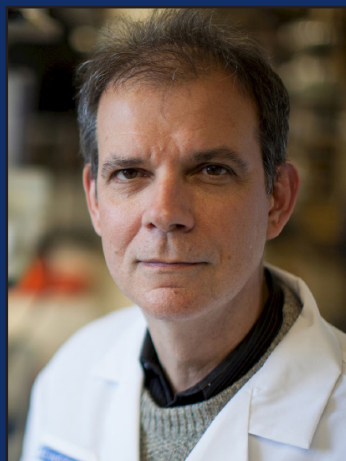


INSIGHTS FROM THE JGI USER MEETING: USING GENOMICS TO TACKLE ENVIRONMENTAL PROBLEMS

BY EMILY HARARI, MATTHEW COLBERT, AND NIKHIL CHARI



Soil microbiology, plant-pathogen interactions, computational biology, and big data all converge in the rising field known as genomics, which involves DNA sequencing and analysis to determine the interactions between organisms integral to biological processes.¹ The Department of Energy Joint Genomics Institute (JGI) aims to solve problems in renewable energy, ecosystem nutrient cycling, and decontamination using genomics techniques.¹ BSJ covered the 2019 JGI User Meeting where we had the opportunity to speak with four distinguished scientists working with genomics to tackle issues as wide-ranging as the fate of carbon in soil systems to the discovery of higher-order combinatorial interactions in cells. BSJ spoke with soil microbiologist Mary Firestone and plant pathologist Mary Wildermuth from UC Berkeley, immunologist Arturo Casadevall from Johns Hopkins, and computational biologist Dan Jacobson from Oak Ridge National Laboratory. Our conversations highlighted some of the widely varied genomics research taking place today.

From left to right: Dr. Mary Firestone, Dr. Mary Wildermuth, Dr. Arturo Casadevall, and Dr. Dan Jacobson.



Figure 1: Grand Prismatic Spring, Yellowstone National Park.⁷ Microbial mats give the spring its vivid colors.

Professor **Mary Firestone** and Professor Wildermuth both spoke about their research in plant and soil microbiology on April 4th. Firestone, a renowned soil microbiologist and biogeochemist, was the day's keynote speaker, and discussed her research in using various genomics techniques to probe the fate of carbon (C) in soil systems. She spent much of her career investigating microbial relationships to nutrient cycling, especially nitrogen (N), in soils, and is more recently diving into C dynamics.

In our interview, Firestone explained that “soil microbes, in a lot of ways, are Mother Earth,” and that they are as “fundamental to life on Earth as anything.” Microbes in soil have long been known to be critical players in the decomposition of soil organic matter (SOM) to carbon dioxide, CO₂. But more recently, they have been recognized for their importance in the stabilization of C in soil, from C sorption to mineral surfaces to the formation of aggregates—clusters of soil particles that bind together and can sequester C within them.

Firestone explained why she chooses to trace each type of DNA found in these soil clusters. There has been little research in the soil viral genome because soil is a solid matrix—viruses are difficult to filter out and can easily sorb to minerals or organic materials. Consequently, very little is known about the importance of viruses to soil ecosystems and nutrient and C cycling. Although Firestone classi-

fies soil viruses as a “big unknown,” she considers fauna to be a “big known.” Much research has been done to determine the roles of soil fauna (small Metazoa like nematodes and mites) in making essential nutrients like N and phosphorus (P) bioavailable to plants, but, according to Firestone, the relationship between faunal food webs and soil C remains under-investigated and “begging for improved molecular tools.”

Among these tools is a genomics technique called primer quantitative PCR, or qPCR. Faunal quantification has typically been performed by visual identification and direct counting in soils. However, Firestone and colleague Javier Ceja-Navarro at Lawrence Berkeley National Lab found that primer qPCR yields largely similar quantitative results.² Firestone explained how qPCR can make quantifying fauna in soil much easier: “You simply have to take a large sample of soil to encompass faunal diversity, extract the DNA, and use specific primers to do PCR amplification of the barcode that you want to quantify. That gave us very, very similar results to the direct counting method, and confirmation that qPCR works quantitatively.”

Firestone's research also tackles the impact of arbuscular mycorrhizal fungi (AMF) on soil ecosystems. AMF are unlike most fungi in that they are not saprotrophs. Where saprotrophs use dead C as an energy source, AMF are biotrophic, or as Firestone said, “they eat

Figure 2: One of Dr. Casadevall's watercolor paintings.
Reprinted with permission.



living C by plugging directly into plant roots.” The exchange is mutual—AMF also pick up essential nutrients like N and P from the surrounding soil and transport them back to plants. Firestone’s recent genomics research focuses on what she calls “helper bacteria,” which associate with AMF. Though Firestone stressed that “this is a story that’s still unraveling,” she believes helper bacteria could work to decompose organic N into ammonia—a bioavailable form that AMF can help transport into plants. By ^{13}C -labeling AMF hyphae and tracing which bacteria around the hyphae are picking up ^{13}C , Firestone and her group are working to “establish a direct link between hyphae supplying C and bacteria living around them that are consuming it.”

Despite her strides in soil metagenomics, Firestone stressed that there is still much work to be done. “Soil is the most diverse microbial habitat on Earth,” she explained, but a “very, very small percentage of the soil genome—less than 1%—has been sequenced.” This is largely because it’s difficult to extract DNA from a solid yet dynamic matrix like soil. But Firestone still believes there are “many, many lives of interesting research for young students in this area of research.”

Professor **Mary Wildermuth** also studies plant-microbe interactions in soil systems, but her path to the field could not have been more different. While she conducted undergraduate research and graduated with a degree in chemical engineering from Cornell, she wasn’t sure she wanted to stick with the field. “I realized as a chemical engineer that I was missing biology,” Wildermuth recounted. “From there, biotech was pretty new, and I realized I didn’t want to go into traditional chemical engineering.” She obtained a position at biotech company Gibco-BRL and helped create the first Hepatitis B DNA

detection probe (for research use) in concert with doctors at the NIH. While contemplating her future options, Wildermuth elected to travel abroad. “I had always wanted to go to Africa to get more of a perspective on what I wanted to study,” she said. After teaching science for two years in Molepolole, Botswana, she returned to the US and worked at the National Center for Atmospheric Research (NCAR), studying plant gas production and climate change. She ultimately obtained her doctorate in Biochemistry at the University of Colorado, Boulder, exploring how plants make the reactive gas isoprene.

In her postdoctoral work at Massachussets General Hospital (Harvard Medical School), Wildermuth then expanded her study of plants to include microbial pathogens, with a special focus on the powdery mildew fungus. “The reason that I love it is that it’s an obligate biotroph, so it can only grow on a living plant,” she said. Powdery mildew is a fungal plant pathogen characterized by white spots on a plant’s leaves. An individual powdery mildew species can typ-

“If you go into a hospital today, the microbes you recover from infected individuals look very different from those that would have been causing disease in the 1900s.”

"I often wonder, if Darwin had written *On the Origin of Species* today, where would he have published?"

ically infect similar host plants, while different species of powdery mildew can infect different sets of plants. "The spores are windborne, and it's almost all asexually reproduced... and they'll make them in five days to a week," Wildermuth explained. This makes them a common nuisance.

Powdery mildew's interactions with the plant host are what most fascinate Wildermuth, who likens them to a mass-balance equation. In this interaction, the powdery mildew establishes a feeding structure in the upper layer of the plant in order to siphon nutrients for itself. The interaction doesn't only touch the upper layer though, as Wildermuth's lab discovered. "Below those, in the mesophyll cells in the plant, the powdery mildew induces endoreduplication," she said. "You go through replication of the DNA, but you don't go through cell division." While it may seem strange to force cells to increase their ploidy (amount of DNA), the mildew has an express goal in doing so. "What we found is that if we compromise the induction of that plant endoreduplication, the powdery mildew doesn't grow as well," Wildermuth recounted. Furthermore, the level of endoreduplication that is induced is positively correlated with the number of spores the mildew can produce.

That isn't the only aspect of the interaction that the Wildermuth lab examined; a protein called MyB3r4 also affects endoreduplication and the growth of the mildew.³ MyB3r4 is a conserved cell cycle regulator that appears in other organisms besides plants, and Wildermuth's team noticed that infected sites demonstrated altered MyB3r4 expression. Wildermuth and her team wanted to find out what happens in the host in the absence of MyB3r4. "We found that when we knocked out that gene, we no longer see this induced endoreduplication," she says, leading to the finding that the MyB3r4 protein could play multiple roles in cell cycle regulation. Currently, Wildermuth's lab is examining how induced endoreduplication in the plant alters plant metabolism to promote the mildew's propagation. She believes her research may have broad-scale agricultural implications in controlling this fungal disease.

On April 5th, the focus of the conference shifted to computational biology. Dr. **Arturo Casadevall**, a Professor of Immunology and Microbiology at Johns Hopkins University, discussed how developments in host-microbe modeling can change the way scientists tackle broad-scale problems. "If you go into a hospital today, the microbes you recover from infected individuals look very different from those that would have been causing disease in the 1900s," Casadevall said. This is because the virulence of a pathogen changes as it mutates and interacts with its human host. But, Casadevall explained, most current microbial models fail to incorporate these adaptations. Without considering the host and other aspects of the microbe's environment, "microbe-centric theories cannot cope with these changes."

In his talk, Casadevall introduced his formula for calculating "pathogenic potential." This formula accounts for the different as-

pects of host-microbe interactions that are neglected in models where all microbes are assumed pathogens. Casadevall wants to tackle the nuances of immunology with a quantitative approach. "If I lived in the ancient world, I'd be called part of the cult of Pythagoras," he said. He believes that "the mathematics of the system will tell you the degree to which anything is predictable," and his current collaboration with biomathematician Aviv Bergman at Yeshiva University turns to dynamical systems as a way to predict trends within biological systems. As Casadevall explained, dynamical systems use a function to describe the time dependence of a point in space. Rather than hold the microbe constant or the host constant, as in most contemporary experimental designs, dynamical systems allow Casadevall and Bergman to look at the interactions of both organisms over time. By analyzing patterns from many simulations, they have concluded that virulence is an emergent property—a characteristic exhibited of a complex system like the host-microbe relationship, rather than an individual microbe. Casadevall explained that microbes "hedge their bets" by randomly varying their behavior in order to better infect their hosts. These small variations can perturb a biological system with dramatic consequences.

Currently, Casadevall studies immuno-compromised individuals who suffer from human immunodeficiency virus (HIV). HIV patients lack adaptive immunity, and are thus more susceptible to fungal infections. But Casadevall explained that immune-compromised individuals are not restricted to patients of disease. "Anybody who ages becomes immuno-compromised," Casadevall explained, "and medical progress is often associated with immuno-suppressed states." For older individuals or those seeking medical treatments, studying fungal infection in compromised immune systems can reveal useful insights into human health. In fact, Casadevall theorizes that fungi are responsible for the emergence of the human species, and mammalian life in general. He cites studies that demonstrate how endothermy, the ability to regulate body temperature, protects mammals and birds against many fungal infections.⁴ Specifically in humans, most fungal diseases are considered "opportunistic"—associated with mutations in the host genome or bacterial infections.

"Instead of asking, 'What killed the dinosaurs?' ask, 'What kept down the reptiles, such that we did not have a second reptilian age?'" Casadevall explained how fossil evidence indicates abundant fungal growth during the time of dinosaur extinction, possibly due to the cool and moist conditions that volcanic activity and resulting ashen skies promoted. Most dinosaurs were probably killed by the cataclysm but the surviving reptiles were ectotherms—their body temperatures were dependent on the temperature of their environments—so fungal diseases from the outside environment may have been able to adapt easily to the dinosaurs' immune systems, potentially killing them. As a result, endotherms with internal body temperatures that exceed the thermal tolerance of fungi might have emerged, leading ultimately to the age of mammals and the evolution of humans. Casadevall illustrated this theory with paintings. "I urge you to draw your science," he said. "It doesn't take that long and it doesn't have to be that good, but I'm really happy with my meteorite."

In addition to our past, Casadevall also illustrated what the future condition of the human race may look like in the context of climate change. "As the planet gets warmer, some of these fungi will

"If we look at a big scary problem, we should stop saying, 'That's astronomical,'" he said, "We should say, 'That's huge. That's *biological*!'"

adapt to higher temperatures." He points out that the average fungal thermal tolerance over the past 30 years has already risen. One such example is *Candida auris*, a fungus that has independently emerged around the globe in the past five years and is killing many immunocompromised individuals.⁵

Though he remains passionate about his ideas, Casadevall has been largely limited to publishing his theories in mini-reviews, "because thought is very difficult to get published."

"I often wonder, if Darwin had written *On the Origin of Species* today, where would he have published?" Casadevall asked. Today, he promotes scientific thought as the Editor-in-Chief of *mBio*, an open access journal sponsored by The American Society for Microbiology that publishes many different forms of science communication including mini-reviews, opinions and hypotheses, commentaries, and perspectives. Casadevall is optimistic about the potential of his research. "I went into science because I like to explore ideas," he said.

Another JGI speaker with a focus in computational modeling was Dr. **Dan Jacobson**, a computational systems biologist at Oak Ridge National Laboratory. Jacobson is a world record holder addressing the world's biggest problems with big data.

Jacobson's team developed the software for the supercomputer Summit, breaking the exascale barrier, which had previously limited computations to a billion floating point operations per second. The use of tensor cores on graphical processing units (GPUs) allowed for a massive boost in computation performance which enabled the discovery of higher-order combinatorial interactions in cells. Modeling these higher-order interactions is leading to a better mechanistic understanding of biological systems. A challenge associated with these tensor cores is the use of mixed precision (16-bit) numbers. However, tensor cores boosted the speed of calculations, allowing scientists to process much larger datasets. "For this application, by going to 16-bit numbers from 64-bit, we actually lost no accuracy whatsoever," Jacobson explained. "There was just a tremendous speed boost." Jacobson would like to increase speed even more by going down to 8- and even 1-bit numbers. He revealed that currently 11% of computing speed is attributed not to calculating, but to communication overhead associated with moving data onto and off of the GPUs. The remaining limitations of computation speed for 16-bit numbers, therefore, may lie not within the calculations themselves, but in the computer's ability to communicate them and relay their information.

Jacobson is using Summit to bring supercomputing to biology. With explainable artificial intelligence (AI), he'd like to create more comprehensive epistatic calculations. Epistasis is the phenomena by which multiple mutations interact to affect a phenotype even each individual mutation has little to no effect by itself. AI looks at many iterations of data and classifies them based on patterns the computer observes. "Most AI methods.... don't tell you the pattern that they're

using," Jacobson said, "What we want for biology is that pattern that the algorithm is using for classification." Explainable AI identifies this pattern that traditional AI algorithms do not normally reveal on their own. Summit can also help geneticists fill in holes in their data by recapturing rare genetic variants. Statistically, these are convenient to ignore. However, "they often have some of the largest effects on phenotypes," according to Jacobson. Supercomputing can also help construct pan-genomes for whole populations, which will more accurately capture what an entire genome of a species looks like.

In addition to improving how scientists interpret data in the lab, Jacobson's work can also benefit agricultural scientists and farmers in the field. "If you design a new plant genotype that's going to make lots of biofuels or food, but then put it someplace where it dies, you have failed," he explains. Jacobson took data from genomes of many different plants and revealed that different climate types correlated with specific genetic adaptations. He is discovering useful genome-wide associations that plant biologists can use to engineer crops for abiotic stresses like drought, and carbon sequestration for bioenergy, both challenges associated with climate change.⁶ "As [biologists] design genotypes that we want to deploy, we want to know whether they'll be successful." Jacobson uses clustering algorithms for different regions of the world to test how successfully different plant genotypes would survive projected climate conditions around the globe.

Jacobson believes these genomic computing techniques have the potential to change the way we look at macroscale biological patterns. "If we look at a big scary problem, we should stop saying, 'That's astronomical,'" he said, "We should say, 'That's huge. That's *biological*.'"

REFERENCES

1. DOE Mission Areas. (n.d.). Retrieved from <https://jgi.doe.gov/our-science/doe-mission-areas/>
2. Firestone, M. (2019, April). Disentangling the Web of Interactions Shaping Microbial Communities in Soil. Presented at 2019 JGI User Meeting, San Francisco, CA.
3. Wildermuth, M. (2019, April). Host Factors in Plant-Pathogen Interactions. Presented at 2019 JGI User Meeting, San Francisco, CA.
4. Casadevall, A. (2012). Fungi and the Rise of Mammals. *PLoS Pathogens*, 8(8). doi:10.1371/journal.ppat.1002808.
5. Richtel, M., & Jacobs, A. (2019, April 06). A Mysterious Infection, Spanning the Globe in a Climate of Secrecy. Retrieved from <https://www.nytimes.com/2019/04/06/health/drug-resistant-candida-auris.html>
6. Jacobson, D. (2019, April). Exascale Biology: From Genome to Climate With a Few Stops Along the Way. Presented at 2019 JGI User Meeting, San Francisco, CA.

IMAGE REFERENCES

7. Brocken Inaglory. (Photographer). (2008). *Grand Prismatic Spring and Midway Geyser Basin* [digital image]. Retrieved from https://commons.wikimedia.org/wiki/File:Grand_Prismatic_Spring_and_Midway_Geyser_Basin_from_above.jpg