

## Identifying growth harness actuator genes for *Pseudomonas putida*

Shara Balakrishnan<sup>1\*</sup> (sbalakrishnan@ucsb.edu), Rob Egbert<sup>2</sup> and Enoch Yeung<sup>3</sup>

<sup>1</sup>Dept. of Electrical and Computer Eng, University of California Santa Barbara, Santa Barbara, California;

<sup>2</sup>Earth & Biological Sciences Directorate, Pacific Northwest National Laboratory, Richland, Washington;

<sup>3</sup>Dept. of Mechanical Eng, University of California Santa Barbara, Santa Barbara, California

**Website:** <https://genomicscience.energy.gov/research/sfas/pnnlbiosystemsdesign.shtml>

*Project Goals: The Pacific Northwest National Laboratory Persistence Control Scientific Focus Area is focused on identifying controllable variables that govern the persistence of engineered microbial functions in the rhizosphere. Phenotype controllability is a new concept, integrating advances in machine learning and control theory with the latest measurement and biotechnological advances in synthetic and systems biology. With rigorous notions of phenotype control established, we elucidate novel design principles to control the environmental niche of candidate rhizosphere microbes. These principles will lead to secure plant-microbe biosystems that promote secure, stress-tolerant, and highly productive biomass crops.*

Advancing a new genetically modified microbial species from the laboratory to the field is a major obstacle in synthetic biology. When a synthetic microbe is introduced into a new environment, it interacts with native microbial fauna, subjecting itself to competitors and predators that constrain proliferation [1]. To control its persistence in the alien environment, we aim to design growth harness actuators that can potentially control the expression of genes to manipulate the microbial growth. A critical challenge in the design process is to identify the “fitness genes” which are the key genes that play a role in the growth phenotype.

Existing approaches primarily center on analysis of differential RNA-sequencing (dRNA-seq) data; comparative models contrast different phenotypes with different transcriptome profiles to create a genotype-to-phenotype map. To identify the fitness genes, dRNA-seq can be used by comparing the RNA expression levels of samples in the growth phase and the stationary phase. Random barcoded transposon sequencing (RB-TnSeq) is an approach that is tailored specifically to identify fitness genes. The fitness genes are identified by creating genetic variants of the base strain by deactivating a single gene in each variant and measuring the fitness values of the variant. The more negative the fitness value, the more important the gene. The fundamental premise of RB-TnSeq methods is that the microbe of interest is genetically tractable and can be engineered with high enough transformation efficiency to make large, pooled libraries of knockdown variants. Thus, for emerging microbial species in the rhizosphere with undefined genetics, there is a need for data-driven approaches.

We propose a novel data-driven technique that integrates dynamical systems theory and machine learning to discover genotype-to-phenotype mathematical models that map transcriptomic signatures to fitness curves. These models quantify how the growth curve changes in response to change in expression of candidate genes, in a purely data-driven fashion, which enables analysis of de

novo genome sequence and transcriptomics data. These models are formulated precisely to solve a never before explored sensor fusion problem in Koopman operator theory, which through spectral techniques yields candidate genes that strongly influence growth phenotype.

We model the proliferation process of the microbe as a state space model from dynamical systems theory by assuming the state of the system to be the RNA expression as it characterizes the genetic activity of the microbe and the output to be the growth curve as that is the phenotype of interest. The state space model comprises the state equation which captures the evolution of the genetic activity as a function of time and the output equation which maps the genetic activity to the growth curve output. We developed an algorithm called the output constrained deep dynamic mode decomposition (OC-DeepDMD) algorithm to identify a high dimensional linear Koopman model of a relatively lower dimensional nonlinear system such that the state space model becomes linear in both the state and output equation. The high dimensional state is obtained by a state inclusive nonlinear transformation identified using multilayer feedforward neural networks as done in [2].

In this work, we consider the growth of the soil bacterium *Pseudomonas putida* in R2A media with varying concentrations of two nutrients - glucose as a carbon source and casein hydrolysate as a source for amino acids. By observing the growth curves of *P. putida* under varying concentrations of the two nutrients, we select the condition under which a maximum growth rate (MAX condition) is observed and the negative control (NC) condition in which the nutrients are absent. We perform a time-series RNA sequencing experiment for the MAX and NC conditions while obtaining optical density (OD) measurements. We identified the Koopman operator representation of the state space model using OC-DeepDMD algorithm. Using modal decomposition, we club the genes into sparse modes and the dominant genes in the mode that contributes maximum to the output are the candidate genes of interest.

In summary, we proposed the OC-DeepDMD algorithm to identify state space models that serve as genotype-to-phenotype maps. Specifically, we used this algorithm to identify a set of genes for *P. putida* in R2A media that correspond to the growth phenotype. These genes form the recommendations for genetic targets for CRISPR-dCas9 knockdown to thereby control persistence of the microbe.

## References/Publications

- [1] Wang, F. and Zhang, W., 2019. Synthetic biology: recent progress, biosafety and biosecurity concerns, and possible solutions. *Journal of Biosafety and Biosecurity*, 1(1), pp.22-30
- [2] Yeung, E., Kundu, S. and Hodas, N., 2019, July. Learning deep neural network representations for Koopman operators of nonlinear dynamical systems. In *2019 American Control Conference (ACC)* (pp. 4832-4839). IEEE.

**Funding statement:** This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0012345 and by Pacific Northwest National Laboratory, grant no. 490521