SAND2019-6082C

Sandia National Laboratories

Development of Computational and Genetic Tools for CRISPRi/a Screening in *Synechococcus* sp. PCC 7002

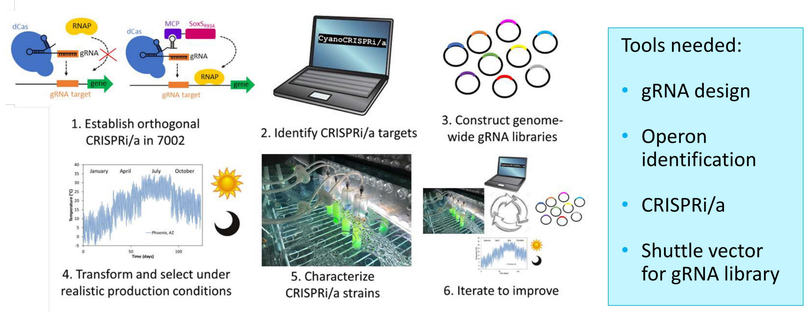
Raga Krishnakumar,[1] Anne M. Ruffing,[2] Chuck R. Smallwood,[3] Joshua Podlevsky,[2] Stephanie Kolker,[2] and Valerie R. Hines[4]

1. Systems Biology, Sandia National Laboratories, Livermore, CA; 2. Bioenergy and Defense Technologies, Sandia National Laboratories, Albuquerque, NM; 3. Nanobiology, Sandia National Laboratories, Albuquerque, NM; 4. Military Academic Collaboration Cadet Intern, United States Coast Guard

Introduction

Natural biomass production rates limit the economic viability of algal biofuel systems. Rational strain development efforts targeting single gene or pathway manipulations in microalgae have led to incremental improvements in laboratory biomass production, yet they have failed to translate to significant improvement in outdoor testing. These deficiencies result from a failure to optimize metabolism at the systems-level and to effectively mimic outdoor production environments during strain development. We propose to develop high-throughput screening of CRISPR interference and CRISPR activation (CRISPRi/a) in a cyanobacterium to identify genetic modifications that lead to enhanced growth phenotypes under realistic production conditions. *Synechococcus* sp. PCC 7002 (subsequently abbreviated 7002) is our cyanobacterial platform due to its natural properties of high salt tolerance, high light tolerance, temperature tolerance, and fast growth rate. We will leverage the CRISPRi system already developed for 7002[1] and develop a modified system for CRISPRa. Through guide RNA (gRNA) library screening for down- and up-regulation of every gene in 7002, we will identify genetic modifications that improve growth of 7002 under simulated outdoor light and temperature conditions.

Project Overview



Tools needed:

- gRNA design
- Operon identification
- CRISPRi/a
- Shuttle vector for gRNA library

Figure 1. Steps for high-throughput gRNA screening of CRISPRi/a targets in 7002 for improved biomass productivity.

gRNA Design

- Protospacer adjacent motif (PAM) sites for 10 different Cas proteins (Table 1) were identified in the 7002 genome and it's 6 native plasmids.
- Spy and Sau2 have the highest number of PAMs per kb and therefore the highest gRNA densities (Figure 2).
- PAM densities in the 7002 plasmids are similar to that of the chromosome.

Table 1. List of common Cas proteins and associated protospacer adjacent motif (PAM) sites

Organism	Nickname	PAM
Streptococcus pyogenes	Spy	NGG
Staphylococcus aureus - 1	Sau1	NGRRT
Staphylococcus aureus - 2	Sau2	NGRRN
Neisseria meningitidis	Nme	NNNNGAAT
Campylobacter jejuni	Cje	NNNNRYAC
Streptococcus thermophilus	Sth	NNAGAAW
Treponema denticola	Tde	NAAAAC
Prevotella/Francisella	Cpf1	TTTV
Prevotella/Francisella	Cpf1_m1	TYCV
Prevotella/Francisella	Cpf1_m2	TATV

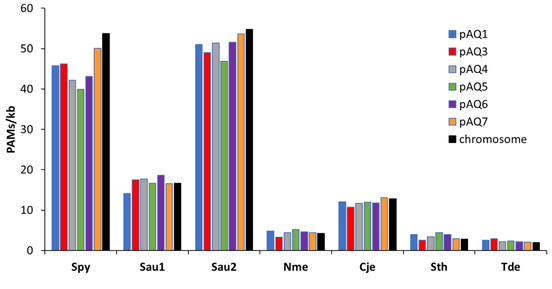


Figure 2. PAM site density per kilobase (kb) for the chromosome and 6 native plasmids of 7002.

gRNA Design (cont.)

gRNA prediction pipeline:

- Restrict gRNA sequences to be -300bp to 300bp relative to the transcription start site of the gene target
- Use Benchling to find gRNA sequences based on the PAM site for each Cas (standard search algorithm)
- Calculate 'blast penalty' based on mismatches to the genome (Figure 3)
  - Blue BLAST penalties indicate a low potential for off-target effects
  - Yellow BLAST penalties have a high potential for off-target effects
- Keep both template and coding strands

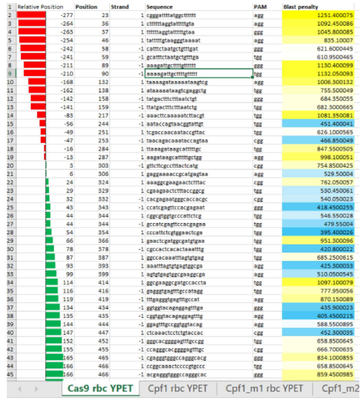


Figure 3. Results from gRNA prediction pipeline.

Operon Identification in 7002

- Operon annotation in 7002 is needed to design gRNAs for CRISPRi/a and to determine potential pleiotropic effects.
- ProOpDB:[2] STRING score (physical and functional protein interactions), orthologs (Clusters of Orthologous Groups [COG] database), intergenic distance
- OperonDB:[3] orthologs (gene pair conservation) – de novo BLAST identification, intergenic distance < 200bp
- RNA-seq based approach: For two adjacent genes in the same direction: Is the expression of gene n+1 within *two standard deviations* of gene n? –YES = OPERON

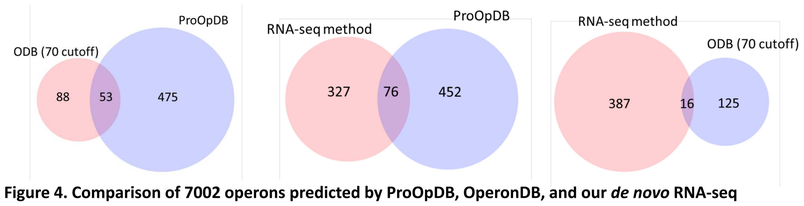


Figure 4. Comparison of 7002 operons predicted by ProOpDB, OperonDB, and our *de novo* RNA-seq based approach. RNA-seq dataset = Gordon *et al.*[5]

Development of CRISPRi/a Tools in 7002

- Spy dCas9 CRISPRi tools from Gordon *et al.* 2016 obtained from Addgene, changed the antibiotic resistance cassette, and genome-integrated into 7002 and two Ypet-expressing strains.
- Inhibition of *cpcB* (phycocyanin production) and *ypet* was confirmed with genome integration of gRNAs targeting these genes (Figure 5), but there was leaky expression from the *tet* promoter.
- Development of CRISPRa system based on 2-hybrid CRISPR-dCas system established in *E. coli* [4] is ongoing (Figure 6). This system uses a single dCas and was shown to have higher activation compared to direct fusion of the transcriptional activator to dCas9.

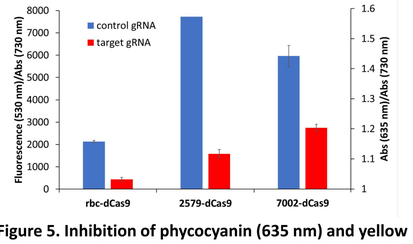


Figure 5. Inhibition of phycocyanin (635 nm) and yellow fluorescence (530 nm) with gRNAs targeting *cpcB* and *ypet* and expression of dCas9 from *tet* promoters, 48h after addition of anhydrotetracycline.

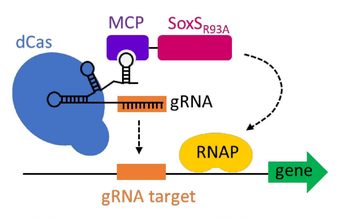


Figure 6. Schematic of 2-hybrid CRISPRa system.

Development of *E. coli* – 7002 Shuttle Vector

- pAQ1 is the smallest native plasmid in 7002.
- 8 overlapping fragments from pAQ1 were transferred into pBISNL2 to test as *E. coli*-7002 shuttle vectors – pBIAQ1 (Figure 7), along with the predicted ori.[5]
- 6 of the 8 putative shuttle vectors produced antibiotic resistant colonies in 7002 after transformation (Table 2).
- Plasmid DNA (pBIAQ1) was isolated from the 7002 transformants and transformed back into *E. coli* to confirm plasmid maintenance in 7002.
  - Only fragment 2 had more than 1 *E. coli* transformant (Table 2).
  - The re-transformed plasmid is similar in size to the original vector (Figure 8).

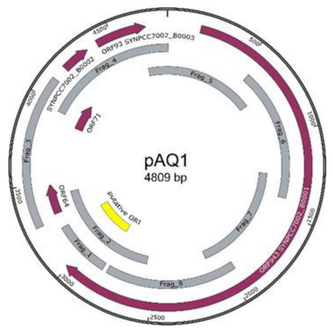


Figure 7. Map for pAQ1 and fragments tested for the origin of replication.

Table 2. Transformants from 7002 shuttle vectors.

Shuttle Vector	# 7002 transformants	# *E. coli* transformants
pBIAQ1-1	0	-
pBIAQ1-2	7	134-300+
pBIAQ1-3	8	3-7
pBIAQ1-4	4	0
pBIAQ1-5	11	0-1
pBIAQ1-6	-	-
pBIAQ1-7	3	0-1
pBIAQ1-8	5	0-1

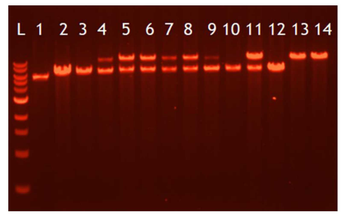


Figure 8. DNA gel electrophoresis of retransformed linearized plasmids. Controls: 1 = pBISNL2, 2 = pBIAQ1-2, and re-transformed shuttle vectors: 3-9 = pBIAQ1-2, 10-12 = pBIAQ1-3, 13 = pBIAQ1-7, and 14 = pBIAQ1-8.

Conclusions and Future Work

- Spy Cas9 will be used for CRISPRi/a due to its high gRNA density in the 7002 genome.
- A gRNA design pipeline was developed for 7002 to account for potential off-target effects.
- Long-read RNA-seq data is needed to improve operon prediction in 7002.
- CRISPRi was validated in 7002, and the 2-hybrid CRISPRa tool is under construction.
- The 7002 origin of replication was identified for pAQ1 and will be used to construct an *E. coli* – 7002 shuttle vector for gRNA library expression. Essential ORFs from pAQ1 will be genome integrated, and pAQ1 will be cured from 7002.
- After tool development is complete, the gRNA library will be designed, constructed, and tested in 7002 to identify CRISPRi/a mutants with improved growth under simulated outdoor conditions.

References and Funding

1. Gordon GC, Korosh TC, Cameron JC, Markley AL, Begemann MB, and Pfleger BF, *CRISPR interference as a titratable, trans-acting regulatory tool for metabolic engineering in the cyanobacterium Synechococcus sp. strain PCC 7002.* Metabolic engineering, 2016. 38: p. 170-179.
2. Taboada B, Ciria R, Martinez-Guerrero CE, Merino E. ProOpDB: Prokaryotic Operon DataBase. *Nucleic Acids Res*. 2011;40(Database issue):D627–D631. doi:10.1093/nar/gkr1020
3. Pertea M, Ayanbule K, Smedinghoff M, Salzberg SL. OperonDB: a comprehensive database of predicted operons in microbial genomes. *Nucleic Acids Res*. 2008;37(Database issue):D479–D482. doi:10.1093/nar/gkn784
4. Dong, et al. *Synthetic CRISPR-Cas gene activators for transcriptional reprogramming in bacteria*. Nature Communications, 2018. 9:2489.
5. Gordon GC, Cameron JC, and Pfleger BF. Global RNA half lives in cyanobacteria. *Gene Expression Omnibus.* GSE109174. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE109174

This work is funded by Laboratory Directed Research and Development (LDRD) at Sandia National Laboratories.