary statistics. Okay, cool, scroll down. The gene that they identify from the overall thing is based on the Flo1 mutant, which is kind of interesting. I wonder what that means. So ideally, you'd be going off and doing a quick study on this and find out what's so special about it. But again, that also shows how this oversimplification that GeneLab did isn't really the right way. You really want to break it down and look at the correct pairwise comparisons.

I also like to have a hypothesis, I'm just going to look at the genes that went up in the wild type. And then we go down and I look at the two Flo mutants, and I'm gonna look at the ones that also went up in the Flo mutants. Because for me, I use some of these data systems like an intuitive filtering system. And oh, yeah, they went up here too. Okay, cool. So now we'll be down. These are the ones I think are probably going to be interesting. Why did I go up? I like systems I can modify. I love the concept of synthetic biology, Okay, knocking out a gene is one thing. And so you can understand a system if a gene is important by knocking out. But I do love the concept where if we add a gene in, we might be able to increase resistance to stuff. And it's a simple perspective. But if the planet if the yeast is stressed out in spaceflight, maybe it's going to switch on some genes that help it adapt. And if it's doing it in the wild type, and it's doing it in the mutant is something which is conserved, I should understand better why they selected that Flo mutant, but that's not my job, that's your job. You're the yeast people, right. Then we need to give these groups a function. This was what I expected if I made a different dashboard.

Okay. And again, there's infinite ways to show these things. The reason I put it in as like a dashboard is because the profs always come back with like, they always want to filter. That's why I tried to make it so you can go in and filter so they can filter themselves afterwards. I'm trying to teach them to like, fish, but they're not taking it, they still make you do it over and over again. But that's the nature of professors in high academia. This is what I call a random prejudice generator. I mean, everyone loves them. I love them too. It's random, kind of where they put the words. So I always worry that it's going to make you misrepresent or misinterpret data. So wherever possible, I just try to do a bar chart, to be honest.

Okay, pie charts, I love them. There's so many better ways to display data than pie charts. And everyone loves them because they're circular. There's something intrinsically beautiful about it. So it's always good to have one there somewhere. But remember, your main reason for doing it is to create beauty within the data display.

**Dr. Richard Barker...**
There's nothing wrong with that- aesthetics, there's a paper on the importance of aesthetics. If you present data in an ugly way, and this dashboard is ugly, if you present it in an ugly way, people often miss the important pattern that comes out of it. So there's nothing wrong with having a pie chart, well positioned, to create structure to something to help people move around your visualization. Okay, cool. There's the filter. So again, you can find your favorite gene family, that happens a lot, your favorite function, maybe those types of things, but I think I've been doing this embedding resources, like the real value in a yeast library program would be to unlock the resources that's in the Ensembl Expression Atlas. They're great because again, they've already processed the data. So there's tables you can just pull in. Depending on how you like to operate, there's a range of different mechanisms. But I like to be able to say, I didn't process it, it's their fault if the results are wrong. So I go to Expression Atlas and GeneLab for my data. Plus, that way I don't have to worry about the high throughput stuff. I like to keep it on my laptop where possible. So the moral of the story is... and the way I've been using this to catch everyone, to impress the GeneLab people, because they're mice researchers, then I go look at a mouse study or a human study.

Humans in the electromagnet magnetic field there's nothing significant in it. When you look at the paper, they make it look like there's something significant. But then when you actually look at the process data, there's nothing. It's really funny, I say amazing how you can tell a story. We selected a bunch of probes based on yeast and now we're in the mouse studies. These are mice genes that have an ortholog in yeast, and so now we can say that this is a Smart Strand spaceflight, just going to select the most significant ones of that. So now we've gone and done comparative genomics or comparative transcriptomics. We're taking what we've learned from yeast. And now we've gone and looked in the mouse study and found overlap between those two studies. We can start to narrow in on the pathway, how many are we down to now? 18.

So this has been the game, I've kind of stopped because there's just so many possibilities of overlaps. As a geneticist and in plant science, I always want to get down to one gene so I can then go, 'Oh, well, this gene comes on this organ, this tissue, this stage of development, and responds to these environmental stressors. Give it all of these different characteristics. Give it a name, call him, Bob, right? But like, there's so many genes and so many different organisms. It gets hard. This system allows that kind of intuitive filtering, applying over each other. I'm just showing some off of yeast. And now we're down to a bunch of fused human genes in bedrest, okay, bedrest is an analogue that's used by NASA to look at the consequences of muscle wasting in spaceflight.

And so we've done mice in spaceflight, we've done humans in bed- muscle and muscle wasting, and we've done yeast, and this is one gene that we would have narrowed it down to is from all these different things.

**Dr. Richard Barker...**
The moral of the story is I now believe this one gene is actually worth going off on, on a tangent to creating a bioinformatics story with. Finding out how it's expressed in humans, where it comes on, where it goes off. Same thing with mice. And again, with the concept that it's something to do with cells, cellular metabolism, because fish out of yeast originally, that's how we first narrowed it down, right? And it's something that's conserved in cellular changes in all these different organisms in spaceflight. Because my prejudice with my plants studies, I'm going to think is probably something to do with hypoxia. But this is again the problem, we should try not to have too much prejudice from prior studies, or should we I don't know. It's a fun game though, right?

**iGEM Concordia 08:05**
It's very interesting. And one thing I wanted to say is that we do want to show your database. It's in our local host. And one thing we're missing because the idea was like we wanted to do exactly what you're just showing and to compare, but not across a microorganism, specifically across studies, and to find a kind of standard. We talked to quite a few microgravity researchers, and we were noticing there isn't a standard. How do you say, okay, like definitively this gene is affected or this pathway is affected? It was suggested to us to add heat maps, but then you have to make sure the studies are comparable. And a lot of times they're not. So this is where it gets a little bit tricky, and I thought we could just show you our database and what it looks like. It's still just a working example. We don't have visualizations yet. But we're working on the heat map. Can you share your screen Maher?

So basically, we were able to gather. Like for MVP we aimed for complete yeast data gathering. We found two main studies, one included, two different generations of studies 5 generations versus 25 generations and the last one was actually a different type of study. So one was like, use the HARV, the other one was space flown. Eventually what we're aiming to do is to have a cross platform search. We initially worked on eight different search possible criteria for users to search for their specific organism, species strain or even specific gene, or platform operating frame. And also fold or regulation, so called change up or down or unaffected, which is something we're still debating on.From the available studies, it's a search for a single gene. So this is the first study which had the 25 and the 5th generations. And this is the second one. They both were taken from NCBI data sets, the samples. So, yeah, so one of the things, as Hajar mentioned, was a low p-value with which log fold change would actually inform us that this expression difference is significant enough and can be taken into consideration for scientists.

**Dr. Richard Barker 11:36**
Yeah. This takes us back to the earlier question, which was, what's the right threshold?

**Dr. Richard Barker...**
It is entirely set by the individual study. This is the terrible thing...there's a wonderful research group Space Plants Florida, and they do everything first. They're like the pioneers like they go to Congress and do everything they want. If you look through their history of spaceflight studies, and for the earliest ones, latest ones, you can see in every single study, they use a different threshold to discuss their favorite top responding genetic pathways. It's not until you do the meta analysis and look at all the different studies, you realize every single one has a different threshold.

And there's a reason for that. I could just say, you know, it doesn't matter if it's naught point naught five, that's the only threshold. My way of getting around it, because I wanted to find overlap. And I'm not a statistician, so I have low statistical model standards. And so I use p-value instead of adjusted p-value instead of adjusted p-value, as my personal mechanism. So I keep naught five, because I'm old fashioned naught point naught five cuts off, right. But there's all these different statistical tools to create different curves, and all the different omics based technologies do that. Be in the microarray or the RNAseq, also called single cell sequencing, they will have different statistical tools to kind of create the curve to kind of normalize.

Because everyones knows the danger of just doing that single t-test. That's kind of why, in order to find a larger overlap, and to then use biological detection to then actually see if it's if there's anything significant in there, I used a naught point naught five cutoff value as a p-value 0.05 as my personal threshold in my TOAST, manuscripts and TOAST database. And that's just because it gave me a reasonably pretty good overlap. I had a few hundred genes and networks that came out of it made biological sense. I  didn't have to get too bogged down into which statistical normalization strategy I was using to create the just the adjusted p-value. And that's exactly what I'm about to run into with the new GeneLab analysis working group. We have this manuscript coming. And everyone wants to use adjusted p-values for that.

The goal of this database, is it to provide insight about the spaceflight or about yeast in general?

**iGEM Concordia 14:18**
It was supposed to be about spaceflight, like basically, microgravity effects on yeast as an MVP.
And of course, we're expanding to different species and organisms.

**Dr. Richard Barker**
Cause this is brilliant, and I have a request. Oh, please, please welcome this. Okay.
So you've said that you pulled in, all of the yeast data and spaceflight stuff to this.
There's not a lot of space flight data in yeast so far, but there's more coming. But the
PCA plots... can I show you something. Yeah, I'll show you the plant version that

**Dr. Richard Barker...**
we did. And when this panned out, we realized we needed to add more kinds of studies to the two seconds and you have got the perfect platform there for this in yeast. I'm trying to find connections between studies, hardware as the major study, trying to get down to one gene. But what I really wanted was to show you the PCA thing. Okay, I know it's busy but bear with me okay. So these are all the first plant studies in spaceflight. The first 16 microarray and RNA seq. So these are the different biological reps, little dots and such.

This study I helped with BRIC-19; BRIC-16 from the space shuttle and BRIC-20. This one up a bit later, the MCS, a modular cultivation system with a centrifuge was spinning. This is so cool, the veggie. This is the next generation. This is the future RNA seq and lights and vegetable growth capacity. But here this is interesting. So hyperbaric chamber, hypoxia, this will be one terrestrial study that was kind of added at random. Similarly, here root tip, this is rooted with microarray and RNA seq. So these two studies here were done by this group, they did this study and this study. So there's a bit of a batch effect based on the group.

But what am I trying to say, you could do this, you could have all of your yeast studies, you've pulled in the arrays from GEO and such, you have these transcriptional arrays there. You can put it through over a PCA or an MDS plot. And you'll have your spaceflight study. And you have all your other studies from all the yeast studies ever. And they will be scattered based on the data distribution of those studies, right. So any of those threshold studies that are similar to spaceflight might cluster together. And so there might be some study on some random mutant on yeast, on Earth that has to happen to have the same transcriptome as that we see in spaceflight. Yeast is one of the few organisms that has enough mutants and enough to do that. Does that kind of make sense?

The real value, in my opinion, I mean, of all you've already done is the fact that you've got that system to pull in all of the yeast arrays. Then you just create that array, I'm kind of bound to do just the GeneLab stuff. So this is just the 16 GeneLab studies. And I've been saying to them that's crazy. We should be doing it with the ensemble for all plants. And a few years now, get rid of their shackles and work on how to do it, but you guys have already built that infrastructure for the yeast stuff. You can tell I'm excited, right.

**iGEM Concordia**
And I'm really excited to be speaking with you honestly.

**Dr. Richard Barker 17:46**
Like, yeah, I mean, it's fascinating stuff. I wanted to show you this to basically say, like, a real quick pivot, rather than trying to fish in those individual genes, which is what professors want to do. So you build tools to keep your main type of audience happy, don't get me wrong. The thing that I would the question that I would have thought, that I think you could answer with it is, could you put all of those matrices from all the different threshold studies and the space ones through like a PCA software and or MDS software, and in that having assigned to them in a particular piece of particular factors.

So what do I mean by factors, and this is what I had to do with the GeneLab system, like, this was a graph they wanted to make. So these are a bunch of genes or whatever, this is a bunch of studies. And these are the factors that went into the initial studies to make these transcriptional. So for them, it was ER, histological study, etc, etc. Okay, this makes a lot more sense for me if I do it with plants, so I'm gonna do a plant. Okay, so this was the dendrogram that I made with those 16 plants studies that we pulled in from GeneLab. They tried to process them as best they could to the same pipeline. But there is this massive structural difference. You couldn't separate out the RNA seq from it, you couldn’t mix their data structure. When we look at the PCAs, we can see the biggest component was microarray versus RNA seq. So that PCA plot that I was showing you earlier.

This is actually PCA two and three. Because the PCA one was basically just saying, it's all microarray.So this is PCA one, and here's the microarray. Here's all the RNA seq. And the thing I did put on this graph was this is research group number one, this is research group number two. But you can't point that kind of stuff out. Because they're the groups that reviewed your grants and you don't tell NASA that they have different research groups doing their studies, everyone's going to get different results, just different methods, because they're different research groups. There's business, there's a need to be stewards here as well. You catch my drift. So you're gonna have this artifact of RNA seq versus microarray. You can’t not mine the microarray, because there's so much of it, that's been done. But the future is going to be here. So let's not worry about the future for now, let's make the most of what we've got.

That's why I've made these factors. So all of your yeast studies, there's going to be metadata factors associated with them. I don't know how well they were described in your yeast database that you're pulling from. When I was pulling from NASA GeneLab. Now, I had to curate it myself, I had to fix it. That's when the real patterns really came out, what I mean by how to fix it. Okay, so you see here, organ versus tissue. So there were some researchers that wrote terms about organs in the tissue column. And there are some that wrote in terms about the tissue column in the organ. When I pulled the data from them, they were mixed up. This is why I can't do this for yeast, because I don't know the yeast studies.

**Dr. Richard Barker...**
This is why you need these biologists to go in and basically create the same kind of matrix. Because if you do that, you can do your PCA. You can find based on the distribution of your different arrays, but it's up to the spaceflight array. You can find which of the threshold studies are most similar to the spaceflight studies. And that will just be mind blowing. Like there'll be something really cool in there. And it says it's hypoxia doesn't matter.

**iGEM Concordia 21:28**
Are these just expression values that you've put in. Or are they something else?

**Dr. Richard Barker**
Yeah, I'll show you how I cheated, so I'm a plant scientist. I've got a coding student who's now going through to kind of extract the code from our Shiny app that I adore. And so essentially, you can just load into it. Like if you've got an RNA seq counts table (sp?), if you've got microarray, you can do a normalized array. Or if you've got microarray and RNA seq at the same time, if you can get your RNA seq into a process format like fpkm, rpkm, CPM, depending on your organization system, you can just plug that in. And so I was able to pull down the Gene Lab process data and then plug it into here and then use this particular tool to provide some of those insights, because it has a really nice system for doing the PCA. And the MDS plots...

The results from this just kind of pop out into this. So this was like the pipeline. ID conversion goes in and produces those big clustering heat maps that don't really show you anything, but everyone loves them. But the thing I liked about this one was the PCA, MDS, and T SNARES. So these four methods for showing the distribution it shows you how similar the different reps are, how similar different studies are. So you want to split up your yeast studies into the different treatments, into the different reps, because it might be the control of your yeast, which is most similar to spaceflight. Or maybe one of these studies, they stress it by giving it high light, and maybe the highlight environment is actually most similar to spaceflight.

And that's what we learn in plants. Turns out you could activate about half of the space flight plant responses by hitting them with really high light. We're trying to work out why- I could go off and on about cellular stuff- but always comes down to ROS (reactive oxygen species). That's the other positive control. You saw the Heat Shock Protein stuff. They're just responding to increased amounts of ROS in the cell. If you have too much ROS in your cytosol, PCD occurs, program cell death in plants, for sure. And of organisms, probably. In mammals and such as this BGL1 system, it's also known as bit one is another pathway that links into it. These genes are basically switches that determine when a cell gives up. In plants, we found that orthologs called BAG6 (?) came up in our like BRIC-19 study.

**Dr. Richard Barker...**
Because in our study, we were really pushing the plants to the edge because we were growing them in darkness, low oxygen environments. We found this stress response genus was really important for them surviving in that situation. Genetics, we want to get down to that one gene. Based on what I saw from you guys earlier, you've got a really good opportunity here, and that is to basically find out collectively, which of the yeast studies on Earth are most similar to those in spaceflight? Have you seen them do this with the microbiome as well? There's a new microbiome study that's come out in which they took all the different microbiomes from the different spaceflight visits, and then they went on to do the threshold microbiomes. And they did hospitals and NASA clean rooms where they build Curiosity rover and a few other places. And then again, they just treated the microbes like diversity as the factor that they did the MDS on the PCA on.

Then they were able to show that this base station is actually really quite similar to a hospital because of how it's cleaned and because of the throughput of people and the different factors. No one's done that with yeast, and yeast transcriptomics. If you do that, please let me see the results. I'd love to collaborate on it, it'd be fascinating.

*[End Part 1]*

## PART 2

**iGEM Concordia**
We wanted to start by showing you the work we've been doing so far. We will discuss some of the suggestions that you talked to us about last time. Go for it, Ben.

I was looking at the yeast transcriptomic studies, sort of putting them together. I have two methods that I found to combine the differential expression, log FC values. Here I started off, I used a package called Meta Volcano R, and I combined them using Fischer's method, and it gave us this nice volcano plot. Then I did random effects modeling. And they gave us this thing, which had a lot of variability in it. What I decided to do next was to take the top 250 from Fischer's algorithm and run it through a heat map to show the log FC values that were different. The data sets that we're using right now is the random positioning one that has Gen25, and Gen5 samples. Then there's the Flo mutation. One. I just took the control from it and did differential expression... it's really different from the other ones. Then I decided to do some k-means clustering. And it gave me this, it looks like two clusters. Very interesting. And then I went on with principal component analysis. The idea is to sort of turn this into a Shiny app. I don't know if that's what you were using before, or is it just a markdown, Richard?

**Dr. Richard Barker 02:47**
I tend to use a Shiny app, but then I have to convert the Shiny app into a markdown to make it readable like this. How to put it. So you can compile this into a Shiny so you can load any other yeast study into it. Yeah, that would be the next goal. What I think you've shown here, this current figure is that the HARV is quite different from the random position machine. I think that's going to be a clear pattern. This could be coming out of the data over the next few years. Right now, NASA has just invested in a lot of terrestrial research on microgravity analogs, and there's going to be a lot of people looking at this type of thing. I'm worried that a few years from now we're going to end up concluding, that the terrestrial analogs are just not the same as flight. And this figure is, I think maybe, we will see whether it's a premonition or not. But that aside, what I would recommend is then trying to find a load of other studies alongside it.

**.iGEM Concordia 03:50**
This might be it for microarrays.

**Dr. Richard Barker 03:54**
There's no other microarrays on yeast. Like it's nothing to have spaceflight. Just in a hypoxic environment...

**iGEM Concordia**
Yeah, I can do that for sure, like different types of stressors.

**Dr. Richard Barker 04:07**
Cuz that would be the next tangent, you can imagine the Heatbox getting wider and more plots being on the PCAs, and eventually clusters starting to form. So maybe life inside a hypoxic chamber is more similar to a space station than it is to life in an RPM machine. That would be my next recommendation, looks like you've done a tremendous job here. And that could just be one way of trying to add value to trying to focus in and see Is there another threshold study, which is more similar to spaceflight than the RPM.

**GEM Concordia 04:47**
Because right now, like the PCA is pretty simple. I think it's just like the K means or just clustering on these first two here, right? That's what I'm assuming. Going forward, how do you normally combine? Do you usually use log FC values and combine them together? Or is it something else?

**Dr. Richard Barker 05:21**
Um, it depends on the types of data that you're pulling together. These are all microarray, if I'm not mistaken? Yeah, so after the normalization, you can bring them all together, at any stage. I mean, you want to bring the samples together before the normalization stage, because the most likely stage of your batch effect is when you come from different studies. So for example, the different GSE like numbers in the database in the crowd, that's the most likely to be the greatest batch effect is which of those studies you pulled together, so you want to pull them together before the normalization. Sometimes those have already been pre-normalized. If they've all been pre normal as the same system it usually works. But you can do it as it was one. And that tends to work out better. But I think that's what you've done here, isn't it?

**iGEM Concordia 06:11**
No, I've pulled the data from GEO and I did the D-analysis right away, took the right groups out of it. And then I put them all together in a list. From what I've read, there's problems with using raw data and putting them all together like that. With microarrays, at least, if the same treatment has been put on the same platforms, then it's difficult to normalize them together.

**Dr. Richard Barker 06:42**
Exactly that yet. And that's one of the things we run into in the GeneLab system. Essentially, everyone tends to just focus on alpha metric arrays. There are a bunch of like older array types that are again just so hard to really normalize across the different platforms.

**iGEM Concordia 07:00**
Alright, so I thought I could avoid it because this package seemed to believe that you can, you can use different studies together if you just use differential expression values.

**Dr. Richard Barker 07:09**
That's kind of how I did in TOAST, just using again, the fold change, which genes go up and down. And like knowing too much about the p-value. If you get enough studies that agree in the same way, it eventually comes convincing,

**iGEM Concordia 07:24**
Right? Because this is the problem with the HARV one is it's also not very good. The reason we're getting this is because the p-value, the adjusted p-values, at least are all close to one.

**Dr. Richard Barker 07:37**
It's almost like yeast doesn't really mind that much being in spaceflight. So this is a good point. But what is the what do you determine from that?

**Dr. Richard Barker...**
You could determine the hardware wasn't right, experimental design wasn't right. And you spend a lot of time looking at the hardware, look the characteristics of it, and looking at hypotheses that are being tested. We could also simply conclude that yeast can fly up into space. The factors up there would have extra background radiation doesn't really change that many genes seems very cool. There's nothing wrong with having a conclusion where like, well, there's not that much changes basically, the question is, how do you go, well, what is being changed, is it the random positioning machine? Like if the purpose of that is to remove like the gravity component? Does it actually do that?

My question is what type of random positioning machine too. One of the things I've seen is based on hardware, like in space-like experiments, you can't really compare this piece of hardware with this piece of hardware. They both have their own experimental designs. When I had some random positioning machine, sometimes they go into a clinostat mode where they turn smoothly, and sometimes they go in random positioning mode, where they have these moments of sheer, because it's completely random, it ultimately averages out over zero over a long period of time. That doesn't mean that there's periods when it has a hyper gravity, where it's swung. We've seen the videos of them swim, spinning.

So you could argue the randomization machines cause shear. This then takes us to the other type of microgravity simulator, these slow rotating vessels that move slowly, they don't have that shear, they just don't really let stuff sediment at the bottom, which is meant to be like the analog. So I'm feeling we are looking for other studies that use a different analog for a microgravity for the yeast.

**iGEM Concordia. 10:02**
Yeah, I mean, I'll start adding in different stressors. I've done that before. But I'll pull more studies and see what that does.

**Dr. Richard Barker 10:09**
Heat Shock. It's a classic. In fact, in the Toast next to the flocculent experiment at the bottom, there is R Shiny out that has a Heat Shock Protein. Heat Shock mutant expression profile, and it was meant to be a pineapple to hint us on, because Heat Shock proteins come up and all the other stuff. And so that's so the model. There is some sort of Heat Shock related transcriptome in yeast. And I thought it'd be really interesting just overlap and see if there's similarity there. The theory is that in microgravity that your mitochondria doesn't quite function, right. You've got lower oxygen and a bit extra $CO_2$, you might come in as working harder in it and it gives off heat. It might just be this internal kind of warming up that occurs. We have sensors in the room, we don’t see any other cause of Heat Shock.

But Heat Shock proteins come up with all the different organisms. Maybe they just really badly named proteins. Who knows. There's loads of things to compare there but Heat Shock could be a good starting point, that and the ROS, hydrogen peroxide. Again it is a very broad term but it comes up in all of the different things. There's bound to be a yeast study where they add hydrogen peroxide or loss of some sort.

**iGEM Concordia 11:46**
What I found interesting about Ben's findings is that we spoke to a researcher. He's a PhD student in the Astromedicine lab at the University of Nottingham, and he did a comprehensive review of microgravity studies in bacteria. And he's publishing a review on that. He found the same thing that there's a difference between spaceflight and simulated microgravity.

**Dr. Richard Barker 12:25**
Yeah, it's definitely a recurring theme. But that's why people were like, trying to create alternative versions that are a bit better. But yeah, to be honest, we're gonna end up realizing just how different the reality is up there relative to any of these things. But five more years of research, we'll know for sure. Then hopefully the launch capacity will be so great that we won't worry anymore. We will just be doing all the experiments up there, that will be the ideal situation.

**iGEM Concordia 14:18**
We wanted to look into RNA seq analysis.  We want to ask for server access, if we could find that somewhere.

**Dr. Richard Barker 14:35**
Using the cloud pieces to save it. Yeah, I think there's a couple options. I personally recommend CyVerse. Have you heard of it? No. Fantastic. Okay. So CyVerse is brilliant. It's the biggest best kept secret ever. And it's free, limitless cloud storage for any student or academic anywhere in the world funded by the National Science Foundation. It used to be called IE plant, which was when I found out about it back when I was learning how to do RNA seq. And they basically had a bunch of petitions, who built some pipelines to help me do it. It was awesome. And then they had to rebrand local CyVerse, now you have a limitless cloud storage that continues to build things onto it.

There's a galaxy interface for now and these other things, but essentially, it's just like a big Dropbox on them. So they have this thing called the discovery environment. That's how you see your cloud store. There's a thing called iduck (?). That is just another software you can use for moving like sinking files from the computer to the cloud. So you can always sync it up there. These are the sorts of things, you usually have to pay loads of money for it like Google.

For research, it's usually I have access to it. So you're trying to provide the infrastructure. But the great thing about the infrastructure there is it's built on top of supercomputers. So you can pass through any of your commands onto there like a stampede cluster. They're in Texas or wherever they have the fancy name for the one in Arizona.

**iGEM Concordia**
I can run R remotely with this platform.

**Dr. Richard Barker 16:09**
Yeah, you can launch R, R Shiny, R Studio, Jupyter lab works off it. So essentially you can have your high throughput sequencing data, if you're doing RNA seq you have these massive fast queue files, you do not want to waste your time with them on your computer. Don't get me wrong, you just come back the next day. But for me, I really enjoyed having the capacity just to set it up on their machine and have it run on the cloud and continue playing computer games or whatever. Then have it send me an email when it's done. A bunch of bioinformaticians have built a bunch of pipelines in there and you can go join and you can launch like notebooks.

They were the newer thing when I first got this before, you had the capacity to launch your R Studio or your like Jupyter lab book. That's there now under these VICE apps That's the other keyword. VICE app is one of the apps that's inside their discovery environment that allows you to launch R studio. And that just leaves really big lifting high throughput, like fast few steps. It has all the other stuff you have to do in the cloud at the high throughput. You can build that with a bunch of different things like your own Jupyter lab work merged with an R. But ultimately, the counts file is the best moment because it's small if something shrunk it down again. And then there's infinite statistics to come out of that. Whereas you use that counts and combinations counts files from other studies. As you go through RNA seq analysis, there's always saved stuff, spots and every saved stuff spot, you get another file of a different name and an infinite number of branches on other people's software, it's the academic formula unless you buy those out of the box software like CLC genomics, where you just trust the Craig Venter Institute got a right and just push it through their thing, which is fine. Nothing wrong with that. But it is fun to go into the full and it's informative, as you go to each of the stages from checking the quality, the QC controls, there's like three or four different software's that do that and they all say the same thing.

Then there's our quality like trimming stages. There's three or four options in there. So it's good to just run them all on someone else's compute time, and have to pay for it and then just see and compare and contrast. And see they all do the same thing. There's not much variation there. But like, after counts, that sort of variation comes in. Once you've got

your RNA experiment, your first few files, you filter them for good quality. You've then got them floating on a cloud space. You either want to do a de novo transcriptome assembly, so software like Trinity tries to work out what's a piece of messenger RNA and it does it all fresh. Nice if you've got a new novel genome, but if you've got like a standardized genome like *Arabidopsis*, or like *Saccharomyces cerevisiae*, then essentially you probably want to just align to the normal genome, but the model that they use at that time, it's quicker, it's more accurate, and it's so much easier to interpret. That's the most important bit- making life easy with low, low hanging fruit is good.

So you've got all your RNA fragments, you filter them, you're aligning them to your genome. There's the FASTA file that's like the standard genome format, but you also have your GTX and GFF files. So these are like the code names for your different loci and where they're located on the FASTA files. They basically guide your RNA fragment alignment, make it quicker, more accurate and more easy to analyze. You can count how many times you have a known fragment next to a known gene or transcript model. That's why that's a really fantastic safe point, which isn't as appreciated by as many people as it should be the counts, it's shared, we've moved things around and from counts, you can move out of your high throughput, and you can get back on your local laptop. I haven't had a chance to share my personal experience. And hopefully, that is helpful.

**iGEM Concordia 20:27**
Yeah, so it runs basically everything on the cloud, that I need to make a count file basically, as a pipeline for all of it. Okay, that's really helpful.

**Dr. Richard Barker 20:36**
Yep. And, at this point, I should also mention, I think, GeneLab now is also building the same system. I tried to persuade them just to work together and use the same thing, and theirs is entirely run through Galaxy. The Galaxy system has a few pipelines built into it, but it's very limited. If you're interested in the Galaxy. Essentially Galaxy then sits on it on a cloud computing infrastructure. And if you ever want to use Galaxy or just check it out, you should fly five minutes it’s kind of fun galaxy.eu.

**iGEM Concordia 22:39**
I also had another question. Our supervisors are really keen on us, including pathway information, but he wants us to visualize it, right? We're kind of tight on time, because our iGEM is going to freeze GitHub in October. And we're just wondering what do you think is the easiest way for us to incorporate that in our software?

**Dr. Richard Barker 23:02**
So Reactome, Kegg, AraSci- which is Arabidopsis Sci, but there are others. I'm pretty sure there's a yeast sci. And there must be a yeast reactor I'm on a search for now...This is basically how to connect the transcriptomics to biochemical pathways. Right? I know that reactome does have API access. Essentially, if you're doing it the old fashioned way, there's a page in there where you get your list of differentially expressed genes, and you stick it into Reactome and it tells you which pathways they're involved in. And if any of those pathways are bridged, and a lot of times you'll see some genes you're interested in, but the p-value just isn't enough. And other people start to cry at this point but don't. Because like metabolic pathways, there's certain bottlenecks in pathways, and that's a really important thing to remember. I overspecialized in this one, a hormone ...

*[poor audio for 10 sec]*

And these two key regulatory steps. If you switch off one of those locations, it doesn't tweak as being significant in the whole metabolism because there's only one gene out of like 50. So when you look at the metabolism genes, there's a bit of detective work that needs to go in. The code where you'll do your split down to those that are just significant. And that makes perfect sense. But that might mean nothing comes out.

That's what we would have looked into what were the individuals loci they came out, which pathways are they in? For me, I really liked AraSci and here, yeast, BioSci is the equivalent. I found when I got down to the individual part by chemical pathways they had a really nice simple clear wave of allowing you to plot a fold change onto the pathway. So that one is really good. Reactome one, I found its plotting wasn't quite as good, in my opinion, is a little bit more artistic, which I kind of liked, but it wasn't good for scientific perspective. So I think putting the two together is really good. Particularly if you get to the point where one of your pathways is not totally significantly differentially altered, but you can see a few key players that there might be that bottleneck in the pathway. That's when I tend to pull out the reactome picture and go look at this. Wouldn't this be cool if this was a real practice. But that's just artistic license storytelling. And sometimes you can really believe it. To me, that's the most important thing. There's times when I've done that because I believe this bottleneck in the pathway is a root cause or known as a real thing. I just wanted to bring that story out, like it depends on what pops out of the stats.

*[End Part 2]*

It was pretty much understanding how we could close off ecological systems for extreme environments and how people can use them to survive in that extreme environment. That was really the connection where I was like, that's where my two passions interlock together. And so that's where everything began. I went through some independent studies because my school didn't really do a lot of space exploration related stuff. A lot of this was me reaching out to professors who were soil scientists or systems ecologists and asking them if they would like to put a twist on what they were doing, and put it in the space environment and think about how these things would change in the space environment.

I started my research with doing these independent studies, majoring in environmental science and biology with a chemistry minor. And out of this research that I was doing, I developed a new system for what I now call a quasi-closed agroecological system, which applies the agroecology, principles and landscape theory and things that have been missing from the conversation that has been primarily dominated by engineers. So that happened and my company Deep Space Ecology came out of my research because when I went to a space conference, nobody was really talking about space agriculture or food security. It was briefly mentioned, but kind of like, we'll handle that when we get there. And in my mind, this is much more complicated than you're making it out to be. I saw this niche that wasn't being filled in the industry, where nobody at that time was really talking about security and space agriculture at these conferences. Deep Space Ecology came out of that and was founded in May of 2016.

We've been working on that ever since I co-founded it with my father, because he has a background in space research, as well as his Master's thesis in physics dealing with high energy particles in space. He was kind of living vicariously through me as well because he's always wanted to do stuff with space as well. It's been a fun time working with him on Deep Space Ecology and moving everything forward. Now I'm a PhD student at Cornell University in soil and crop sciences and continuing my side of my research, but also the Art side of my research as well.

**iGEM Concordia 6:04**
That's amazing. I connect to what you've said, people will think that this isn't a problem for now. But when you hear things like SpaceX's first cruise flight is scheduled for 2024. I mean, it is in, what, four years?

**Morgan Irons 6:22**
Oh, goodness, yeah. When we're talking about long duration spaceflight and long duration on planet, habitation, and everything that could go wrong with that, and how the only way we'll be adaptable is by creating an environment where we have the