Accelerating engineered microbe optimization through machine learning and multi-omics datasets

DE-EE0008489, PI: Jeffrey Dietrich

Project Type

- DE-FOA-0001916
- Decreasing the development cost and time for cost-competitive production of new bioproducts from lignocellulosic biomass.

Timeline

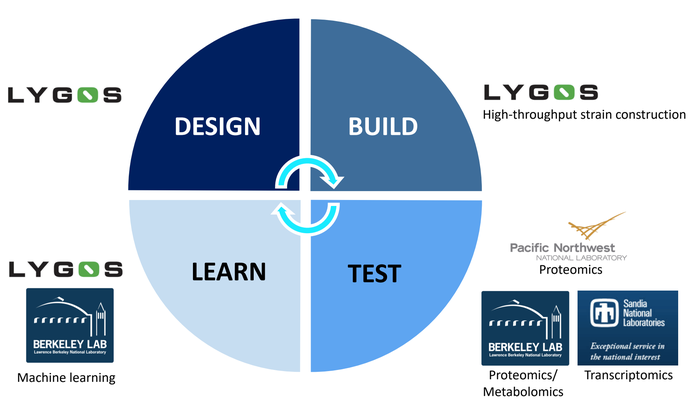
- **Project Start Date:** Jan. 1 , 2019
- **Project End Date:** Mar. 31, 2021
- **Percent Completed:** 5% Completed

	Budget Period 1 Funding	Budget Period 2 Funding	Budget Period 3 Funding	Total Funding
DOE Funded	$18,558	$286,285	$295,157	$600,000
Project Cost Share	$7,953	$358,408	$490,782	$857,143

Additional DOE funding in the amount of $1,400,000 to FFRDC partners as follows:
Lawrence Berkeley National Labs (62.5%)
Sandia National Labs (21.4%)
Pacific Northwest National Labs (16.1%)

Management Approach

- This project is a collaboration between Lygos and the Agile BioFoundry (ABF) labs, namely Sandia National Lab (SNL), Lawrence Berkeley National Labs (LBNL) and Pacific Northwest National Labs (PNNL) with an appointed technical lead from each of these institutions.
- The roles and responsibilities of each of the institutions are summarized in the figure below and have been chosen based on the expertise and capabilities of the respective institutions.



- All data will be stored and processed through the Experimental Data Depot (EDD), a web-based software tool developed by the Agile Biofoundry (ABF).
- Monthly meetings will be held with participation from all technical leads to discuss and address any technical, management or performance related issues.
- All members will also contribute to the quarterly progress reports vis-à-vis the performance milestones.

Technical Approach

- Demonstrate via a high-throughput Design-Build-Test-Learn (DBTL) cycle workflow that malonic acid production can be improved in engineered *Pichia kudriavzeii.*
- In the Test Phase, time-series, multi-omics data will be collected and used in the Learn phase for machine learning to generate a predictive model of the pathway dynamics.
- *P. kudriavzeii* is a non-model microbe; therefore developing a sufficiently rich understanding of cellular metabolism is needed to accurately inform metabolic engineering designs.
- Improving the operational efficiencies of the DBTL cycle will be key to accelerating the biomanufacturing cycle and shorten the time needed to bring a product to commercial market.
- As the number of DBTL cycles performed increases, the predictive accuracy of the machine learning algorithms will also improve and therefore quicken the pace of metabolic engineering R&D.
- Strain performance will be confirmed in benchtop scale fermentations using cellulosic glucose as the feedstock

Relevance and Impact

- The goal of this project is to decrease the development cost and time for cost-competitive production of new bioproducts from lignocellulosic biomass.
- Production of bioproducts from non-food biomass sustains domestic production of biofuels, reduces greenhouse gas emissions, and helps displace the whole barrel of oil.
- Demonstrate an industrially relevant bioprocess using machine learning to accelerate the biomanufacturing cycle by shortening metabolic engineering R&D.

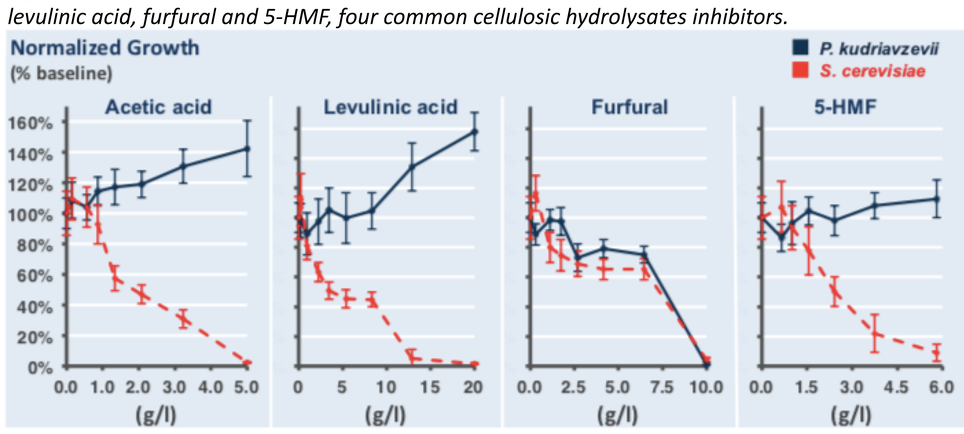
Future Plans

- Baseline malonic acid fermentation metrics at 1-L or 2-L scale using cellulosic glucose as the feedstock will first be validated by Department of Energy technical representatives.
- Once baseline metrics are validated, multi-omics method development will be performed for *P. kudriavzevii* and a protocol established for high-throughput implementation.

Technical Progress

- Work related to this project will not commence until the Department of Energy (DOE) has performed an initial validation to establish baseline fermentation performance data, which will be completed Q1-2019.
- It has been previously established that *P. kudriavzevii* has a multi-stress tolerant physiology, which makes it a good host from an industrial biotechnology perspective.
- As shown in Figure 1, *P. kudriavzevii* has demonstrated tolerance to many byproducts commonly found in cellulosic hydrolysates, making it an amenable host to cellulosic glucose as a feedstock.

Figure 1. Growth comparison between P. kudravzeii and S. cerevisiae in the presence of acetic acid, levulinic acid, furfural and 5-HMF, four common cellulosic hydrolysates inhibitors.



- Lygos is leveraging its expertise in the genetic engineering of *P. kudriavzeii* and has invested in a dedicated strain Build team tasked with efficiently converting designs to completed, sequence verified strains.
- Significant strides have been made in FY18 in increasing the total number of strains built, and decreasing the overall cost and turnaround time for strain generation (Figure 2).
- Continual efforts towards improving these metrics will remain a major focus as part of the larger endeavor to shortening the overall DBTL cycle.

Figure 2. Lygos strain Build team operation progress and upcoming targets for (A) number of strains built, (B) cost per strain and (C) average strain build time.