Title: Enabling adaptive biological experiments through the integration of laboratory automation with machine learning

Background: Laboratory automation is rapidly evolving to the point where whole experiments are performed without human involvement. At the same time machine learning and increasing computational resources are automating data analysis. Together, these disruptive technologies will enable experimental designs not that are not currently possible.

Illustrative use cases:

Designing experiments based on plant developmental milestones. Understanding of plant biology and plant-microbe interactions are central to DOE missions in sustainable energy and understanding the carbon cycle. Currently, plant and plant-microbe interaction studies are predominately performed using time-based sampling strategies. However, there are typically large variance in plant development necessitating large numbers of biological replicates. Integration of experimental capabilities such as automated plant and microbial cultivation, hyperspectral imaging, and sampling with machine learning and high-performance computing (HPC) would enable researchers to specify plant developmental milestones (e.g. secondary root formation, as assessed using imaging) to guide sampling and interventions. This would dramatically reduce experimental sizes and enable investigation of new biological phenomena. Given the large size of the imaging data (>10TB/study) this will require the capability to detect and analyze features, reduce data, perform image segmentation and registration, among others.

Enabling adaptive experimental designs for discovery and optimization. Integration of automated experimental capabilities with machine learning and high-performance computing will enable new experimental that adapt based on intermediate results. For example, integration of the automated plant-microbe experimental capability with machine learning and HPC described above can screen for microbiomes that result in the target attributes (e.g. plant biomass) and then to set-up new experiments based on the best performing microbiomes. In the short-term, this capability could be used to discover and optimize high performing microbiomes. Longer-term, experimental capabilities using CRISPR-Cas and synthetic biology could be used to rapidly optimize bioenergy crops, etc.

Together the integration of HPC, machine learning, and automated biological experimentation will greatly accelerate science by increasing reproducibility and enabling new, more efficient experimental designs.