

TO: Committee on Science, Space, and Technology  
FROM: Kelly C Wrighton, PhD  
DATE: September 09, 2020  
RE: Written Testimony, “Biological research at the Department of Energy: Leveraging DOE’s unique capabilities to respond to the COVID-19 pandemic”

Below I provide my written testimony to the three main questions I was asked to address in response to the U.S. Department of Energy Office of Science’s (DOE) Biological and Environmental Research (BER) program.

***(i) What are benefits of the Department’s Bioscience Research Capabilities?***

As you have heard from our other panelists, the DOE has a long history of biological research. This tradition of pioneering scientific innovation continues today and is manifested by ongoing DOE investments in The Joint Genome Institute (JGI) and Environmental Molecular Sciences Laboratory (EMSL). Although I am not directly affiliated with these facilities, I am a self-proclaimed super-user, meaning that since my laboratory’s inception, I have 9 different project and awards with EMSL and JGI that I have led (this does not account for other collaborative efforts). In sharing my narrative with you, I want you to take away how these facilities propel science innovation in this country, and especially benefit those like myself at the earliest stages of their independent research programs. Simply, these user facilities provide foundational biological resources that support scientific advances across environmental, energy and human health sectors for the United States.

Starting a research program at a R1 university is much like starting your own small business. Essentially your job is to take the universities investment in you, called start-up funding, to finance a research program doing impactful, innovative science. The goal is that short term investment from the University can enable one to compete for federal research grant dollars that sustain, and further fuel expanded scientific endeavors by your research program.

User facilities play a vital role in that transitional phase of early career scientists. The facilities allow scientific researchers like myself to maximize their investment by (1) scaling the scope of their science to levels unachievable by a single, academic laboratory, (2) providing accessibility [minimizing privilege] to technical resources located beyond the walls of any building or campus, and (3) networking motivated, early career scientists with trained experts who are at the very frontier of this technology. In return, the user facilities create large amounts of environmental data for databases and develop expertise beyond that retained within the DOE national laboratory systems. This is a symbiotic relationship between individual researchers and the user facility that benefits the entire scientific community.

To better articulate this value, let us use a single example (of many) from my research portfolio. The year is 2014 and I have just started my brand-new research program at The Ohio State University. Due to my geographic proximity to the emerging shale oil and gas industry in the Appalachian Basin, I became interested in interrogating the microorganisms that colonize hydraulically fractured shale wells. These contaminating microbes from the earth’s surface can prosper 2,500 meters below the surface in shale energy wells. Here this unexpected microbial growth can sour the valuable hydrocarbon commodity, corrode the well infrastructure, and maybe, alternatively if managed in a different fashion, can be harnessed for greener, energy efficiency or yield.

My university start-up funds allowed me to hire the personnel and collect and process 10 or so samples from a single well collected over a year period. At the same time, I competed and was awarded EMSL and JGI resources to further advance this research. This collateral resource investment by the Biological arm of DOE allowed me to scale my research enterprise from this handful of samples in a single West Virginian

well, to over 50 wells across North America, with sample collection occurring over thousands of days. Moreover, through my collaboration with DOE user facilities, I was exposed to new, higher resolution analytical chemical methods that had not been applied to these systems and were not readily accessible at my institution. This technology led to new insights about how and what these microbes were living off in the deep subsurface in this new, engineered system. Moreover, EMSL and JGI scientists became my collaborators, providing invaluable insights into the project from their multi-disciplinary perspectives.

Those early investments by DOE user facilities led to scientific publications that helped develop me as a research leader in this domain in a few short years, with my laboratory's findings awarded the top 100 discoveries by Discovery magazine in 2016. More important, this collaboration generated data that facilitated my future fiscal support from applied projects supported by US industry, as well as more foundational scientific efforts by the National Sciences Foundation. Research collaborations provided by user facilities like JGI and EMSL catalyzed impactful biological science, with recognition culminating in my PECASE (the presidential early career award for scientists and engineers) award that represents the highest honor bestowed by the US government to scientists in the early stages of their independent research careers. In summary, DOE investments in user facilities amplify the innovation and discoveries of our US scientific enterprise.

***(ii) How has BER expertise and their advanced tools been leveraged to respond to the COVID-19 pandemic?***

A quick scan of others in this panel led me to believe the direct contributions of the DOE to COVID-19 research will be better covered by others like my esteemed colleague Dr. Mary Maxon. An area that I am uniquely qualified to speak to is the translational value of DOE investment in environmental research. Essentially, how does investment in one scientific arena energize or cross-pollinate other parallel scientific discovery?

Let me contextualize this quickly. The hydraulic fracturing example that I described, where we gained insights into the chemical diet that enabled microbes to adapt and persist to life thousands of meters below the surface, has direct applications to the microbes living in your gut. In fact, leveraging our newly developed technologies for working with complex samples, and our newfound understanding of the microbial reactions catalyzed, it turns out those same processes predict why one person is more susceptible to cardiovascular disease than another person. The microbes in your gut use the same enzymes and produce the waste products as those in shale rock. As such I was able to take my DOE and NSF supported research and compete and earn support from the National Institutes of Health for research in medicine. Biological discovery is seamless and transcends disciplinary boundaries.

You don't have to know much about biology to know that extracting microbes from shale rock and human feces is not the cleanest or easiest of work. Work by myself and others in environmental sciences involving extraction of DNA and RNA from complex matrices has yielded knowledge that is leveraged by my colleagues working with RNA viruses in wastewater for Sars-CoV-2 research. Moreover, even prior to Covid pandemic, DOE BER was leading investments in viral mechanistic ecology - advancing research into how and where viruses establish and function. Through their funding calls that academics like myself compete for, a critical area of research has been dedicated to understanding viral roles in the environment. Recent discoveries by my group and others have uncovered how critical viruses are to maintain the functional balance in our soils, our rivers, our own guts, and even the deep subsurface. As such, DOE investments have led research for development of viral genomic tools that enable insights ranging from identifying and categorizing the viral world, to elucidating the biochemistry contained within it. For example, earlier this year I was funded through the DOE Genomic Sciences program to participate in a multi-disciplinary research team containing microbiologists, chemical engineers, and computer scientists. Here we are devising new genomic algorithms for defining the chemical reactions encoded, but currently enigmatic with viral genomes. This new software will yield translational content facilitating viral

environment and health research avenues. Beyond direct investments and repurposing of DOE infrastructure for virtual interaction as discussed earlier, DOE enabled biological platforms and technological insights support this pandemic. In summary, BER supported expertise and tools are immediately transferable to solutions of existing and emerging environmental, energy, and health challenges.

***(iii) What are my perceptions of possible future directions for the Departments biological research activities?***

Despite the advanced capabilities the User Facilities shepherd there are gaps in our nations capabilities to ensure we remain as the world's premier scientific enterprise. We need to address the increasingly complex challenges residing in the biological systems of present. Currently genomics information is being generated faster than the corresponding proteomics, metabolomics and structural imaging capabilities can keep up, and more so that our computational infrastructure can mine.

The result is a list of 10's or 100's of thousands of proteins that have potentially critical yet unresolved function. Our recent DOE-supported discovery in *Science*, a premier research journal, last week demonstrated how even within fairly well characterized bacterium, we identified a mis-annotated pathway that can generate ethylene without using oxygen. While this discovery has important ramifications for industry, as it represents a non-combustible pathway for generation of this industrial precursor, it also directly explains many environmental phenomena such as the observation that ethylene accumulates in soils after rain events with toxic outcomes for crops. Given that this process is catalyzed by less than 5 genes in a pathway, and bacteria encode several thousands more genes that have this same unresolved content, think of how much discovery is awaiting us.

What is needed to expedite the speed in which researchers can translate genomics information into actionable knowledge?

- (1) A coordinated, organized computational infrastructure that enables computer aided pattern recognition of this deluge of environmental genomics/microbiome data. Efforts led by DOE supported national microbiome data collective (NMDC) are paving this way. The NMDC is a first of its kind microbiome central repository that links the environmental context, the genomics, and the environmental chemistry across a range of ecosystems that DOE supports. This indexing and cross-linking of data will allow for machine learning and artificial intelligence data mining of content that can steer new experimentation.
- (2) Research automation and scale that extend beyond the resources of any one laboratory. Imagine a future, where you can be shipped modular labs to your workspace or ship your samples and login to a command center and run samples through robotic systems. This virtual, shared lab allows for resource use efficiency, data generation that is customized, reproducible, productive. The same could be said of computational resources. While shared computational resources are supported by DOE and NSF, they aren't readily accessible for common workflows, timelines and data storage to enable efficient genomics research. The future of science extends beyond 4 walls and personalized compute power. There is redundant investment being made with federal dollars, and perhaps the future will be streamlined workspaces needed to advance and capitalize on our genomic investment.
- (3) More so, the heart of future innovation lies in multi-disciplinary collaborative nexus. Examples like you have heard from my other colleagues today, where new teams and resources rapidly emerged to address knowledge gaps for COVID19 research. Research incentives that drive this cross-agency, multi-disciplinary, higher risk innovation are needed, and ideally prior to, not in spite of, the next global challenge.

This streamlined and cross-disciplinary scientific space will allow us to embark on a new era for scientific discovery, leveraging the recently developed foundational genomics infrastructure. Human, physical, and computational capital is required to position the US to lead breakthrough applications in biological sciences in the next decade.