Reinforcement Learning to Optimize Medium Chain Fatty Acid Production from Lignocellulosic Stillage using Anaerobic Microbial Communities

Madeline Hayes[1,2] * (mmhayes4@wisc.edu), Jaron Thompson,[1] Jun Feng,[1,2] **Ophelia S. Venturelli**[1,2]

[1]University of Wisconsin-Madison, Madison, WI; [2]DOE Great Lakes Bioenergy Research Center, Madison, WI

Project Goals: The goal of this research is to apply a data-driven approach to guide the design of microbial communities that maximize production of target metabolites.

Microbes exist in complex communities in every environment on earth, leveraging their flexible metabolic capabilities to perform a wide variety of chemical transformations. Anaerobic fermentative communities perform many functions of interest with collective metabolism, but complex natural communities have proven very difficult to explore due to increasing dimensionality. Understanding the ecological dynamics of these communities has proven challenging; even with the advent of advanced sequencing techniques that allow unprecedented characterization of community composition, frameworks for understanding competition and cooperation as they relate to metabolic flux are still relatively undeveloped.

To this end, the integration of machine learning into biological systems to help inform complex and high-dimensional problems is an emerging technique with particularly useful applications in microbiome and microbial community ecology research. Exploration of complex communities using machine learning to build computational models that can handle high-dimensional problems and predict target metabolic functions holds much promise for engineering productive microbiomes in a wide variety of environments, and for a wide variety of production targets. Medium chain fatty acids (MCFA's, fatty acid chains of length C6-C10) are high value compounds used in industrial processes, including as fuels and specialty pharmaceutical precursors. Currently, these compounds are produced or harvested from non-renewable or unsustainable sources, such as petroleum products or palm oil, and therefore represent a desirable target for sustainable production. Leveraging microbial community metabolism to convert plant-based carbon to these valuable products has proven promising but remains a complex problem as community size and dimensionality increases.

We seek a reduced-complexity synthetic consortium of anaerobic bacteria that optimizes the conversion of carbon from lignocellulosic stillage waste into high-value MCFA products. We propose a design-test-learn approach that leverages principles from reinforcement learning to seek optimal experiments that systematically explore and exploit microbial community functions. This iterative approach involves collecting data, building computational models from the data that predict target production, generating rational hypotheses to narrow future experimental spaces, and testing these hypotheses by collecting more data. The modelling approaches employed here include the generalized Lotka-Volterra model, which predicts community dynamics based on monospecies growth parameters and pairwise interactions, and an ensemble of machine learning methods including linear regression, Gaussian processes, random forest regression, and feed-forward neural networks. These models provide an ensemble of predictions that serve as features for a final Bayesian linear regression model that provides probabilistic estimates of MCFA production. Previously untested microbial communities that are predicted to maximize MCFA production become candidates for future experiments.

This material is based upon work supported by the Great Lakes Bioenergy Research Center, U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Number DE-SC0018409.