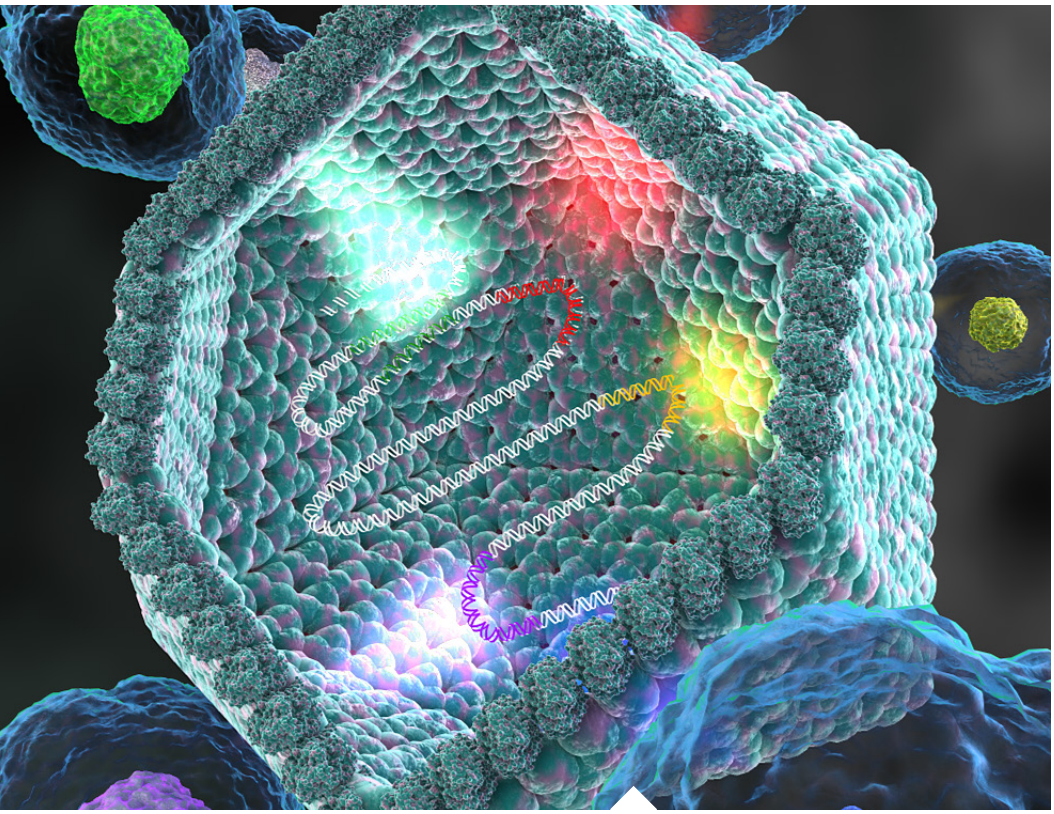
THE PRIMER

JGI JOINT GENOME INSTITUTE DEPARTMENT OF ENERGY

spring 2017
volume 14
issue 2

Novel Giant Viruses Shed Light on Evolutionary Origins



Giant virus acquiring genes from different eukaryotic host cells. (Artistic rendering by Ella Maru Studio)

KRISTINE WONG

Viruses are ubiquitous. They outnumber both the microbes on the planet and the stars in the Milky Way. Giant viruses are characterized by disproportionately large genomes and virions that house the viruses' genetic material. After discovering a novel group of giant viruses with a more complete set of translation machinery genes than any other virus known to date, scientists at the U.S. Department of Energy Joint Genome Institute (DOE JGI), a DOE Office of Science User Facility, named the group Klosneuviruses and believe they significantly increase our understanding of viral evolution.

DOE JGI postdoctoral researcher Frederik Schulz and Microbial Genomics Program head Tanja Woyke unearthed a Klosneuvirus while analyzing data from an Austrian wastewater treatment plant sample. "We expected genome sequences of nitrifying bacteria in the microcolony sequence data," Woyke said. "Finding a giant virus genome took the project into a completely new and unexpected, yet very exciting direction."

The predicted hosts for the Klosneuviruses are protists, and while their direct impacts on protists are not yet worked out, these giant viruses are thought to have a significant impact on these protists that help regulate the *continued on page 5*

in this issue

Surviving the Datapocalypse

We're drowning in data. Thanks to advances in sequencing and computing technology, microbial genomics researchers have so much at their fingertips that they struggle just figuring out what to do with it all.

Welcome to the datapocalypse.

In his closing keynote at the 2017 DOE JGI Genomics of Energy & Environment User Meeting, biological data analytics specialist C. Titus Brown of the University of California (UC), Davis, spoke about his personal journey wrestling with the unique challenges posed by genomic sequence analysis. While sharing strategies and resources that researchers can use to scale the massive size of today's datasets to their purposes, he also reflected on how the user community can work with the DOE JGI to best harness this ever-increasing amount of data.

"The frequency and scale of these inquiries are astonishing," Brown said regarding the number of researchers who ask him how to analyze sky-high piles of sequencing data. "We don't *continued on page 3*

Synthesizing CO_2 Fixation Pathways

Tobias Erb

Many scientists see reducing levels of CO_2 in the atmosphere as a challenge. But for Tobias Erb, a microbial synthetic biologist from the Max Planck Institute of Technology in Munich, Germany, it's a chance.

"Carbon dioxide is a simple [and] ubiquitous carbon source," he said during his opening keynote at the 12th Annual DOE JGI Genomics of Energy & Environment Meeting. "We as scientists and as a society should be able to harvest this molecule and convert it into useful compounds."

Nature, of course, has already covered that job to some degree. Among the many enzymes that fix CO_2, RuBisCO is the most common. Found in plants, algae, and bacteria, it plays a key role in the first stages of the Calvin cycle by converting inorganic CO_2 to the organic form of carbon. But RuBisCO is a slow worker. And from an engineering perspective, Erb said, the additional CO_2 fixation pathways that have evolved in nature are also suboptimal.

So he turned to synthetic biology as a potential solution. For the last few years, Erb and his research team have been working with the DOE JGI to engineer an approach to fixing carbon that would be as good or better than the ones that have evolved from nature.

"Our lab has started to ask different questions," Erb said. "What if we were able to find or design new enzymes for CO_2 fixation? And what if we were able to actually design new pathways that would give us faster access to biomass productivity and convert CO_2 to useful compounds?"

By understanding how biochemical pathways are made in nature first, he emphasized, it's possible to construct synthetic approaches that are more efficient. Using this knowledge as a foundation, Erb's lab designed the so-called CETCH cycle and found a combination of enzymes that fixed CO_2 20 times faster. The experience showed Erb that understanding biological design principles are just as important as the chemistry behind the process. (For more about the work that was described in *Science*, go to jgi.doe.gov/engineering-efficient-system-harnessing-co2/.)

"In principle, we should be able to draw pathways on paper, find enzymes, and implement these paths into living organisms to create designer metabolites," Erb said. He and his team have also implemented a synthetic CO_2 fixation pathway in a living organism, reprogramming it to fix CO_2 *de novo*. He hopes his lab will be able to implement the CETCH cycle into Methylobacterium in the future. "I hope that we can see if our artificial path can be realized in living organisms," he said.

Watch the full talk on the DOE JGI YouTube channel at bit.ly/JGI2017Erb. —KW

Safety in Synthetic Biology

Gene editing has revolutionized biotechnology. Despite its benefits, it also presents new vulnerabilities for genomes of all sizes—even those of entire ecosystems, said Renee Wegrzyn, a program manager at the U.S. Department of Defense's Defense Advanced Research Projects Agency (DARPA). "We can't retrofit safety onto these tools," she said, noting that the technologies are "outpacing our advancement in biosecurity and biosafety." Promoting a safe gene space, Wegrzyn spoke about DARPA's decision to make pivotal early investments in technology aimed at increasing national security, and plans to develop programs that address genomic security and living foundries that advance the level of security in its biomanufacturing capabilities. See the full talk at bit.ly/JGI2017Wegrzyn. —KW

Renee Wegrzyn

The Importance of Diversifying Science

Cat Adams

Most scientists studying the natural world don't delve too deeply into how bias affects the people around them. But research shows it can play a role in hiring, grading, and academic publishing rates—not to mention daily interactions with others.

Now one fungal microbiologist wants to change those dynamics through raising awareness among her fellow scientists. At the DOE JGI Genomics of Energy & Environment Meeting, fungal researcher and UC Berkeley graduate student Cat Adams reminded the audience that countering unconscious bias helps promote more diverse collaborations, leading to more innovative work.

"What's really insidious about unconscious biases is that we can harbor them even when we consciously believe that sexism or racism is wrong," she said. "All our lives we're exposed to negative stereotypes, [which are] widely held but fixed and oversimplified ideas of a particular type of person or thing. And because of this social programming, even really well-meaning intelligent folks like us can act unconsciously with unconscious bias."

But there is some good news. The literature shows that there are measures people can take that are empirically shown to reliably reduce bias and increase diversity instead. "I see three sorts of approaches on how to address bias," Adams said. "We can: 1) enact institutional policy changes to limit the effects of bias; 2) work to reduce our own unconscious biases; and, 3) speak up when other people speak and act in unconsciously biased ways."

For the past two years, Adams has been tackling the issue through the Unconscious Bias Project (UBP), a group comprised of fellow STEM professionals. They aim to educate others in STEM fields through talks and workshops. She left meeting participants with some strategies to overcome automatic biases, all of which researchers have shown as effective: "Replace stereotypes. Individuate instead of generalizing, and get to know them. Take perspective and think about how you'd feel if it happened to you."

See the full talk on the DOE JGI YouTube channel at bit.ly/JGI2017Adams. —KW

Datapocalypse

continued from page 1

have computational tools to deal with all of this. This is a problem for everybody."

Brown decided to tackle the problem by assuming the data for any one project is infinite. Because infinite amounts can never be completely analyzed, he said, researchers must think about what they need to get out of the data, analyze that, and discard the rest. Additionally, he suggested that researchers take both past and future analyses of large-scale datasets to a higher level—through improving reference databases. For example, after reassembling and reannotating the Marine Microbial Eukaryote Transcriptome Sequencing Project using a trio of new bioinformatic tools, Brown's lab found 30 percent new annotations compared to the first time around.

"Tools improve, and we should be thinking about reanalyzing old datasets that are of value with new tools," he advised.

In closing, Brown offered suggestions on what DOE JGI users can do to take better advantage of what it's offering the community: integrate multiple data types; develop better metadata exploration tools; and, boost computational training.

"The only way we're going to escape this trap we've gotten ourselves into with the datapocalypse is to train our graduate students and post docs to do this work with us in the future," he said.

See the full talk on the DOE JGI YouTube channel at bit.ly/JGI2017Brown. —KW

Calling for Shifted Perspectives

MASSIE SANTOS BALLON

"JGI's not stopping its sequencing [of] data, so we're going to have all the data we want, and it's going to be our job to make sense of it," C. Titus Brown of UC Davis said in his closing keynote. "What can we do as community to better take advantage of it?" (See page 1.)

Nearly 500 registered attendees participated in this year's meeting. The opening keynote, delivered by Tobias Erb of the Max Planck Institute of Technology, talked about the challenges of designing or discovering new CO_2 fixation reactions that are faster and more efficient (more on page 2). Renee Wegrzyn spoke of DARPA's contributions towards anticipating control over genome editing and developing the safest ways to innovate in synthetic biology (more on page 2).

Brown was not the only one who talked about the challenges of tackling datasets. Gloria Coruzzi of New York University described an automated phylogenomics pipeline applied to a Chilean systems biology study. Patsy Babbitt of UC San Francisco talked about the challenges of mapping the enzyme universe. Abhishek Biswas from the University of Tennessee talked about the genome assembler Disco. Building on the poplar reference genome, Jay Chen of Oak Ridge National Laboratory promoted QTLs for lignin biosynthesis genes.

"If we can assemble genomes, we can practice bias reduction," said UC Berkeley graduate student Cat Adams, reminding the audience that countering unconscious bias helps promote more diverse collaborations, leading to more innovative work (see page 3).

Several talks focused on how organisms interact with their environments at the micro and macroscale. Jan Leach from Colorado State University spoke about how bacteria influence Russian wheat aphid virulence. Jonathan Lynch from Penn State University advocated for going "cheap, steep, and deep," to develop crops with deeper roots in order to access nutrients and sequester carbon. Philipp Zerbe from the University of California, Davis focused on plant terpene metabolism in both model and non-model species. Among the fungal researchers, Sunny Liao of the University of Florida described Pine-*Suillus* studies while Teresa Pawlowska of Cornell University spoke about mutualistic relationships between fungi and bacteria.

Mary Lipton from the Environmental Molecular Science Laboratory (EMSL) promoted the national user facility's capabilities that enable

Posters on display at the 12th Annual DOE JGI User Meeting

Qingpeng Zhang, Titus Brown, Lisa Cohen, and Harriet Alexander

plant-soil ecosystem studies at a number of scales. Since 2014, the collaborative science initiative between EMSL and the DOE JGI has enabled researchers to harness the capabilities at both user facilities. One beneficiary is Colleen Hansel of the Woods Hole Oceanographic Institute, who has been able to deconstruct the process by which fungi are involved in manganese oxidation.

Some researchers shared their scientific quests. Harry Beller of Joint BioEnergy Institute focused on the hunt for a sustainable way to produce toluene. Stephen Mondo of the DOE JGI described a new epigenetic modification in the fungal kingdom while Thomas Mock from the University of East Anglia spoke about how the diatom *Fragiliaropsis cylindrus* has adapted to the Southern Ocean (more on page 6.) Andrew Roger of Dalhousie University talked about the protist *Blastocystis*.

Finally, several talks covered the question of scale. Trina McMahon of the University of Wisconsin-Madison promoted long-term time series studies while describing microbial population studies in freshwater lakes. Echoing her call, Lauren Alteio of UMass Amherst described soil and plant studies that are making use of decades of data collected from a long-term ecological research site at Harvard forest. Mikael Andersen of the Technical University of Denmark talked about a genus-wide study of *Aspergillus* (more on page 7.) J. Chris Pires of the University of Missouri talked about a long-term Brassica comparative genomics project. Alex Greenspan of UC Davis focused on finding geographical patterns of bacterial symbionts of chickpea.

Selected videos from the 2017 Genomics of Energy & Environment Meeting are on the DOE JGI Youtube channel: bit.ly/JGI2017_Videos. For the social media recap of this Meeting, go to bit.ly/JGI2017Storify.

Novel Giant Viruses

continued from page 1

planet's biogeochemical cycles. The DOE JGI published the findings in the journal *Science* on April 7, 2017 with collaborators from the National Institutes of Health (NIH), University of Vienna, and California Institute of Technology.

Scientists have been fascinated by giant viruses since 2003. A handful of other giant virus groups have been found, leading to the emergence of two evolutionary hypotheses regarding their origins. One posits that giant viruses evolved from an ancient cell, perhaps one from an extinct fourth domain of cellular life. Another presents the idea that giant viruses arose from smaller viruses.

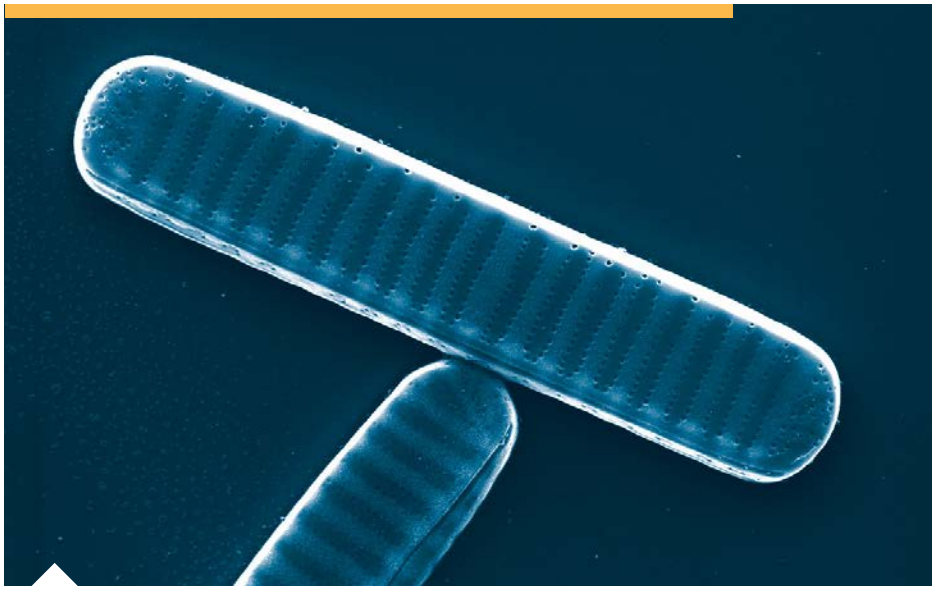
The discovery of Klosneuvirus supports the latter idea, said Woyke, senior author of the paper. "In this scenario, a smaller virus infected different eukaryote hosts and picked up genes encoding translational machinery components from independent sources over long periods of time through piecemeal acquisition," she said.

Co-author and NIH senior investigator Eugene Koonin believes there are more giant viruses waiting to be discovered in metagenomic data. "The discovery presents virus evolution for us in new ways, vastly expanding our understanding of how many essential host genes viruses can capture during their evolution." More about this story at bit.ly/JGIKlosneuvirus.

Frederik Schulz spoke about the team's discovery of the novel group of giant viruses at the DOE JGI 12th Annual Genomics of Energy & Environment Meeting. Watch his talk at bit.ly/JGI2017Schulz. Or listen to his Science Friday interview at bit.ly/SciFriKlosneuvirus.

DOE JGI authors on the Klosneuviruses paper: (left to right) Prokaryote Super Program Head Nikos Kyrpides; Functional Annotation Group Head Natalia Ivanova; study senior author and Microbial Genomics Program Head Tanja Woyke; and, study first author Frederik Schulz.

A Diatom's Adaptation to the Southern Ocean



Scanning electron micrograph of two cells of *Fragilariopsis cylindrus*. Shown are two silica shells (Frustules) in valve view. Magnification: 15,000X; scale bar: 5 µm (Gerhard S. Dieckmann)

In both freshwater and marine ecosystems, the base of the food web is comprised of a diverse community of phytoplankton that includes diatoms who can thrive in a wide range of temperatures. In the Southern or Antarctic Ocean, large populations of a particular diatom, *Fragilariopsis cylindrus*, dominate the phytoplankton communities.

To learn more about how *F. cylindrus* adapted to its extremely cold environment, a team led by University of East Anglia (UEA) scientists in Norwich, England conducted a comparative genomic analysis involving three diatoms by tapping experts from the DOE JGI, who conducted all sequencing and annotation. The results, reported January 26, 2017 in the journal *Nature*, provided insights into the genome structure and evolution of *F. cylindrus*, as well as this diatom's role in the Southern Ocean. Of particular interest was that *F. cylindrus*, which is diploid (it has two copies of each chromosome, thus two versions of each gene) can selectively express the variant best suited to helping it deal with its environment. This provides additional genome-rooted resilience to the organism as its environment changes.

"Many species including phytoplankton are endemic to the Southern Ocean," said Thomas Mock of UEA, who led the study. "They have evolved over millions of years to be able to cope with this extreme and very variable environment. How they did that is largely unknown. Thus our data provide first insights into how these key organisms underpinning one of the largest and unique marine ecosystems on Earth have evolved." Mock spoke about his work on diatoms at the 12th Annual Genomics of Energy & Environment Meeting. Watch his talk at bit.ly/JGI2017Mock. —MSB

IMG Reveals Previously Unknown Protein Structures

Proteins largely form a cell's structures and carry out its functions: they control growth and influence mobility, serve as catalysts, and transport or store other molecules. Comprised of long amino acid chains, the one-dimensional amino acid sequence may seem meaningless on paper. When viewed in three dimensions, researchers can see the long amino acid chains that comprise a protein's structure, and more importantly, how the protein's folds determine its functions.

There are close to 15,000 protein families—groups of families that share an evolutionary origin—in the database Pfam. For nearly a third (4,752) of these protein families, there is at least one protein in each family that already has an experimentally determined structure. For another third (4,886) of the protein families, comparative models could be built with some degree of confidence. For the final third (5,211) of the protein families in the database, however, no structural information exists.

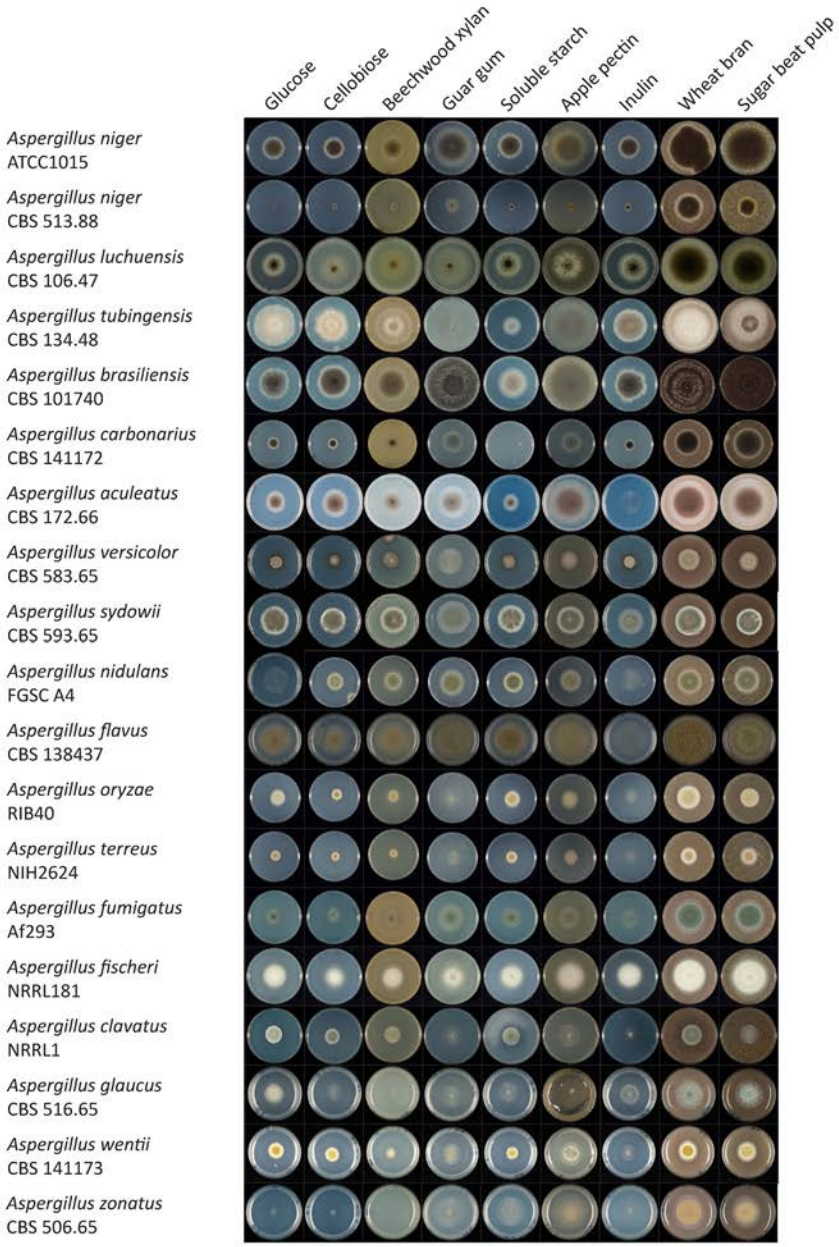
In the January 20, 2017 issue of *Science*, a team led by University of Washington's David Baker in collaboration with DOE JGI researchers reported that structural models have been generated for 614 or 12 percent of the protein families that had previously had no structural information available. "That this could be accomplished using computational modeling methods was not at all apparent 5 years ago," the team noted in their paper. This accomplishment was made possible through a collaboration in which the Baker lab's protein structure prediction server Rosetta analyzed the metagenomic sequences publicly available on the Integrated Microbial Genomes (IMG) system run by the DOE JGI. —MSB

Diversity for Industry: Sequencing the Genus Aspergillus

As the majority of its 350 species have yet to be sequenced and analyzed, researchers are still at the tip of the iceberg when it comes to understanding *Aspergillus*' full potential and the spectrum of useful compounds they may generate. Many species in this genus play critical roles in industrial applications, including biofuel production, and in plant and human health.

In a study published February 14, 2017 in the journal *Genome Biology*, an international team including DOE JGI researchers reports sequencing the genomes of 10 novel *Aspergillus* species, more than doubling the number of *Aspergillus* species sequenced to date. The newly sequenced genomes were compared with the eight other sequenced *Aspergillus* species. With this first ever genus-wide view, the international consortium found that *Aspergillus* has a greater genomic and functional diversity than previously understood, broadening the range of potential applications for the fungi considered one of the most important workhorses in biotechnology.

"Several *Aspergillus* species have already established status as cell factories for enzymes and metabolites. However, little is known about the diversity in the species at the genomic level and this paper demonstrates how diverse the species of this genus are," said study lead author Ronald de Vries of the Westerdijk Fungal Biodiversity Institute in the Netherlands. "One can't assume that an *Aspergillus* species will have the same physiology as a better studied species of the genus." The study, conducted through the DOE JGI's Community Science Program, also demonstrates the importance of evaluating biodiversity within a genus to understand how fungi can be greater utilized to solve a variety of problems. Study co-author Mikael Andersen spoke about this project at the 12th Annual Genomics of Energy & Environment Meeting. Watch his talk at bit.ly/JGI2017Andersen. —KW



Comparative growth of Aspergilli. (Ad Wiebenga & Ronald de Vries, Westerdijk Fungal Biodiversity Institute, Utrecht, The Netherlands)

To learn more about each of these stories, go to the JGI's News & Publications page: jgi.doe.gov/news-publications/.

Mouncey Named DOE JGI Director

Nigel Mouncey

After a 9-month national search, Nigel Mouncey was selected as the next DOE JGI Director. Mouncey is the DOE JGI's fourth Director in its 20-year history. He succeeded Eddy Rubin, who announced his retirement in March 2016 after 14 years as Director, and DOE JGI Science Deputy Axel Visel, who served as Interim Director. Mouncey joined the DOE JGI in time for the DOE JGI's 12th Annual Genomics of Energy & Environment User Meeting.

Mouncey came to the DOE JGI from Dow AgroSciences in Indianapolis. There, among his many achievements, Mouncey directed a 70-member R&D team that supported the growth of a highly successful natural product insecticide that has since generated hundreds of millions of dollars of revenue and significant societal benefit through isolating, optimizing, and scaling-up of new production strains for commercial manufacturing by fermentation. He also built an integrated and highly effective bioprocessing team that in turn developed production strains and fermentation processes for other molecules such as a new fungicide and propionic acid, as well as supporting the discovery of new crop traits. More about Mouncey at bit.ly/JGIMounceyhelm.

Breaking Ground on the DOE JGI's New Home

On January 31, 2017 at Berkeley Lab, a groundbreaking ceremony was held for the Integrative Genomics Building (IGB) that will be the home of DOE JGI and DOE Systems Biology Knowledgebase (KBase) in 2019. Flanked by dignitaries representing the Berkeley Lab, the Department of Energy Office of Science, and the University of California, DOE JGI Director Nigel Mouncey and KBase Principal Investigator Adam Arkin clinked beakers of soil in a toast to their future colocation.

"This is just the start of what I think is a really exciting time for the Biosciences," said Sharlene Weatherwax, Associate Director of Science for Biological and Environmental Research (BER) at DOE Office of Science. Full story at bit.ly/IGBgb.

Furthering JGI-UC Merced Joint Training Opportunities

In 2014, DOE JGI Science Deputy Axel Visel and Genome Analysis Group Lead Zhong Wang, both adjunct faculty members with the University of California, Merced, conceived a summer program that challenges UC Merced graduate students with individual projects at the DOE JGI, offering hands-on experience in cutting-edge genome research. Continuing joint training opportunities between these two institutions, in Summer 2016, Fabian V. Filipp, Director of Systems Biology and Cancer Metabolism and UC Merced's Program for Quantitative Systems Biology, conducted an advanced training course that included in-depth training from manipulation and stable isotope labeling of eukaryotic cells in tissue culture to metabolomics data acquisition and network based data modeling. The main goal of this advanced training course was to bring together systems biologists with wet lab scientists to share and present their work on state-of-the-art concepts in cancer metabolism. The faculty that provided hands-on training for 20 students, many of them first generation underrepresented academics, included Berkeley Lab scientists Trent Northen and Ben Bowen. Plans for Summer 2017 include practical workshops and the opportunity for UC graduate students to visit the DOE JGI. For more information please visit systemsbiology.ucmerced.edu.

Register Now!

National Microbiome Data Collaborative Town Hall ASM Microbe 2017
June 3, 2017
bit.ly/ASMMicrobeNMDC

Microbial and Plant Systems Modulated by Secondary Metabolites Meeting
July 24–26, 2017
bit.ly/2ndry-metas

Mining Microbial Genomes and Metagenomes for Biotechnological Applications Workshop SIMB Meeting
July 30, 2017
bit.ly/IMGatSIMB17

Contact The Primer
David Gilbert, Managing Editor
DEGilbert@lbl.gov
Massie Santos Ballon, Editor

PAC 17-JGI-3944