Title: Expanding the Utility of Integrases for Genome Editing and Stabilizing Gene Modules in Target Bacteria

Authors: Lauren G. Clark[2] (lauren.clark@northwestern.edu), Jesse Cahill[1], Catherine M. Mageeney[1], Grant Rybnicky[2], Michael Jewett[2], Joseph Schoeniger[1], **Kelly P. Williams[1]**

Institutions: [1] Sandia National Laboratories, Livermore, CA; [2] Northwestern University, Evanston, IL

Project Goals:

The overall goal for the Intrinsic Control for Genome and Transcriptome Editing in Communities (InCoGenTEC) project funded under the Secure Biosystems Design initiative is to expand our mechanistic and practical understanding of horizontal gene transfer mechanisms in bacterial communities, and to harness mobile elements to create and deliver constructs to transform, control and detect the genetic and biochemical state of bacteria. Improved ability to engineer genomes of both isolatable and non-isolatable species will enable better scientific understanding of bacterial communities and facilitate biotechnology applications that promote the growth of the bioeconomy. However, biocontainment and biosafety concerns must be addressed. We are creating modular synthetic genetic elements (SGEs) that can sense, and control altered states of microorganisms. These synthetic genetic elements can be used to transform community members *in situ* provided selective delivery and transformation mechanisms if delivery vectors can be created in an agile manner for new target strains. We have developed software that can precisely identify genomic islands in bacteria, which has yielded both thousands of unique DNA integrase insertion sites, and hundreds of thousands of prophages These phages can be recovered and used to deliver modular SGEs to target species within a community, even non-isolable members.

Integrases are key enzymes that can be used in phage-based or conjugative delivery systems for the site-specific insertion of delivered SGE modules into target microbe genomes, or for excising SGEs. Our goals for this project include developing a better understanding of integrase biology and building a broad set of tools that can be used for genome manipulation in a wide range of microbial species including many that are challenging to modify. We will accomplish this through the more immediate goal of conducting a high throughput screen of integrases and their att sites identified by the TIGER and Islander genome mining tools developed at Sandia National Labs. Next, a comprehensive mutational analysis of the natural attB sites will enable targeting of variant attB sites in target organism genomes in addition to providing a syntax of attB sites for the two families of integrase proteins. These goals will enable integrases, delivered via in vitro assembled phage, to be used in the larger framework of the project to edit the genomes of specific targeted non-model bacterial organisms within mixed bacterial communities. They will also inform the degree to which integrase sequence specificity requirements exclude or permit horizontal gene transfer between bacterial genomes.

Abstract Text: Integrases catalyze the splicing of large multi-gene DNAs (genomic islands) into specific sites (attBs) in chromosomes and have thus proven useful in genome editing. Expanding the set of attBs that integrases target will further increase their utility. We have applied our

software[1,2] for finding genomic islands to hundreds of thousands of bacterial genomes, while speeding the software with species-tailored databases and extending its reach to cross-contig islands. Because mapping is precise, each island links an integrase enzyme to the chromosomal site (attB) that it targets. We have thus developed a very large database of integrase/attB pairs. The main organizing principle to apply is a phylogenetic tree of the integrases, which will help us enumerate the number of truly different site-specificities in our database. The tree is challenging because it is large and a typical aid to its assessment - the bacterial tree of life - is not applicable, since integrase genes are on mobile DNAs that effectively jump from branch to branch on the bacterial tree. We have chosen an initial set of integrases to study biochemically, based on their occurrence among the study organisms of our project, and with a balance between the two main integrase protein families. Both in vitro (transcription/translation of integrase genes in cell-free extracts) and in vivo approaches are being explored to demonstrate integrase activity and application, with the goal of characterizing a large number of candidate integrase/attB pairs and conducting a comprehensive mutational analysis on a selected subset of the natural attB sites, to enable the targeting of much larger numbers of variant attB sites present in target organism genomes.

References/Publications

1. C. M. Mageeney *et al.*, New candidates for regulated gene integrity revealed through precise mapping of integrative genetic elements. *Nucleic Acids Res* **48**, 4052-4065 (2020).
2. C. M. Hudson, B. Y. Lau, K. P. Williams, Islander: a database of precisely mapped genomic islands in tRNA and tmRNA genes. *Nucleic Acids Res* **43**, D48-53 (2015).

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 under the Secure Biosystems Design Initiative. Sandia National Laboratories is a multi-mission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. SAND number: SAND2022-0188 A*