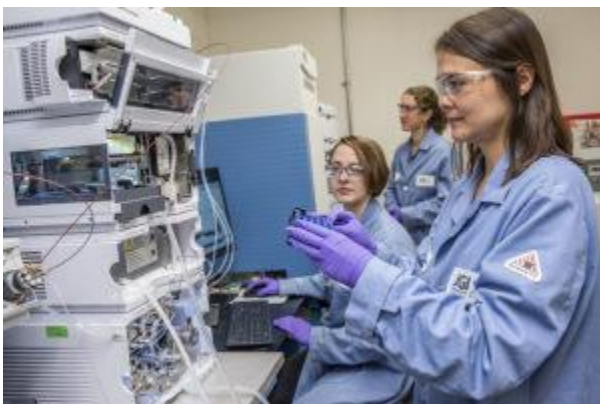# JGI Partnership Case Study: Brightseed

Katherine Louie (foreground), JGI Metabolomics Platform Group Lead.

In response to varying environmental conditions, plants produce a panoply of chemical compounds, the vast majority of which have yet to be characterized. Therein lies an opportunity: to identify specialized metabolic pathways that can be engineered for ecological, agronomic, specialty chemical, or human health applications. The question is: can the analysis of advanced multi-omic data, such as those generated by high-throughput DNA sequencing and liquid chromatography-mass spectrometry (LC/MS), be brought to bear on natural product discovery?

Brightseed, a San Francisco-based biosciences company founded in 2017, has developed Forager™, a computational intelligence whose learning objective is to understand nature more profoundly than humans can. Forager is fueled by Brightseed's growing proprietary data on the world's plant life and deep knowledge of human biology. Forager interrogates and learns with unprecedented speed and scale, discovering natural compounds that solve unmet human needs, including addressing chronic health indications and other emerging market opportunities. To drive this effort forward, Brightseed sought a partner that could help them determine how best to integrate large-scale metabolomic and transcriptomic data sets into Brightseed's discovery process. That's where the U.S. Department of Energy Joint Genome Institute (JGI) factored into the equation.

"We were in need of a unique combination of expert domain knowledge of plant natural products, with access to mass spectrometry characterization of plant tissue, and the specialized bioinformatics tools to interpret the most complex data sets," said Lee Chae (PhD, UC Berkeley, Plant Biology, Computational and Genomic Biology, 2008), Brightseed Co-Founder and Chief Technology Officer. He turned to Trent Northen and his colleagues in the JGI's Metabolomics Technology Group.

JGI performed a secondary metabolite analysis on a subset of Brightseed's plant compound library to enable the company to gauge detection capabilities and limitations, compound abundance, and to identify and annotate secondary metabolites from plant extracts from three underrepresented flowering plant families with a history of human consumption.

"What particularly accelerated progress toward achieving Brightseed's goals was gaining access to MAGI, and the sage interpretation of the results by Ben," said Gabriel Navarro, (PhD, UC Santa Cruz, Natural Products Chemistry, 2013), Brightseed's program lead for the collaboration with JGI. MAGI, the Metabolite Annotation and Gene Integration tool, is the brainchild of Ben Bowen in the Northen Lab. MAGI provides a fundamentally different approach for directly linking novel sequences to their biochemical functions and products by generating a metabolite-gene association score using a biochemical reaction network.

"Metagenomics and single-cell sequencing have provided us a glimpse into the vast metabolic potential of Earth's complex biological systems," Bowen said. "Yet, we've been stymied in our ability to accurately predict and identify the products of most biosynthetic pathways. Most of what we have known of microbial biochemistry was based on characterization of a few model microorganisms, and through sequence correlations based on publicly-available data. We saw this as an opportunity, so we built MAGI to make connecting metabolomics data with genes easier for researchers."

MAGI also enabled Brightseed to identify critical knowledge gaps and to quickly close them.

"JGI's team was very knowledgeable in guiding Brightseed through the project, helping us understand their technology platform and providing us what we needed to know so that we could consider investing in our own instrument," Navarro said. "Our overall positive experience with JGI—clear communication, rapid turn-around, and high-quality results—encourages us to pursue a future project in the near-term."

Bowen, Northen and other colleagues from the Lawrence Berkeley National Laboratory's Environmental Genomics and Systems Biology Division in the Biosciences Area, the Data Analytics and Visualization Group of the Computational Research Division, and the National Energy Research Scientific Computing Center (NERSC) recently published an article about MAGI in *ACS Chemical Biology*.

MAGI is freely available for academic use, both as an online tool at https://magi.nersc.gov, and with source code available at https://github.com/biorack/magi. For more information about how to partner with the JGI, see the different mechanisms described here. To learn more about Brightseed here.