

## Systems analysis of a fast growing N<sub>2</sub>-fixing cyanobacterium for production of advanced biofuels and nitrogen-containing petrochemical replacement compounds

Anindita Banerjee<sup>1\*</sup> (anindita@wustl.edu), Zi Ye,<sup>1</sup> John I. Hendry<sup>2</sup>, Costas D. Maranas<sup>2</sup>, Maciek R. Antoniewicz<sup>3</sup> and Himadri B. Pakrasi<sup>1</sup>

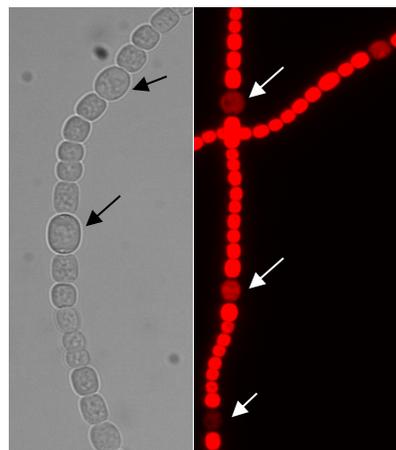
<sup>1</sup>Washington University in St. Louis, MO; <sup>2</sup>Pennsylvania State University, University Park, PA, <sup>3</sup>University of Delaware, Newark, DE

[https://sites.wustl.edu/photosynthbio/anabaena\\_33047/](https://sites.wustl.edu/photosynthbio/anabaena_33047/)

### Project Goals:

The overall objective of this project is to use an integrated systems biology approach to develop the filamentous cyanobacterium *Anabaena* sp. PCC 33047 as a model fast-growing, photosynthetic, diazotrophic production platform. The specific goals for this project are: 1) Construct a genome-scale metabolic model and predict genetic alterations that optimally direct fixed CO<sub>2</sub> and N<sub>2</sub> into target products. 2) Apply <sup>13</sup>C and <sup>15</sup>N assisted metabolomics and metabolic flux analysis to dissect the metabolism of the strain. 3) Develop an efficient genetic toolkit. 4) Demonstrate production of caprolactam and valerolactam in engineered *Anabaena* 33047. 5) Establish a stable consortium between *Anabaena* 33047 and a heterotroph for cost-effective bioproduction.

Cyanobacteria are oxygenic photosynthetic prokaryotes that rely solely on atmospheric CO<sub>2</sub> and sunlight for growth and biomass generation. These microbes are gaining importance as platforms for bioproduction. However, the commonly used cyanobacterial production strains are relatively slow growing and high-light intolerant, prohibiting high product yields. To meet the emerging need for high-light tolerant strains with high photosynthetic CO<sub>2</sub> fixation rates, we are developing a fast-growing, diazotrophic cyanobacterium, *Anabaena* 33047, as a model production platform. *Anabaena* 33047 displays a doubling time of 3.8 hours under photoautotrophic and N<sub>2</sub>-fixing condition, utilizes high light (3000 μE m<sup>-2</sup> s<sup>-1</sup>) to convert CO<sub>2</sub> into biomass at a rate that is the highest known so far among oxygenic photosynthetic organisms (3.0 g L<sup>-1</sup> day<sup>-1</sup>)<sup>1</sup>. However, this strain is not well studied and tools for its genetic modifications are lacking. In the current phase of our project we are focusing on developing a genome-scale metabolic model (GSM) for *Anabaena* 33047 and collecting physiological data that will be used to constrain the model. We are also working on developing a tool kit for the genetic transformation of this strain.



Images showing a high frequency of heterocysts (arrows) in the *Anabaena* 33047 filaments. The strain achieves a 2-fold higher heterocyst frequency compared to the model *Anabaena* 7120

Our GSM construction strategy seeks to consolidate the biochemical information from different pathway databases and available genome scale models to obtain a comprehensive description of the metabolism of *Anabaena* sp. ATCC 33047. We used the genome sequence of *Anabaena* sp. ATCC 33047<sup>2</sup> to construct draft genome scale reconstructions from three pathway databases: KEGG, MetaCyc and ModelSEED. These three draft reconstructions contained 1712,

1024 and 1521 reactions and accounted for 683, 1229 and 859 genes respectively. Additionally, based on sequence homology we constructed another draft reconstruction from the recently published model for *Anabaena* sp. PCC 7120<sup>3</sup>. This fourth draft network contained 964 reactions and accounted for 663 genes. Together, these four draft networks represent predicted reaction annotations for 1544 genes and the process of merging these four draft networks into a single draft network is underway. The synthesized draft network will form the basis for developing the final curated model.

We are characterizing the growth physiology of *Anabaena* 33047 and generating data that will be used to constrain genome-scale mathematical models of *Anabaena* metabolism. Specifically, we are measuring the dynamics of biomass composition during cell culture, which includes measuring macromolecular composition of cells (i.e. percentages of proteins, carbohydrates, RNA, lipids in cell dry weight), as well as quantifying nutrient uptake rates and product secretion rates. These data will then be combined with <sup>13</sup>C-tracing studies to generate a detailed baseline map of metabolism of *Anabaena* 33047.

As part of an effort to build a genetic toolkit for *Anabaena* 33047, we have developed a conjugation system for gene transfer into the strain. We mined the genome sequence of *Anabaena* 33047 for methylases of its restriction modification system. These methylases were then cloned into a helper plasmid with the anticipation that their expression would protect the exogenous DNA from degradation by the host. Multiple gene deletion constructs for *Anabaena* 33047 were generated and used to successfully conjugate the strain. Surprisingly, unlike other filamentous cyanobacterial strains where single homologous recombination is known to be prevalent, all the colonies tested in *Anabaena* 33047 for the gene deletions were double recombinant gene replacement mutants. We are also developing systems for gene integration and CRISPR/Cpf1 mediated genome editing.

## References

1. Gonzalez Lopez CV, Acien Fernandez FG, Fernandez Sevilla JM, Sanchez Fernandez JF, Ceron Garcia MC, Molina Grima E. Utilization of the cyanobacteria *Anabaena* sp. ATCC 33047 in CO<sub>2</sub> removal processes. *Bioresource technology*. 2009;100(23):5904-10.
2. Pfeffer S, Brown RM. Complete Genome Sequence of the Cyanobacterium *Anabaena* sp. 33047. *Genome Announcements*. 2016;4(4):e00809-16.
3. Malatinszky D, Steuer R, Jones PR. A Comprehensively Curated Genome-Scale Two-Cell Model for the Heterocystous Cyanobacterium *Anabaena* sp. PCC 7120. *Plant Physiol*. 2017;173(1):509-23.

*Funding statement: This study is being supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomic Science Program under Award Number DE-SC0019386*