Title: Combinatorial Biocontainment Design and DNA Barcode Genotyping

Authors: Ayako Murao,[1] Diána Hernández Hernández,[1] Melissa Amezola,[1] Lin Ding,[1] Jacob Sebesta,[2] Bin Yang,[2] Gabriella Li,[2] Katie Arnolds,[2] Cristal Zuniga,[3] Rodrigo Santibanez Palominos,[3] Marcus Bray,[3] Christopher Johnson,[2] Karsten Zengler,[3] Wei Xiong,[2] Jeffrey Linger,[2] Jianping Yu,[2] **Michael Guarnieri**,[2] and Yo Suzuki[1*] (ysuzuki@jcvi.org)

Institutions: [1]J. Craig Venter Institute, La Jolla, California; [2]National Renewable Energy Laboratory, Golden, Colorado; and [3]University of California San Diego, La Jolla, California

Project Goals: Synthetic biology promises a transformative bioeconomy, but much work remains in ensuring biocontainment for engineered organisms. A unique challenge for biocontainment in an industrial setting is that massive culture-scale necessitates ultra-low escape frequencies, while it is undesirable if genetic mechanisms for control interfere with organism fitness and biosynthesis capabilities. The IMAGINE team aims to solve this challenge with combinatorics and synergy, where multiple mechanisms that control viability in different ways with little adverse effect on the valuable properties of the microbes are combined to generate robust biocontainment for ensuring biosafety. The overall goal of our project is to examine the incorporation of synergy among mechanisms for biocontainment as a widely applicable strategy for establishing safe industrial organisms without sacrificing efficiencies for bioproduction.

Abstract Text: To facilitate the analysis of combinatorial constructs in our target organisms, a method termed combinatorial genetics *en masse* (CombiGEM; Wong et al., 2016) was implemented. In this method, the recursive cloning of DNA-barcoded modules via designated restriction sites within the modules results in the accumulation of constructs of interest on one side and the DNA barcodes on the other side of the restriction sites. The concatenated barcodes reveal the identity of the multiple modules in each strain. To conduct the CombiGEM process for biocontainment modules, we generated the base plasmid pCombi-CC (copy control) that can be maintained in *E. coli* as a single-copy entity to cope with any toxicity from biocontainment genes and can also be induced to increase its copy number (Lucigen). pCombi-CC enables the recursive assembly of modules using the BamHI, BglII, EcoRI, and MfeI sites. We demonstrated this in an experiment with four mock biocontainment modules. To integrate next-generation sequencing (NGS) into the analysis of concatenated DNA barcodes that mark our combinatorial constructs, we incorporated both a widely used amplicon sequencing approach (Caporaso et al., 2011) and the standard Illumina TruSeq process. The compatible pCombi-CC-Next vectors contain part of the TruSeq Index Adapter sequence, so that a single primer without any strong secondary structure can be used to introduce an index for multiplexing and complete the construction of a library to be sequenced with universal primers. With this approach, we can obtain 1 million reads for a concatenated DNA barcode library for $7 using the Illumina NovaSeq platform. With PCR optimization, artifactual reads were reduced to 0.01% of the whole. A computational analysis pipeline for rapid data analysis was also established. Toward applying the CombiGEM strategy to the analysis of biocontainment mechanisms, we completed

the initial design of the vector, regulatory sequences, and six toxin systems for our five target industrially relevant organisms, *Saccharomyces cerevisiae*, *Pseudomonas putida*, *Synechocystis* sp. PCC 6803, *Clostridium ljungdahlii*, and *Mycoplasma mycoides*. This design reflects the current organism-specific strategies the IMAGINE team is taking, and some of the molecular tools are being developed. In *P. putida*, there are effective tools for inducible gene expression. We are using the AraC-P_{BAD} system (Gauttam et al., 2021) for driving the expression of antitoxin genes. However, there is a shortage of tools for inducible repression needed for inactivating toxin genes when the organism is in a growth-permissible space. Our group is developing a system with nickel-activated Nik repressor for this purpose. Transitioning from the designed sequences to the actual DNA constructs and strains will be our next challenge. With the combinatorial approach, we expect to generate hundreds of genotypes, each corresponding to a particular combination of biocontainment modules, in each organism and use the abundances of concatenated DNA barcodes in mixed populations to elucidate the synergistic interactions among the biocontainment modules tested.

References/Publications

Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Lozupone, C.A., Turnbaugh, P.J., Fierer, N., and Knight, R. (2011). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proceedings of the National Academy of Sciences, USA *108*, 4516–4522.

Gauttam, R., Mukhopadhyay, A., Simmons, B.A., and Singer, S.W. (2021). Development of dual-inducible duet-expression vectors for tunable gene expression control and CRISPR interference-based gene repression in *Pseudomonas putida* KT2440. Microbial Biotechnology *14*, 2659–2678.

Wong, A.S.L., Choi, G.C.G., Cui, C.H., Pregernig, G., Milani, P., Adam, M., Perli, S.D., Kazer, S.W., Gaillard, A., Hermann, M., et al. (2016). Multiplexed barcoded CRISPR-Cas9 screening enabled by CombiGEM. Proceedings of the National Academy of Sciences, USA *113*, 2544–2549.

Funding Statement: This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomic Science Program under Secure Biosystems Design Science Focus Area IMAGINE BioSecurity: Integrative Modeling and Genome-scale Engineering for Biosystems Security, under contract number DE-AC36-08GO28308.