MinGenome: An *in silico* top-down approach for the synthesis of minimized genomes

Lin Wang[1*] (lqw5322@psu.edu), Costas D. Maranas[1], and **Himadri B. Pakrasi**[2]

[1]Department of Chemical Engineering, The Pennsylvania State University of State, University Park, PA 16802; [2]Department of Biology, Washington University, St. Louis, MO 63130

http://www.maranasgroup.com/

Project Goals: The goal of this project is to design a computation tool (i.e., MinGenome [1]) to support top-down genome minimization by shedding functions unnecessary for fast growth and other desirable traits. The hypothesis is that genome-minimized organisms will behave in a more predictable and controllable manner. MinGenome also aids in the identification of previously unknown essential functions whenever a new deleted stretch causes lethality.

Genome-scale metabolic models have led us to a greater understanding of cellular metabolism. However, many cellular and metabolic processes in a cell are still not well understood, placing limits on the extent to which the systems can be predictably engineered. Genome minimized strains are obtained by removing genome segments associated with genes or processes either detrimental or un-needed under bioproduction conditions. They offer advantages as production chassis [2] by reducing transcriptional cost, eliminating competing functions and limiting unwanted regulatory interactions. Existing approaches for identifying stretches of DNA to remove are largely *ad hoc* based on information on presumably dispensable regions through experimentally determined non-essential genes and comparative genomics. As more sophisticated genome editing tools (e.g., CRISPR) are becoming commonplace, the need for a computational aid that will help successively minimize genomes consistent with a set of performance criteria beyond simply growth rate is becoming more pressing.

Herein we present a versatile genome reduction algorithm MinGenome [1] (Figure 1) that implements a mixed integer linear programming (MILP) algorithm to iteratively identify the largest dispensable contiguous sequences without affecting the organism's growth or other desirable traits. Known essential genes or genes that cause significant fitness or performance loss are flagged and their deletion is thus prohibited. MinGenome also preserves needed transcription factors and promoter regions ensuring that retained genes will be properly transcribed while also avoiding the simultaneous deletion of synthetic lethal pairs. The potential benefit of removing even larger contiguous stretches of DNA if only one or two essential genes (to be re-inserted elsewhere) are within the deleted sequence is explored. We apply the algorithm to design minimized *E. coli* strain and *B. subtilis* strains and find that we are able to recapitulate the long deletions identified in previous experimental studies [3] and discover alternative combinations of deletions which have not yet been explored *in vivo*. MinGenome is a versatile computational tool to guide genome reduction. It can be accessed at Maranas group website (https://github.com/maranasgroup). Efforts are currently underway to integrate the tool within KBase.

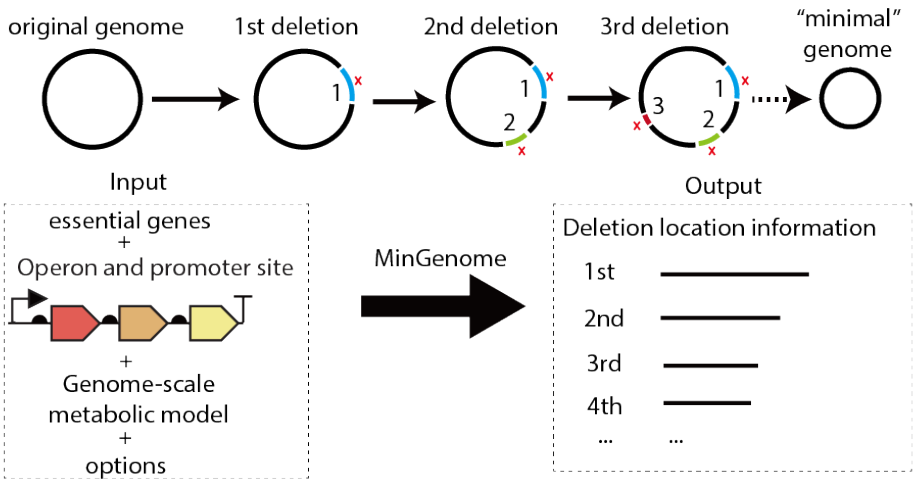

Figure 1. Schematic representation of the MinGenome algorithm. The MinGenome algorithm requires genome sequence information, gene annotation, the presence of a GSM model and information on essential genes, gene and promoter positions. It identifies the sequence of deletions starting with the largest dispensable region and proceeding monotonically to shorter ones.

References

[1] Wang, L., Maranas, C.D. (2017), MinGenome: An *in silico* top-down approach for the synthesis of minimized genomes. *ACS Synthetic Biology*.

[2] Vickers, C. E., Blank, L. M., & Krömer, J. O. (2010). Grand challenge commentary: Chassis cells for industrial biochemical production. *Nature chemical biology*, 6(12), 875-877.

[3] Juhas, M., Reuß, D. R., Zhu, B., & Commichau, F. M. (2014). *Bacillus subtilis* and *Escherichia coli* essential genes and minimal cell factories after one decade of genome engineering. *Microbiology*, 160(11), 2341-2351.

Supported by funding from the U.S. Department of Energy by grant DE-SC0012722.