Title

Comprehensive Comparative Genomics to Study Complex Phenotypes in Cyanobacteria

Author

Adomako, Marie A

Publication Date

2021

UNIVERSITY OF CALIFORNIA SAN DIEGO

Comprehensive Comparative Genomics to Study Complex Phenotypes in Cyanobacteria

A dissertation submitted in partial satisfaction of the requirements for
the degree Doctor of Philosophy

in

Biology

by

Marie A. Adomako

Committee in charge:

Professor Susan S. Golden, Chair
Professor Eric E. Allen
Professor Paul R. Jensen
Professor Scott A. Rifkin
Professor Gurol Suel

2021

The dissertation of Marie A. Adomako is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2021

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SUPPLEMENTAL FIGURES

LIST OF SUPPLEMENTAL TABLES

ACKNOWLEDGEMENTS

My experience as a PhD student has been shaped entirely by my personal identity as a mother. I entered this program with many insecurities and doubts about taking on a huge intellectual endeavor that most individuals embark on when they are relatively young and unencumbered. I have heard hundreds of versions of the sentiment "How do you do this while being a mother? I can barely do this and take care of myself!" and every time has been an opportunity to reflect and be grateful for the grounding and lessons and support I have had *because* I took on this challenge as a mother. Andrew and Michael, you have kept me going even while you made me crazy.

I want to thank my husband, Rees, for being an unflagging supporter and sure of me, especially when I wasn't sure of myself. Honestly, it was quite infuriating sometimes to be confronted by his steadfast confidence in my capabilities and have no choice but to continue on, rather than giving up. I thought I would fail many times. He never considered the possibility.

I want to thank my parents for their gentle, unflappable confidence in me, their willingness to listen to me worry at any time, and their roles as head counselors at Gramma and Grampa's Summer Camp for Adomako Boys every single summer.

I want to thank the incredible friends I made at UCSD- I never expected to be blessed with such a generous, kind, strange, and freakishly smart friend group: the kind of friend that gives you their car keys in an emergency, no questions asked, after knowing you for all of a week, the kind of friend that offers to catch the last of the season's flower beetles so you can take them home to delight your 2-year-old, the kind of friend who gives you a place to stay when the world seems like it is ending.

I want to thank the Golden Lab- all members past and present- for being such fantastic friends and colleagues. It has been such a gift to do research with you all, and to become your friend. I especially want to thank Ryan Simkovsky, my mentor and office mate, who I'm sure will not miss the dulcet tones of "Hey..Ryan?" that inevitably preclude a 30 minute interlude into the mysteries of the R programming language.

And finally, I want to thank my advisor Susan Golden. There is so much to say, and no way to truly explain how grateful I am for your support. Simply put, I don't think I could have done this without you. At every bump (or sinkhole) in the road you have never done anything less than let me know that you support and value me, personally. So many times, I would come to you with failure, scientific or personal, and expect you to be as disappointed in me as I myself was. And yet, you were never disappointed in me. I have left every conversation with you with the confidence that I could do better, be better, and achieve my goals. Thank you for always believing in me, especially when I couldn't believe in myself.

Chapter 2, in full, is a reprint of the published manuscript: Yang Y, Lam V, Adomako M, Simkovsky R, Jakob A, Rockwell NC, Cohen SE, Taton A, Wang J, Lagarias JC, Wilde A, Nobles DR, Brand JJ, Golden SS. 2018. Phototaxis in a wild isolate of the cyanobacterium Synechococcus elongatus. *Proc Natl Acad Sci* 115(52). The dissertation author was the tertiary author of this paper.

Chapter 3, in full, is in preparation for publication: Adomako M, Ernst D, Simkovsky R, Chao Y-Y, Wang J, Fang M, Bouchier C, Lopez-Igual R, Mazel D, Gugger M, Golden SS. Comparative genomics of *Synechococcus elongatus* explains the phenotypic diversity of the strains. (*In preparation*). The dissertation author is the primary investigator and author of this paper.

Chapter 4 is coauthored with Emily Pierce and Susan S. Golden. The dissertation author is the primary investigator and author of this chapter.

Chapter 5 is coauthored with Elliot Weiss, Ryan Simkovsky, and Susan S. Golden. The dissertation author is the primary investigator and author of this chapter.

VITA

EDUCATION

University of California San Diego, San Diego, CA **2021**
PhD, Biology
Idaho State University, Pocatello, ID **2010**
MS, Microbiology
Idaho State University, Pocatello, ID **2008**
BS, Microbiology

PUBLICATIONS

Adomako M, Ernst D, Simkovsky R, Chao Y-Y, Wang J, Fang M, Bouchier C, Lopez-Igual R, Mazel D, Gugger M, Golden SS. Comparative genomics of *Synechococcus elongatus* explains the phenotypic diversity of the strains. (*In preparation*).

Yang Y, Lam V, **Adomako M**, Simkovsky R, Jakob A, Rockwell NC, et al. (2018) Phototaxis in a wild isolate of the cyanobacterium Synechococcus elongatus. *Proc Natl Acad Sci U S A.* 2018;115: E12378–E12387.

Adomako M, St-Hilaire S, Zheng Y, Eley J, Marcum RD, Sealey, W, Donahower BC, Lapatra S, Sheridan PP. (2012) Oral DNA vaccination of rainbow trout, *Oncorhynchus mykiss* (Walbaum), against infectious haematopoietic necrosis virus using PLGA [Poly(D,L-Lactic-Co-Glycolic Acid)] nanoparticles. *Journal of Fish Diseases* 35(3), 203-14.

Hillis AM, Prange ST, **Adomako M**, Christensen LM, St-Hilaire S, Sheridan PP. (2015) Characterization of the bacterial microflora on the skin of boreal toads, *Anaxyrus (Bufo) boreas boreas,* and Columbia spotted frogs, *Rana luteiventris*, in Grand Teton National Park, Wyoming USA. *International Journal of Microbiology Research*, 7(1), 588-97.

GRANTS

Gilliam Fellowship for Advanced Study 2017-2020
Howard Hughes Medical Institute
Cell and Molecular Genetics Training Program 2016-2018
University of California, San Diego

AWARDS AND PRESENTATIONS

- Session Chair; 13th Workshop on Cyanobacteria, Boulder, CO 2019
- Oral presentation; 13th Workshop on Cyanobacteria, Boulder, CO 2019
- Oral presentation; 2018 Annual Gilliam Fellows Meeting, Ashburn, VA 2018
- Advanced Bacterial Genetics Course, Cold Spring Harbor, NY 2018
- Invited seminar; California State Los Angeles 2018
- Finalist, GradSLAM UCSD 2018

TEACHING

California State San Marcos	**San Marcos, CA**
Guest seminar, Microbial Genomics	September 2020
University of California San Diego	**La Jolla, CA**
Instructor, BILD 1 (The Cell- Introductory Molecular and Cell Biology)	Summer I 2019
Instructional Assistant, BIMM 116 (Biological Clocks)	Fall 2018
Instructional Assistant, BIMM 120 (Microbiology)	Winter 2018
Instructional Assistant, BIMM 121 (Laboratory in Microbiology)	Fall 2016
California State Los Angeles	**Los Angeles, CA**
Guest lecture on Biofilms, MICR 3500 (Microbial Physiology)	May 3, 2018

BROADER IMPACTS

UCSD GradHAC Events Officer	2018
Young Chefs After School Club	2018
UCSD Biological Sciences Ph.D. Admissions Committee	2017/2018

ABSTRACT OF THE DISSERTATION

Comprehensive Comparative Genomics to Study Complex Phenotypes in Cyanobacteria

by

Marie A. Adomako

Doctor of Philosophy in Biology

University of California San Diego, 2021

Professor Susan S. Golden, Chair

Cyanobacterial biofilms are not only important as critical components of ecological habitats, but have applications in wastewater treatment systems, bioremediation efforts, and prevention of biofouling. Strains of the freshwater cyanobacterium *Synechococcus elongatus* were first isolated approximately 60 years ago, and PCC 7942 is well established as a model for photosynthesis, circadian biology, and biotechnology research. PCC 7942 is planktonic in lab conditions, but studies of biofilming mutants support a model of constitutive repression of biofilms in PCC 7942. A recent environmental isolate of *S. elongatus*, UTEX 3055, shares 98.46% average nucleotide identity with PCC 7942, but has unique phenotypes of phototaxis and robust biofilm formation in laboratory conditions. This genetic similarity and the constitutive repression of biofilm formation suggests that lab strains of *S. elongatus* may be domesticated. An approach combining comparative genomics analysis with the use of random barcoded transposon sequencing (RB-TnSeq) library screens was used to find the genetic basis of biofilm and phototaxis phenotypes in *S. elongatus*. This work describes the sequencing and annotation of UTEX 3055 as well as the characterization of a novel phototaxis operon in the strain; a comprehensive genome comparison of *S. elongatus* strains that provides a pangenome annotation, a corrected sequence for the type strain PCC 6301, and identification of genes controlling pigmentation and phototaxis phenotypes; the creation of an RB-TnSeq library in UTEX 3055 that can be used in future fitness screening experiments; and an IRB-Seq experiment in PCC 7942 that provides genetic targets to expand the current model of biofilm formation in *S. elongatus.*

CHAPTER 1: Introduction

Biofilms are community structures enclosed in a matrix that are produced by bacterial consortia. The molecular basis of formation, persistence, and dispersal of biofilms has been extensively studied in single-species biofilms of heterotrophic bacteria because of their impact on human health (1, 2). A biofilm begins with the active or passive migration of cells to a surface (3–5), attachment with cellular appendages (3, 6) or adhesive proteins (7–9), and the formation of microcolonies. As these communities grow and mature, the cells are enveloped by an extracellular matrix that is composed of proteins, polysaccharides and extracellular DNA (10, 11). This matrix protects the cells from environmental stresses such as predators (12), antibiotics (13), and even UV light (14). In addition, the matrix can trap nutrients and communication signals (11). Dispersal of cells from the biofilm can be induced by diverse environmental cues and using a variety of mechanisms (1). Although there is an accepted model of a biofilm "life cycle", multiple mechanisms regulate each stage of the cycle, even within the same species (15), and there is no consensus mechanism for biofilm formation.

While heterotrophic biofilms have been extensively studied, there has been much less research focused on phototrophic biofilms (16). Cyanobacteria are important photosynthetic members of biofilm communities in diverse environments (17–19), but can pose human health risks through the release of toxins from harmful algal blooms (20). Cyanobacterial biofilms have potential applications in wastewater purification (21), bioremediation efforts (22), and in combating biofouling (23, 24). The majority of research in cyanobacterial biofilms focuses on describing their roles in environmental nutrient cycling (25) and biofouling (19), but much less research has been performed that describes the mechanisms or genetics cyanobacterial biofilms at the level of detail available for heterotrophic species. Mechanistic research in biofilm formation and regulation in cyanobacteria has found some parallels with heterotrophic biofilms, such as the key signaling molecule cyclic di-GMP that controls the transition between the biofilm state and a free-living motile state in heterotrophic bacteria also controls biofilm formation in the cyanobacterium *Synechocystis* PCC 6803 (26). In PCC 6803 exopolysaccharides contribute to cell buoyancy and protection against environmental stresses, but their role in biofilm formation is less clear (27).

Synechococcus elongatus PCC 7942 is a well-studied member of the cyanobacteria and has provided a foundation for research in photosynthesis and circadian rhythms in prokaryotes (28–30). Additionally, PCC 7942 is naturally transformable and has robust homologous recombination machinery (31, 32). This genetic tractability has made it an attractive and viable production platform for biofuels and other high-value chemicals (33). Under laboratory culturing conditions PCC 7942 exhibits a persistent planktonic phenotype, even in the absence of agitation or bubbling, with no evidence of biofilm formation on the culture vessel. Schatz et al. subsequently identified and characterized a biofilming mutant of PCC 7942, T2SEΩ, in which the PilB homolog of the Type II secretion system/Type IV pili is inactivated (34). Studies using conditioned media show that the wild-type (WT) strain secretes an unknown repressor of biofilm formation that is not secreted from the T2SEΩ mutant, thus allowing biofilms to form. Additionally, small proteins with a double-glycine motif and their corresponding cysteine peptidase transporter support biofilm formation (35). The transcription of these genes is typically repressed in WT cells by the secretion of inhibitor(s), but is significantly upregulated in the secretion mutant T2SEΩ. The authors, including members of our lab, concluded that *S. elongatus* naturally forms biofilms in a regulated manner, and predicted that an environmental isolate would exhibit biofilm formation. In collaboration with Jerry Brand's lab at UT Austin, our lab obtained a novel strain of *S. elongatus*, UTEX 3055, from environmental sampling of Waller Creek in Austin, TX. In support of the hypothesis that domestication has abolished a number of phenotypes, this isolate is interesting because it is motile, displays phototaxis, and forms robust biofilms, all of which are phenotypes the lab adapted PCC 7942 strain does not display (36). I hypothesized that a comparative genomics analysis of PCC 7942 and UTEX 3055 could elucidate the genetic basis of the biofilm and phototaxis phenotype in *S. elongatus* because the high genetic similarity between the two strains narrows the focus for which genes and nucleotide differences between the two strains may contribute to these phenotypes.

I undertook a comprehensive genome comparison among UTEX 3055 and the previously characterized *S. elongatus* isolates PCC 6301, PCC 6311, PCC 7942, PCC 7943, and UTEX 2973 ("legacy strains") to test the hypothesis that laboratory strains may have become domesticated, and that differences between the strains could be used to find genes related to biofilm formation and phototaxis. I used the isolation history of *S. elongatus* strains as the context for understanding the connection between their phenotypes and genotypes. The legacy strains of *S. elongatus* comprise the earliest isolations from freshwater sources in Texas, including the type strain PCC 6301 (37), and strains later isolated from freshwater near San Francisco, California, including PCC 7942 (38). The strains characterized most recently include UTEX 2973, a recent re-isolation from a frozen archive of PCC 6301 (39) and UTEX 3055, isolated from Waller Creek, Texas about 60 years after PCC 6301 was sampled from the same source (36). The first results of this analysis are sequence and annotation refinement through a re-sequencing of the type strain PCC 6301 and sequencing of PCC 6311 and PCC 7943, as well as the creation of a curated pangenome annotation for all *S. elongatus* strains. Examination of the genome differences at different scales revealed large genome regions that control pigmentation phenotypes, a putative operon of UTEX 3055 necessary for phototaxis, and patterns of SNPs in legacy strains that led to a re-evaluation of the relationships among strains, and an explanation of a perplexing SNP in a core circadian clock component, rpaA, that has previously caused confusion in the literature.

In addition to comparative genomics, a randomly barcoded transposon insertion sequencing (RB-TnSeq) library in *S. elongatus* PCC 7942 is a powerful tool that can be used to connect phenotypes and genotypes. In an RB-TnSeq library, a transposon that carries an antibiotic-resistance cassette paired with a unique barcode is inserted randomly into the genome, potentially disrupting gene functions. A fully saturated and pooled library will have representative mutants for each non-essential gene or DNA region. Once next generation sequencing (NGS) identifies the location of each insertion and its associated barcode, the relative fitness of each mutant in the pooled library can be ascertained for any experimental condition relative to a control condition by quantifying the barcodes through NGS. RB-TnSeq enables large-scale, cost-effective mutant fitness screening (40). RB-TnSeq in PCC 7942 has been used to determine the essential gene set of the organism (41), screen for genes that affect survival in tens of conditions (42), and has improved the construction of a genome-scale metabolic model of PCC 7942 (43).

The constitutive repression of biofilm formation in PCC 7942 complicates the interpretation of RB-TnSeq library biofilm screening experiments. I instead took two alternative approaches that leverage the comparative analysis of UTEX 3055 and PCC 7942 as well as the known pathway of biofilm regulation in PCC 7942. In the first approach, I created a complementary RB-TnSeq library in UTEX 3055 and compared the essential gene set analysis of this library with that of the PCC 7942 library. This UTEX 3055 library will be used in future screening experiments to find genes related to biofilm formation and phototaxis. In the second approach, I used interaction RB-TnSeq (IRB-Seq), a method that can determine the fitness effect of synthetic genetic interactions in specific environmental conditions. In IRB-Seq, a known mutation is introduced to the pooled library via transformation, and the resulting interaction library is subjected to an environmental perturbation.

In this work, I have used comparative genomics and RB-TnSeq screens to examine the genetic differences between UTEX 3055 and PCC 7942 and find genes responsible for biofilm formation and phototaxis in *S. elongatus*. In Chapter 2, my coauthors and I characterize UTEX 3055 as a strain of *S. elongatus* that forms robust biofilms in laboratory conditions and contains a unique photoreceptor that controls both negative and positive phototaxis. In Chapter 3, I describe how the comparative analysis of UTEX 3055 with legacy strains of *S. elongatus* has led to refined genome data for all *S. elongatus* strains and the discovery of genes responsible for pigmentation and phototaxis phenotypes. In Chapter 4, I describe the creation of an RB-TnSeq library in UTEX 3055 and how it compares to the library of PCC 7942. In Chapter 5, I use an IRB-Seq approach to expand the known model of biofilm regulation in PCC 7942. This work illustrates how the comparison of two closely related strains with differing phenotypes can be leveraged to examine complex phenotypes.

References

1. McDougald D, Rice SA, Barraud N, Steinberg PD, Kjelleberg S. 2011. Should we stay or should we go: mechanisms and ecological consequences for biofilm dispersal. Nat Rev Microbiol 10:39–50.

2. Stanley NR, Lazazzera BA. 2004. Environmental signals and regulatory pathways that influence biofilm formation. Mol Microbiol 52:917–924.

3. Pratt LA, Kolter R. 1998. Genetic analysis *of Escherichia coli* biofilm formation: roles of flagella, motility, chemotaxis and type I pili. Mol Microbiol 30:285–293.

4. Merritt PM, Danhorn T, Fuqua C. 2007. Motility and chemotaxis in *Agrobacterium tumefaciens* surface attachment and biofilm formation. J Bacteriol 189:8005–8014.

5. Caiazza NC, O'Toole GA. 2003. Alpha-toxin is required for biofilm formation by *Staphylococcus aureus*. J Bacteriol 185:3214–3217.

6. Klausen M, Heydorn A, Ragas P, Lambertsen L, Aaes-Jørgensen A, Molin S, Tolker-Nielsen T. 2003. Biofilm formation by *Pseudomonas aeruginosa* wild type, flagella and type IV pili mutants. Mol Microbiol 48:1511–1524.

7. Cucarella C, Solano C, Valle J, Amorena B, Lasa I, Penadés JR. 2001. Bap, a *Staphylococcus aureus* surface protein involved in biofilm formation. J Bacteriol 183:2888–2896.

8. Toledo-Arana A, Valle J, Solano C, Arrizubieta MJ, Cucarella C, Lamata M, Amorena B, Leiva J, Penadés JR, Lasa I. 2001. The enterococcal surface protein, Esp, is involved in *Enterococcus faecalis* biofilm formation. Appl Environ Microbiol 67:4538–4545.

9. Shanks RMQ, Stella NA, Kalivoda EJ, Doe MR, O'Dee DM, Lathrop KL, Guo FL, Nau GJ. 2007. A *Serratia marcescens* OxyR homolog mediates surface attachment and biofilm formation. J Bacteriol 189:7262–7272.

10. Flemming H-C, Wingender J. 2010. The biofilm matrix. Nat Rev Microbiol 8:623–633.

11. Flemming H-C, Neu TR, Wozniak DJ. 2007. The EPS Matrix: The "House of Biofilm Cells." J Bacteriol 189:7945–7947.

12. DePas WH, Syed AK, Sifuentes M, Lee JS, Warshaw D, Saggar V, Csankovszki G, Boles BR, Chapman MR. 2014. Biofilm formation protects *Escherichia coli* against killing by *Caenorhabditis elegans* and *Myxococcus xanthus*. Appl Environ Microbiol 80:7079–7087.

13. Mah TF, O'Toole GA. 2001. Mechanisms of biofilm resistance to antimicrobial agents. Trends Microbiol 9:34–39.

14. Wright DJ, Smith SC, Joardar V, Scherer S, Jervis J, Warren A, Helm RF, Potts M. 2005. UV irradiation and desiccation modulate the three-dimensional extracellular matrix of *Nostoc commune*. J Biol Chem 280:40271–40281.

15. O'Toole G, Kaplan HB, Kolter R. 2000. Biofilm formation as microbial development. Annu Rev Microbiol 54:49–79.

16. Bharti A, Velmourougane K, Prasanna R. 2017. Phototrophic biofilms: diversity, ecology and applications. J Appl Phycol 29:2729–2744.

17. Cole JK, Hutchison JR, Renslow RS, Kim Y-M, Chrisler WB, Engelmann HE, Dohnalkova AC, Hu D, Metz TO, Fredrickson JK, Lindemann SR. 2014. Phototrophic biofilm assembly in microbial-mat-derived unicyanobacterial consortia: model systems for the study of autotroph-heterotroph interactions. Front Microbiol 5:109.

18. de los Ríos A, Cary C, Cowan D. 2014. The spatial structures of hypolithic communities in the Dry Valleys of East Antarctica. Polar Biol 37:1823–1833.

19. Crispim CA, Gaylarde PM, Gaylarde CC. 2003. Algal and cyanobacterial biofilms on calcareous historic buildings. Curr Microbiol 46:79–82.

20. Huisman J, Codd GA, Paerl HW, Ibelings BW, Verspagen JMH, Visser PM. 2018. Cyanobacterial blooms. Nat Rev Microbiol 16:471–483.

21. Egan S, Thomas T, Kjelleberg S. 2008. Unlocking the diversity and biotechnological potential of marine surface associated microbial communities. Curr Opin Microbiol 11:219–225.

22. Velasco Ayuso S, Giraldo Silva A, Nelson C, Barger NN, Garcia-Pichel F. 2017. Microbial nursery production of high-quality biological soil crust biomass for restoration of degraded dryland soils. Appl Environ Microbiol 83.

23. Ivnitsky H, Katz I, Minz D, Volvovic G, Shimoni E, Kesselman E, Semiat R, Dosoretz CG. 2007. Bacterial community composition and structure of biofilms developing on nanofiltration membranes applied to wastewater treatment. Water Res 41:3924–3935.

24. Gademann K. 2007. Cyanobacterial natural products for the inhibition of biofilm formation and biofouling. CHIMIA International Journal for Chemistry https://doi.org/10.2533/chimia.2007.373.

25. Arp G, Reimer A, Reitner J. 2001. Photosynthesis-induced biofilm calcification and calcium concentrations in Phanerozoic oceans. Science 292:1701–1704.

26. Agostoni M, Waters CM, Montgomery BL. 2016. Regulation of biofilm formation and cellular buoyancy through modulating intracellular cyclic di-GMP levels in engineered cyanobacteria. Biotechnol Bioeng 113:311–319.

27. Jittawuttipoka T, Planchon M, Spalla O, Benzerara K, Guyot F, Cassier-Chauvat C, Chauvat F. 2013. Multidisciplinary evidences that *Synechocystis* PCC6803 exopolysaccharides operate in cell sedimentation and protection against salt and metal stresses. PLoS One 8:e55564.

28. Kondo T, Strayer CA, Kulkarni RD, Taylor W, Ishiura M, Golden SS, Johnson CH. 1993. Circadian rhythms in prokaryotes: luciferase as a reporter of circadian gene expression in cyanobacteria. Proc Natl Acad Sci U S A 90:5672–5676.

29. Zouni A, Witt HT, Kern J, Fromme P, Krauss N, Saenger W, Orth P. 2001. Crystal structure of photosystem II from *Synechococcus elongatus* at 3.8 A resolution. Nature 409:739–743.

30. Jordan P, Fromme P, Witt HT, Klukas O, Saenger W, Krauss N. 2001. Three-dimensional structure of cyanobacterial photosystem I at 2.5 A resolution. Nature 411:909–917.

31. Golden SS, Brusslan J, Haselkorn R. 1987. [12] Genetic engineering of the cyanobacterial chromosome, p. 215–231. In Methods in Enzymology. Academic Press.

32. Taton A, Erikson C, Yang Y, Rubin BE, Rifkin SA, Golden JW, Golden SS. 2020. The circadian clock and darkness control natural competence in cyanobacteria. Nat Commun 11:1688.

33. Angermayr SA, Gorchs Rovira A, Hellingwerf KJ. 2015. Metabolic engineering of cyanobacteria for the synthesis of commodity products. Trends Biotechnol 33:352–361.

34. Schatz D, Nagar E, Sendersky E, Parnasa R, Zilberman S, Carmeli S, Mastai Y, Shimoni E, Klein E, Yeger O, Reich Z, Schwarz R. 2013. Self-suppression of biofilm formation in the cyanobacterium *Synechococcus elongatus*. Environ Microbiol 15:1786–1794.

35. Parnasa R, Nagar E, Sendersky E, Reich Z, Simkovsky R, Golden S, Schwarz R. 2016. Small secreted proteins enable biofilm development in the cyanobacterium *Synechococcus elongatus*. Sci Rep 6:32209.

36. Yang Y, Lam V, Adomako M, Simkovsky R, Jakob A, Rockwell NC, Cohen SE, Taton A, Wang J, Lagarias JC, Wilde A, Nobles DR, Brand JJ, Golden SS. 2018. Phototaxis in a wild isolate of the cyanobacterium *Synechococcus elongatus*. Proc Natl Acad Sci U S A 115:E12378–E12387.

37. Stanier RY, Kunisawa R, Mandel M, Cohen-Bazire G. 1971. Purification and properties of unicellular blue-green algae (order *Chroococcales*). Bacteriol Rev 35:171–205.

38. Golden SS. 2018. The international journeys and aliases of Synechococcus elongatus. N Z J Bot 1–6.

39. Yu J, Liberton M, Cliften PF, Head RD, Jacobs JM, Smith RD, Koppenaal DW, Brand JJ, Pakrasi HB. 2015. *Synechococcus elongatus* UTEX 2973, a fast growing cyanobacterial chassis for biosynthesis using light and CO2. Sci Rep 5:8132.

40. Wetmore KM, Price MN, Waters RJ, Lamson JS, He J, Hoover CA, Blow MJ, Bristow J, Butland G, Arkin AP, Deutschbauer A. 2015. Rapid quantification of mutant fitness in diverse bacteria by sequencing randomly bar-coded transposons. MBio 6:e00306–15.

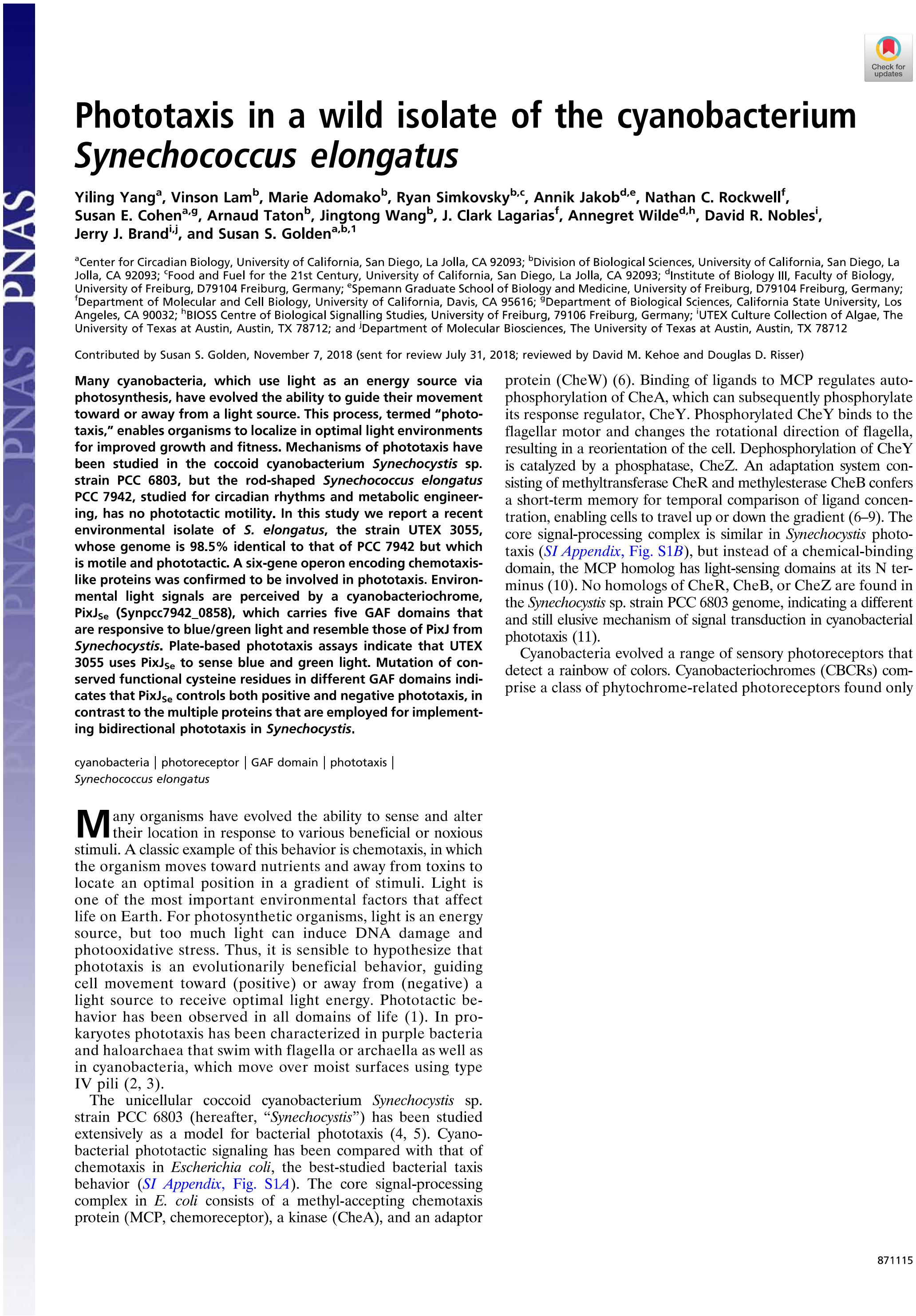
41. Rubin BE, Wetmore KM, Price MN, Diamond S, Shultzaberger RK, Lowe LC, Curtin G, Arkin AP, Deutschbauer A, Golden SS. 2015. The essential gene set of a photosynthetic organism. Proc Natl Acad Sci U S A 112:E6634–43.

42. Price MN, Wetmore KM, Waters RJ, Callaghan M, Ray J, Liu H, Kuehl JV, Melnyk RA, Lamson JS, Suh Y, Carlson HK, Esquivel Z, Sadeeshkumar H, Chakraborty R, Zane GM, Rubin BE, Wall JD, Visel A, Bristow J, Blow MJ, Arkin AP, Deutschbauer AM. 2018. Mutant phenotypes for thousands of bacterial genes of unknown function. Nature 557:503–509.

43. Broddrick JT, Rubin BE, Welkie DG, Du N, Mih N, Diamond S, Lee JJ, Golden SS, Palsson BO. 2016. Unique attributes of cyanobacterial metabolism revealed by improved genome-scale metabolic modeling and essential gene analysis. Proc Natl Acad Sci U S A 113:E8344–E8353.

CHAPTER 2: Phototaxis in a wild isolate of *Synechococcus elongatus*

2.1 Physiology of *S. elongatus* UTEX 3055



Phototaxis in a wild isolate of the cyanobacterium *Synechococcus elongatus*

Yiling Yang[a], Vinson Lam[b], Marie Adomako[b], Ryan Simkovsky[b,c], Annik Jakob[d,e], Nathan C. Rockwell[f], Susan E. Cohen[a,g], Arnaud Taton[b], Jingtong Wang[b], J. Clark Lagarias[f], Annegret Wilde[d,h], David R. Nobles[i], Jerry J. Brand[i,j], and Susan S. Golden[a,b,1]

[a]Center for Circadian Biology, University of California, San Diego, La Jolla, CA 92093; [b]Division of Biological Sciences, University of California, San Diego, La Jolla, CA 92093; [c]Food and Fuel for the 21st Century, University of California, San Diego, La Jolla, CA 92093; [d]Institute of Biology III, Faculty of Biology, University of Freiburg, D79104 Freiburg, Germany; [e]Spemann Graduate School of Biology and Medicine, University of Freiburg, D79104 Freiburg, Germany; [f]Department of Molecular and Cell Biology, University of California, Davis, CA 95616; [g]Department of Biological Sciences, California State University, Los Angeles, CA 90032; [h]BIOSS Centre of Biological Signalling Studies, University of Freiburg, 79106 Freiburg, Germany; [i]UTEX Culture Collection of Algae, The University of Texas at Austin, Austin, TX 78712; and [j]Department of Molecular Biosciences, The University of Texas at Austin, Austin, TX 78712

Contributed by Susan S. Golden, November 7, 2018 (sent for review July 31, 2018; reviewed by David M. Kehoe and Douglas D. Risser)

Many cyanobacteria, which use light as an energy source via photosynthesis, have evolved the ability to guide their movement toward or away from a light source. This process, termed "phototaxis," enables organisms to localize in optimal light environments for improved growth and fitness. Mechanisms of phototaxis have been studied in the coccoid cyanobacterium *Synechocystis* sp. strain PCC 6803, but the rod-shaped *Synechococcus elongatus* PCC 7942, studied for circadian rhythms and metabolic engineering, has no phototactic motility. In this study we report a recent environmental isolate of *S. elongatus*, the strain UTEX 3055, whose genome is 98.5% identical to that of PCC 7942 but which is motile and phototactic. A six-gene operon encoding chemotaxis-like proteins was confirmed to be involved in phototaxis. Environmental light signals are perceived by a cyanobacteriochrome, PixJ$_{Se}$ (Synpcc7942_0858), which carries five GAF domains that are responsive to blue/green light and resemble those of PixJ from *Synechocystis*. Plate-based phototaxis assays indicate that UTEX 3055 uses PixJ$_{Se}$ to sense blue and green light. Mutation of conserved functional cysteine residues in different GAF domains indicates that PixJ$_{Se}$ controls both positive and negative phototaxis, in contrast to the multiple proteins that are employed for implementing bidirectional phototaxis in *Synechocystis*.

cyanobacteria | photoreceptor | GAF domain | phototaxis | *Synechococcus elongatus*

Many organisms have evolved the ability to sense and alter their location in response to various beneficial or noxious stimuli. A classic example of this behavior is chemotaxis, in which the organism moves toward nutrients and away from toxins to locate an optimal position in a gradient of stimuli. Light is one of the most important environmental factors that affect life on Earth. For photosynthetic organisms, light is an energy source, but too much light can induce DNA damage and photooxidative stress. Thus, it is sensible to hypothesize that phototaxis is an evolutionarily beneficial behavior, guiding cell movement toward (positive) or away from (negative) a light source to receive optimal light energy. Phototactic behavior has been observed in all domains of life (1). In prokaryotes phototaxis has been characterized in purple bacteria and haloarchaea that swim with flagella or archaella as well as in cyanobacteria, which move over moist surfaces using type IV pili (2, 3).

The unicellular coccoid cyanobacterium *Synechocystis* sp. strain PCC 6803 (hereafter, "*Synechocystis*") has been studied extensively as a model for bacterial phototaxis (4, 5). Cyanobacterial phototactic signaling has been compared with that of chemotaxis in *Escherichia coli*, the best-studied bacterial taxis behavior (*SI Appendix*, Fig. S1*A*). The core signal-processing complex in *E. coli* consists of a methyl-accepting chemotaxis protein (MCP, chemoreceptor), a kinase (CheA), and an adaptor protein (CheW) (6). Binding of ligands to MCP regulates autophosphorylation of CheA, which can subsequently phosphorylate its response regulator, CheY. Phosphorylated CheY binds to the flagellar motor and changes the rotational direction of flagella, resulting in a reorientation of the cell. Dephosphorylation of CheY is catalyzed by a phosphatase, CheZ. An adaptation system consisting of methyltransferase CheR and methylesterase CheB confers a short-term memory for temporal comparison of ligand concentration, enabling cells to travel up or down the gradient (6–9). The core signal-processing complex is similar in *Synechocystis* phototaxis (*SI Appendix*, Fig. S1*B*), but instead of a chemical-binding domain, the MCP homolog has light-sensing domains at its N terminus (10). No homologs of CheR, CheB, or CheZ are found in the *Synechocystis* sp. strain PCC 6803 genome, indicating a different and still elusive mechanism of signal transduction in cyanobacterial phototaxis (11).

Cyanobacteria evolved a range of sensory photoreceptors that detect a rainbow of colors. Cyanobacteriochromes (CBCRs) comprise a class of phytochrome-related photoreceptors found only in cyanobacteria that have been reported to sense wavelengths that span the entire visible spectrum (12–15). CBCRs use GAF (cGMP phosphodiesterase/adenylate cyclase/FhlA) domains for photoreception and a C-terminal domain, such as a histidine kinase or an MCP domain, for signal output (12, 16). PixJ1 is a CBCR that mediates positive phototaxis in *Synechocystis*, and its inactivation results in negative phototaxis (11, 17). PixJ1 contains two GAF domains, but only the second carries the conserved Cys residue that covalently binds the bilin chromophore phycoviolobilin (PVB) (10). Purified PixJ1-GAF2 (PixJg2) switches between two conformational states that absorb either blue or green wavelengths (blue/green photocycle) when exposed to light, as does the GAF domain from the *Thermosynechococcus elongatus* photoreceptor TePixJ (18, 19).

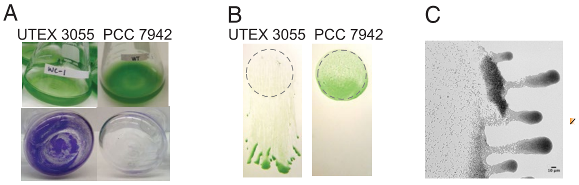
Other photoreceptors that regulate positive or negative phototaxis have been reported in *Synechocystis*. PixD, a sensor that uses flavin adenine dinucleotide (FAD) as a cofactor in a domain called "BLUF" (sensors of blue light using FAD), regulates the direction of phototaxis through binding to its interaction partner PixE, which inhibits positive phototaxis in its monomeric state (20, 21). The photoreceptor UirS/PixA is a UV-A sensor that, together with its response regulator UirR/NixB and a PatA-like protein called "LsiR/NixC," is involved in switching between positive and negative phototaxis (22). Moreover, UirS/UirR-regulated LsiR is required for negative phototaxis (23). The photoreceptor Cph2 is involved in the inhibition of phototaxis in the presence of blue light. Upon blue-light detection by its third GAF domain, Cph2 catalyzes c-di-GMP formation through a C-terminal GGDEF domain, which results in the inhibition of pili-based motility (*SI Appendix*, Fig. S1*B*) (24–26). It is still not clear whether any of the receptors described above is genuinely required for detecting light direction, because disruption of any of them results in a reversal of the direction of movement or general inhibition of motility, but not random movement, under directional light (3). A recent study demonstrated that *Synechocystis* cells sense a light source by acting as a microlens that focuses incoming light onto the membrane at the opposite side of the cell. Cells undergoing positive phototaxis move away from the focused-light spot and therefore move toward the light source (27). This model implies that the spherical shape is important for the mechanism of phototaxis for coccoid species and does not explain how phototaxis would occur in nonspherical species.

Although widely used for studies of circadian rhythms (28), light-regulated gene expression (29, 30), and metabolic engineering (31), the rod-shaped, unicellular model species *Synechococcus elongatus* PCC 7942 (hereafter, "PCC 7942") does not exhibit some environmentally relevant and often interrelated behavioral responses, such as biofilm formation and phototaxis. It has been shown that the ability to form biofilms is encoded in the PCC 7942 genome but is locked in a permanently repressed state unless mutations in the type II secretion or type IV pilus assembly system are made that prevent the production or secretion of a repressing agent (32, 33). Phototaxis in the thermophilic rod-shaped *T. elongatus* has been reported (34), but the paucity of genetic tools for that organism has limited investigation. Here, we report a recent wild isolate of *S. elongatus*, the strain UTEX 3055, that readily forms biofilms under laboratory conditions and shows strong phototactic behavior under directional light exposure. The mechanism of phototaxis in this cyanobacterium relies on a single 5-GAF domain containing an MCP-like photoreceptor that localizes at the cell poles. Despite structural and spectroscopic similarities to the blue/green photoreceptor PixJ of other cyanobacteria, this receptor alone is responsible for both negative and positive directional orientation. Like *Synechocystis*, UTEX 3055 focuses light with its rod-shaped cell, roughly opposite to the direction of incidence.

Results

A Wild *S. elongatus* Isolate Shows Photo-Induced Migration and Biofilm Formation. To evaluate whether PCC 7942, which has been a domesticated laboratory strain for more than four decades (35), was once capable of phototaxis or other environmentally relevant behaviors, we isolated a wild strain of *S. elongatus*. Cyanobacterial cells were enriched from rock-attached scum in Waller Creek at the University of Texas, Austin, where the species had been isolated previously (36, 37). Cells of the wild isolate have a morphology, as observed by light microscope, very similar to that of PCC 7942 (*SI Appendix*, Fig. S2). Sequencing and assembly of the genome of this wild isolate (hereafter, "UTEX 3055") revealed a 98.46% average nucleotide identity (ANI) with PCC 7942. Based on their identical 16S rRNA sequences and an ANI >95%, we concluded that they are the same species. UTEX 3055 has chromosomal insertions and deletions relative to PCC 7942, a large plasmid homologous to pANL of PCC 7942 but with expansions, and a 24-kb plasmid (GenBank accession nos.: CP033061, CP033062, and CP033063). UTEX 3055 also has an inversion in the chromosome relative to PCC 7942 that has been previously described for another domesticated laboratory strain, *S. elongatus* PCC 6301 (38). UTEX 3055 has 38,774 SNPs compared with PCC 7942, of which 9,446 predict nonsynonymous amino acid substitutions (*SI Appendix*, Table S1). Because PCC 7942 is easily manipulated genetically and because it is the premier model organism for bacterial circadian rhythms, we first examined whether the genetic tools used for PCC 7942 would work in the wild isolate and whether UTEX 3055 exhibits similar circadian rhythms. The UTEX 3055 genome includes the entire set of *S. elongatus* clock genes (*SI Appendix*, Table S2). When transformed with the same luciferase-reporter vector used for PCC 7942 circadian studies and with the same transformation protocol developed for PCC 7942, UTEX 3055 produced rhythmic patterns of bioluminescence, demonstrating both natural competence and the presence of a functional circadian clock system (*SI Appendix*, Fig. S3).

Despite sharing high nucleotide identity, UTEX 3055 shows some interesting phenotypes that are absent in the laboratory strain. Wild-type PCC 7942 exhibits an entirely planktonic phenotype when grown in liquid medium and remains suspended indefinitely in the absence of agitation. Only when specific mutations are introduced does PCC 7942 tend to settle and form biofilms (32). In contrast, wild-type UTEX 3055 cells flocculate and form biofilms on the wall of a culture flask under the same conditions (Fig. 1*A*). Moreover, UTEX 3055 exhibits strong taxis toward a lateral directional light source, whereas PCC 7942 does not (Fig. 1*B*). As observed in *Synechocystis* (39, 40), UTEX 3055 forms finger-like projections directed toward the light source on soft agarose plates (Fig. 1*C* and Movies S1–S3). Because bacterial movement can be monitored by microscopy (*SI Appendix*, Fig. S4), we tested whether individual UTEX 3055 cells respond to a light-intensity gradient projected onto the surface by placing a dark filter on one side of the light field. The light gradient that was generated parallel to the plane of the surface by a perpendicular light source resulted in migration that is significantly different from the biased movement under lateral directional light (Fig. 1 *D* and *E* and compare Movie S4 with Movie S5) but still exhibits directionality compared with a uniform circular distribution (Hodges–Ajne test). The results suggest that cells may sense light direction rather than a gradient of intensity during phototaxis, but more refined experiments are needed to exclude the effect of light scattering caused by the agarose medium. In contrast, PCC 7942 cells were minimally motile and had a random direction of movement under directional light when observed in the microscope (Movie S6), which is consistent with the lack of phototaxis on agarose plates (Fig. 1*B*).

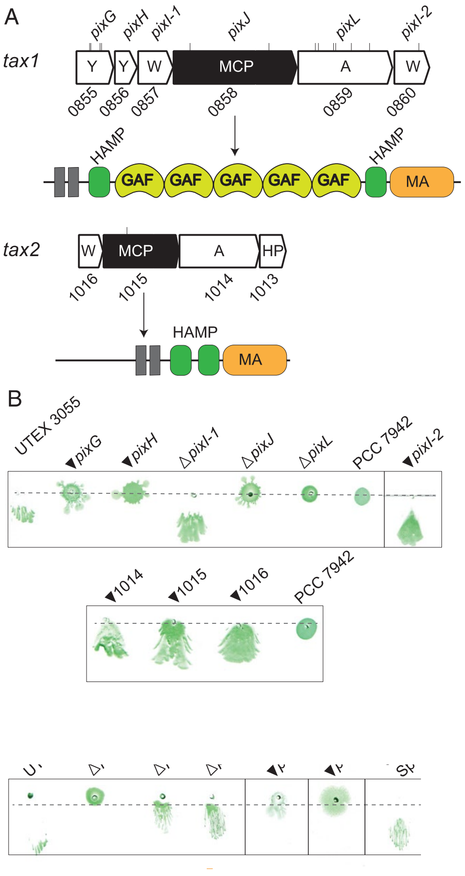


o plots in *D*; $r = 0$ indicates perfect nondirectionality, and $r = 1$ indicates maximal clustering in one direction. The apparent oscillation pattern may represent noise in the measurements.

Examination of the distribution of cell speed after transferring UTEX 3055 from a 12-h light:12-h dark environment to a constant-light environment showed no time-of-day variation (*SI Appendix*, Fig. S5), indicating that UTEX 3055 motility is not under circadian control.

Genetic Identification of a Phototaxis Operon. The ability to manipulate UTEX 3055 genetically with tools and techniques developed for PCC 7942 enables the dissection of the molecular mechanism of UTEX 3055 phototaxis. In other organisms phototaxis-signaling components are encoded by chemotaxis-like (Che) genes that are often encoded in operons. Two Che operons, named "*tax1*" and "*tax2*" to follow the nomenclature used in *Synechocystis*, were found in both the PCC 7942 and UTEX 3055 genomes (Fig. 2*A*). Both operons encode homologs of an MCP, CheA and CheW, whereas *cheY* is present only in *tax1*. One gene in the *tax1* operon (Synpcc7942_0858/UTEX3055_0948) is predicted to encode a protein with N-terminal GAF domains and a C-terminal MCP domain, similar to the *Synechocystis* protein PixJ1 that mediates positive phototaxis (Fig. 2*A*) (10). However, the version of this gene in UTEX 3055, also present in PCC 7942, encodes five contiguous GAF domains that all have the potential to bind a chromophore, whereas *Synechocystis* PixJ1 carries a single bilin-binding GAF domain (10). Thus, it is reasonable to speculate that the Synpcc7942_0858 protein (hereafter, "PixJ_{Se}") functions as a photoreceptor that mediates phototaxis in UTEX 3055. To test this hypothesis, a *pixJ* disruption mutant was created by transforming UTEX 3055 with a mutagenic cosmid that carries a transposon insertion in *pixJ* from the PCC 7942 uni-gene set (UGS) library (41, 42). Indeed, inactivation of *pixJ* resulted in loss of phototaxis in UTEX 3055 (Fig. 2*B*). Disruption of other genes in the *tax1* operon, such as Synpcc7942_0859 (CheA-like; hereafter, "PixL_{Se}"), Synpcc0855 (CheY-like; PixG_{Se}) or Synpcc0856 (CheY-like; PixH_{Se}), also led to nonphototactic phenotypes (Fig. 2*B*). However, disruption of Synpcc7942_0857 or Synpcc0860 (both are CheW-like; hereafter "PixI_{Se}-1" and "PixI_{Se}-2") individually did not affect phototaxis, nor did individual disruptions of the MCP-like (Synpcc7942_1015), CheA-like (Synpcc7942_1014), or CheW-like (Synpcc7942_1016) genes in the *tax2* operon (Fig. 2*B*). Proficiency for phototaxis in the mutants that are disrupted upstream of *pixJ* argues against polar effects contributing to the *pixJ* phenotype. To verify that the nonphototactic phenotype is caused by mutations in *tax1* genes, a shuttle vector that carries the corresponding intact UTEX 3055 ORF was recombined into the chromosome of respective mutants at neutral site I (NS1) (43). Introduction of *pixJ*, *pixL*, *pixG*, or *pixH* restored the respective mutant's phototaxis; however, complementation by *pixJ* and *pixH* was observed only at specific expression levels (Fig. 2*C* and *SI Appendix*, Fig. S6 *B* and *C*). The extent of phototaxis was rarely restored to wild-type levels by expression of a gene from an ectopic site, but because the mutants designated as nonphototactic showed no biased movement in more than 10 assays, even modest restoration of biased movement was scored as complementation.

Because wild-type UTEX 3055 cells switch from directional movement under directional light to nondirectional movement in the dark (*SI Appendix*, Fig. S7), we tested whether motility that is unrelated to direction is also affected in the phototaxis mutants by assessing the speed of movement of the *pixJ* mutant while applying a dark pulse. The results showed that the mutant cells are motile and move in random directions under both light and dark conditions (*SI Appendix*, Fig. S7). This experiment demonstrates that *tax1* does not control type-IV pilus biogenesis and function in *S. elongatus*. Throughout this study we found that motile but nonphototactic mutants yield varying colony phenotypes

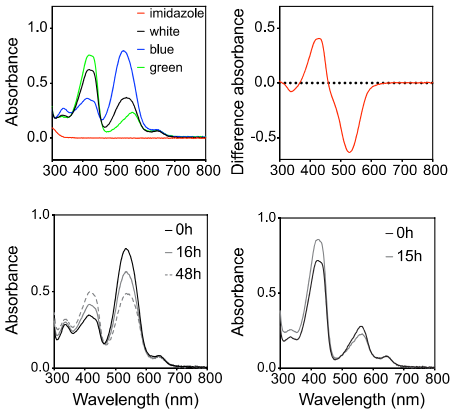


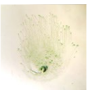
55 strains expressing an SpSm cassette and lacking *pixJ* served as positive and negative controls, respectively. See *SI Appendix*, Fig. S6*A* for the experimental setup.

in phototaxis assays. Sometimes all the cells appeared to be trapped at the inoculation site, forming a spot with a smooth edge; alternatively, some cells appeared to burst out in a few random or in all directions and sometimes formed twisted tree-like branches (*SI Appendix*, Fig. S8). We speculate that this phenotypic variability relates to differences in agarose surface conditions. Because disruption of *pixJ* or *pixL* causes loss of phototaxis in UTEX 3055, we asked whether the nonphototactic phenotype of PCC 7942 is due to impairment of key proteins encoded in the *tax1* operon. However, we found that both *pixJ* and *pixL* from PCC 7942 restore phototaxis when expressed in their respective UTEX 3055 mutants (*SI Appendix*, Fig. S9*A*), indicating that loss of phototaxis in PCC 7942 is caused by factors other than genetic differences in the photoreceptor or kinase genes.

PixJ$_{Se}$ Is a Blue/Green-Sensing DXCF Subfamily CBCR Photoreceptor. PixJ$_{Se}$ harbors five contiguous GAF domains at its N terminus. Alignment and clustering of the sequences of these domains with GAF domains from other CBCR and phytochrome proteins demonstrates that the GAFs of PixJ$_{Se}$ are most similar to each other, followed by a high degree of similarity with the previously described DXCF/CBCR/GAF domains PixJg2 from *Synechocystis* and the single GAF from TePixJ of *T. elongatus*, all of which contain two conserved Cys residues (*SI Appendix*, Fig. S10 *A* and *B*). The DXCF subclass of CBCRs detects blue/green light via its PVB chromophore (12); PVB is derived by chemical isomerization of the precursor phycocyanobilin (PCB) following its covalent attachment to the protein via thioether linkages to the two conserved Cys residues (44). The first Cys is essential for binding the chromophore, and the second Cys is essential for detecting light at shorter wavelengths (12, 45). Light excitation triggers photoisomerization of the chromophore's C15, C16 double bond, which leads to changes in chromophore–protein interaction that control signal transmission to the output domain. To determine whether any of the PixJ$_{Se}$ GAF domains also binds a chromophore, we assayed for fluorescence of covalently bound bilin chromophores in SDS/PAGE gels under UV light in the presence of zinc acetate (46). Protein gel electrophoresis of a wild-type UTEX 3055 lysate showed multiple fluorescent bands, indicating the presence of chromophore-bound proteins, including the abundant light-harvesting phycobiliproteins and minor species (*SI Appendix*, Fig. S10*C*). We identified one of these minor bands as PixJ$_{Se}$ based on the protein's predicted size of 153 kDa, the absence of this band in the *pixJ*-knockout strain, and the band's reappearance upon complementation (*SI Appendix*, Fig. S10*C*). Consistent with complementation data demonstrating that PixJ$_{Se}$ from PCC 7942 is functional (*SI Appendix*, Fig. S9*A*), a PCC 7942 lysate also produces a fluorescent PixJ band in this assay (*SI Appendix*, Fig. S9*B*). These results demonstrate that PixJ$_{Se}$ is a chromophore-bound protein that has the potential for light sensing.

To determine the absorption spectrum of PixJ$_{Se}$, we heterologously expressed a single GAF domain, GAF2, to avoid the difficulties of purifying a large membrane protein. The well-studied cyanobacterial phytochrome Cph1 was also purified as a positive control (47). Recombinant C-terminally tagged GAF2 (PixJ$_{Se}$GAF2-His) was expressed in PCB-producing *E. coli* and purified by nickel-affinity chromatography (*SI Appendix*, Fig. S10*D*). The presence of a covalently bound chromophore in both proteins was confirmed by zinc-induced fluorescence on SDS/PAGE (*SI Appendix*, Fig. S10*E*). The UV-VIS absorption and difference spectra of PixJ$_{Se}$GAF2-His show two major peaks at 429 nm and 529 nm (Fig. 3 *A* and *B*), corresponding to blue-absorbing (Pb) and green-absorbing (Pg) states, respectively. The Pb state of PixJ$_{Se}$GAF2-His appeared yellow after green-light exposure, which was efficiently converted to magenta after blue-light irradiation (Fig. 3*A*, *Inset*). A dark-reversion experiment showed that, after blue-light irradiation, the Pg state is slowly converted to the Pb state in the dark (Fig. 3*C*). In contrast, the green-light–induced Pb state was stable in the dark (Fig. 3*D*). These experiments demonstrate that PixJ$_{Se}$GAF2-His photoconverts between the Pb and Pg states, where Pb is the dark-adapted state and Pg is the photoproduct, similar to the previously studied blue/green sensors PixJg2,



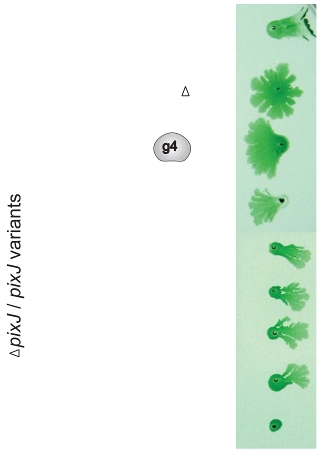


TePixJg, and cce_4193g2 (10, 18, 48). Based on the sequence conservation of the five GAF domains in PixJ$_{Se}$, we hypothesize that the characterized photoresponse of GAF2 is common for all five domains.

Based on the blue/green photocycle of PixJ$_{Se}$GAF2-His, we speculated that the in vivo photosensory behavior of PixJ$_{Se}$ involves sensing blue and green light during phototactic movement. To test this hypothesis, phototaxis assays were performed on wild-type UTEX 3055 under lateral directional illumination from different colors of lights. Blue light alone resulted in bidirectional migration of UTEX 3055 cells, with high intensities causing repulsion and low intensities causing attraction (Fig. 3*E*). In contrast, green-light exposure did not elicit directed cell movement, regardless of light intensity. However, the combination of blue and green light induced positive phototaxis in a manner that resembles that of white-light exposure. Lateral directional red-light illumination of wild-type UTEX 3055 significantly stimulated cell growth but did not induce any phototactic movement, and the same effect was observed for a red-and-green combination (*SI Appendix*, Fig. S11). Red and blue light together stimulated migration similar to that of blue light alone but with better cell growth and enhanced positive migration (*SI Appendix*, Fig. S11*A*). As a negative control, *pixJ*-knockout mutants did not respond to either blue or blue-and-green light (*SI Appendix*, Fig. S11*B*). These results are consistent with the blue/green cycles of PixJ$_{Se}$ and further demonstrate the role of this photoreceptor in determining the direction of motility in response to blue and green illumination.

GAF Domain Signaling Is Important for Regulating the Direction of Cell Movement. To better understand the role of the multiple GAF domains (GAF1–5) in PixJ$_{Se}$, we made Cys → Ala mutations in the first Cys residue, which is essential for chromophore binding (10, 49), in various GAF domains, singly and in combination. Most mutants that retain at least one bilin-binding GAF domain exhibited positive phototaxis to white light, like the wild type (*SI Appendix*, Fig. S12). These include single mutations in GAF1, GAF3, or GAF5 and almost all other double, triple, or quadruple mutants tested (*SI Appendix*, Fig. S12). Except for the quintuple GAF1–5 mutant, all mutants produced fluorescent bands upon SDS/PAGE zinc staining, demonstrating that PixJ$_{Se}$ mutant proteins with at least one functional GAF domain continue to bind chromophores (*SI Appendix*, Fig. S13). The quintuple mutation of all GAF domains resulted in a nonphototactic phenotype similar to that of the *pixJ*-knockout mutant (Fig. 4 and *SI Appendix*, Fig. S12). Mutation of the GAF4 domain alone or both GAF4 and GAF5 resulted in negative phototaxis under light conditions where the wild type shows positive phototaxis (Fig. 4). Mutation of GAF domains other than GAF5 in addition to GAF4 restored positive phototaxis, indicating that integration of signals among the domains is important for signaling. Thus, PixJ$_{Se}$-mediated phototaxis requires at least one chromophore-bound GAF domain; the presence of multiple GAF domains enables switching between positive and negative directions of movement; and GAF4 is specifically important for positive phototaxis.

Bipolar Localization of Photoreceptor PixJ$_{Se}$. To explore the function of PixJ$_{Se}$ further, we investigated whether this photoreceptor accumulates at a specific location within the cell. YFP was fused to the C terminus of PixJ$_{Se}$ with a GSGGG linker and introduced into the UTEX 3055 *pixJ* mutant. An immunoblot showed the expression of full-length fused protein (*SI Appendix*, Fig. S14*A*), while zinc staining confirmed the presence of covalently bound chromophore (*SI Appendix*, Fig. S14*B*). YFP-tagged PixJ$_{Se}$ appears to be fully functional in the cell because PixJ$_{Se}$-YFP restored the *pixJ* mutant's phototaxis at uninduced (leaky) expression levels (Fig. 5*A*, *Left*). Notably, tight fluorescent clusters were observed on membranes primarily at cell poles (Fig. 5*B*, *Left*), similar to those seen for chemoreceptor complexes in *E. coli* (8). Note that this polar localization is not caused by the intrinsic localization of YFP, because previous studies showed that unfused YFP distributes homogeneously throughout the cytoplasm (50). However, overexpression of PixJ$_{Se}$-YFP caused a nonphototactic phenotype (Fig. 5*A*, *Right*), possibly due to the disruption of polar localization or stoichiometry of the signaling complex (Fig. 5*B*, *Right*). PixJ$_{Se}$-YFP



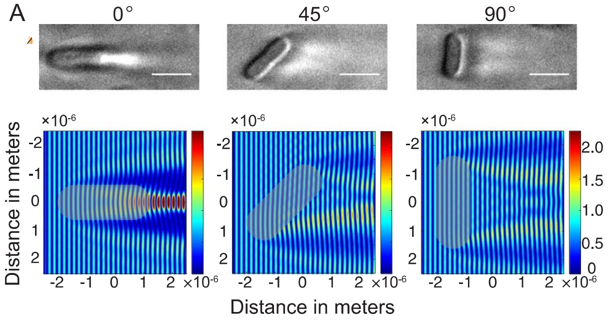
also showed bipolar localization when heterologously expressed in *E. coli* (*SI Appendix*, Fig. S14*C*), indicating that polar localization either may use an evolutionarily conserved mechanism or is an inherent property of the protein.

Lensing Effect of Rod-Shaped *S. elongatus* Cells. A compact complex is consistent with the hypothesis that PixJ_{Se} acts in conjunction with a cellular lens. The round cells of the cyanobacterium *Synechocystis* act as microlenses that focus incoming light onto the membrane at the back of the cell, enabling cells to sense the light direction (27). We tested whether such a lensing mechanism also operates in rod-shaped *S. elongatus* cells. Both PCC 7942 and UTEX 3055 exhibited a strong lensing effect under directional illumination (Fig. 6*A*, *SI Appendix*, Fig. S15, and Movie S7). Although the focusing effect was greatly reduced at an orientation of 45° and 90° to the light direction, this result shows that a rod-shaped cell is also capable of lensing, focusing incoming light onto the opposite side of the cell. Thus, the optical properties of the cell resemble those known from *Synechocystis*. Mathematical finite difference time domain (FDTD) simulations with an assumed uniform refractive index of 1.4 (51) of the cell immersed in water reproduced this lensing effect (Fig. 6*B*). Although a uniform refractive index of the cells is only an approximation, it seems that the microscopically observed lensing effects can be modeled sufficiently by this simulation approach. Notably, lensing was not affected in cells that lack PixJ_{Se} (Movie S8). As shown for *Synechocystis*, the cells did not respond to the projection of a light gradient, indicating that the stimulus is a directional light source rather than a spatial change in light intensity (Fig. 1*D*).

Discussion

UTEX 3055 Is a Useful Model Cyanobacterium for Studying Environmentally Relevant Behaviors. The discovery of robust biofilm formation and phototaxis in an isolate of *S. elongatus* that can be studied using the extensive genetic tools developed for PCC 7942 provides an opportunity to understand environmentally relevant cyanobacterial behaviors. The functionality of *pixJ* and *pixL* genes of PCC 7942, when expressed in UTEX 3055 mutants, argues strongly that original isolates of PCC 7942 were also phototactic before laboratory propagation. Similarly, although wild-type PCC 7942 does not form biofilms under laboratory growth conditions, a number of mutations can enable biofilm formation by triggering the expression of proteins that contribute to the biofilm phenotype (32, 33). Taken together, the data suggest that domestication over four decades may have selected for either the loss or repression of these phenotypes. These losses have limited the use of PCC 7942 for studying complex behaviors that are both environmentally and perhaps commercially important. UTEX 3055 preserves the circadian clock system characterized in PCC 7942, is naturally competent, and performs homologous recombination in a manner similar to PCC 7942. In passaging the strain, we have been

Scale bars: 2 μm.)



mindful of our culturing techniques to prevent the loss of these traits that are shared with PCC 7942 as well as the wild traits of biofilm formation and phototaxis (*Methods*). Because of the high sequence identity between the strains, a wide array of genetic tools developed for PCC 7942 can be directly applied to UTEX 3055, establishing the wild strain as a model organism.

Phototaxis in *S. elongatus* Is Encoded by a Single Phototaxis Operon. Mutagenic disruption of phototaxis pathway genes in the *tax1* operon leads to loss of directional light-dependent migration in UTEX 3055. However, these phototaxis mutants and cells of PCC 7942 are still motile and migrate in random directions independent of the location of a directional light source (*SI Appendix*, Fig. S7 and Movie S6). These results indicate that the *tax1* operon is responsible only for phototactic signaling. Mutation in *tax2* genes did not affect phototaxis, and the stimuli to which this operon may be related are unknown. Notably, neither operon is required for cell motility, which is most likely mediated by peritrichous type-IV pili based on the moving behavior of UTEX 3055 on the agarose surface (Movie S7) and the observed distribution of pili in PCC 7942 (52). We suspect that the motility of phototaxis mutants causes the observed variation in colony shapes on the semisolid agarose surface in different phototaxis assays, frequently observed as out-grown bursts (Figs. 2*B* and 4 and *SI Appendix*, Fig. S8). These bursts occur in random directions with respect to the orientation of lateral illumination, suggesting these are not phototactic migrations caused by second-site suppressor mutations. A reconstruction experiment showed that a minimum of 1 in 100 cells must be phototactic to form functional moving structures toward a light source. This ratio is unlikely to be reached by a spontaneous second-site suppression mutant that emerges during the course of the assay. *Synechocystis* contains a similar phototaxis operon in its genome, but disruption of the genes in the homologous operons of *Synechocystis* results in either reversal of phototactic direction or a nonmotile phenotype due to the loss of type-IV pili (11, 17). This comparison suggests that *Synechocystis* integrates stimuli through competing directional sensing or signaling pathways, one for positive phototaxis and others for negative phototaxis, whereas the single *tax1* operon in UTEX 3055 is solely responsible for directional migration. The PixJ$_{Se}$ and PixL$_{Se}$ homologs encoded in the genome of PCC 7942 are functional, indicating that the loss of directional movement results from other defects, such as decreased motility (Movie S6) or exopolysaccharide secretion or in downstream signaling to relay the directionality of light.

Bidirectional Phototaxis Is Regulated by Integration of Light Stimuli in a Single Photoreceptor. *S. elongatus* UTEX 3055 shows bidirectional responses to lateral directional blue-light stimulation, with strong intensities repelling and weak intensities attracting the cells. Repulsion by high blue-light intensities is reasonable because blue wavelengths can cause cellular damage. However, insufficient light is also detrimental to the photosynthetic organism, consistent with positive phototaxis under weak blue-light intensities. The results show that *S. elongatus* seeks optimal light conditions by adjusting its location to tune the intensity of blue wavelengths contained in the light source. *S. elongatus* inhabits aquatic environments, where the components and intensity of light vary throughout the diel cycle and in a depth-dependent manner. In such a complex environment, cells in a microbial consortium use phototaxis to microadjust their locations in a biofilm to enhance exposure to sunlight for photosynthesis while avoiding damage caused by too much light. We hypothesize that the intensity of blue light relative to other colors, which are attenuated by depth in the water column, acts as a signal to determine the direction of phototactic movement. Short-wavelength light-induced bidirectional phototaxis has also been observed in *Synechocystis*, probably mediated by the blue-light–sensing photoreceptors PixJ1, Cph2, and PixD/PixE (10, 21, 24, 25) and a UV-A–responsive UirS/UirR-controlled LsiR expression system (23). The presence of multiple sensing systems indicates the importance of blue/UV sensing, which may be a common strategy used by cyanobacteria to find optimal light conditions. Similar bidirectional regulation strategies are observed in other taxis systems, such as pH taxis and thermotaxis (53, 54). Green light alone did not induce any phototactic movement in UTEX 3055, but blue and green light presented together stimulated strong positive phototaxis, similar to that of white light. This result correlates well with the blue/green absorbance of purified PixJ$_{Se}$GAF2. Notably, *Synechocystis* shows positive phototaxis to green-light stimulation (23, 55). Such different responses to blue and green light in *Synechocystis* and *S. elongatus* may reflect differences in their natural habitats and/or the use of multiple photoreceptors to guide taxis in *Synechocystis*.

The most striking difference among the photoreceptors PixJ$_{Se}$, PixJ1 of *Synechocystis*, and TePixJ of *T. elongatus* is the presence of five DXCF/GAF domains at the N terminus of the UTEX 3055 protein as compared with only one DXCF/GAF domain in PixJ1 and TePixJ. An additional functional complexity is evident in the *S. elongatus* protein, in which removal of chromophore attachment through Cys mutations in GAF4 or GAF4/GAF5 resulted in negative phototaxis, indicating that bidirectional phototaxis can be achieved through the integration of signals detected by a multi-GAF–containing photoreceptor. However, we cannot exclude the presence of other photoreceptors that contribute to phototaxis. No other combination of GAF4 and other GAF-domain mutations resulted in negative phototaxis, suggesting that GAF4-mediated conformational changes play a critical role in reversing the output signal from more N-terminal GAF domains. It should be noted that we have not presented direct evidence to show that the other four GAF domains (GAF1, -3, -4, and -5) have the same photocycle as GAF2; however, given the sequence similarity of these domains to each other, to TePixJ, and to PixJg2 (*SI Appendix*, Fig. S10*B*), it is likely that all the PixJ$_{Se}$ GAF domains have blue/green photocycles and bind PVB. This result also demonstrates that the presence of multiple GAF domains in PixJ$_{Se}$ is neither a matter of simple redundancy nor solely for the purpose of signal addition and amplification. Instead, the multiple GAF domains appear to play a regulatory role in controlling the direction of movement. Numerous multiple-GAF CBCRs are encoded in genomes throughout the cyanobacterial clade. In filamentous *Nostoc punctiforme*, photoreceptor NpF1883 contains three chromophore-bound GAF domains (NpF1883g2/3/4) that all show similar blue/teal photocycles (14). An MCP-like photoreceptor in *Nostoc punctiforme*, PtxD, contains six chromophore-bound GAF domains, all with different photocycles (15). Loss of PtxD has been reported to cause loss of phototaxis of motile filaments called "hormogonia" (56), a phenotype similar to that observed for PixJ$_{Se}$ in this study. These examples suggest that multi-GAF photoreceptors present a common mechanism for regulating the direction of phototaxis.

How Do Rod-Shaped Cells Determine the Direction of a Light Source? Spherical *Synechocystis* cells physically sense light direction by acting as a microlens that focuses the incoming directional light on the portion of the membrane at the opposite side of the cell. Cells then actively move away from the bright spot, resulting in positive phototaxis (27). Although not spherical in shape, *S. elongatus* also lenses light. However, it is not clear yet whether this lensing effect is used by the cell to determine the direction of incoming light. Interestingly, PixJ$_{Se}$ localizes to the cell poles in a pattern that is similar to that of chemoreceptors in *E. coli* and some other prokaryotic species (57). Delocalization of PixJ$_{Se}$ and its redistribution throughout the cell membrane was concurrent with the abolition of phototaxis (Fig. 5). Similar polar localization of the PixJ homolog was reported in *T. elongatus* (58). We hypothesize that lensing of incoming light on localized photoreceptor complexes enhances directional light detection and orientation to enable phototaxis.

Methods

Bacterial Strains, Growth Conditions, and DNA Manipulation. The plasmids and strains used in this study are described in *SI Appendix*, Tables S2 and S3. All *S. elongatus* strains were cultured in BG-11 medium as previously described (43), illuminated with 70–150 μmol photons·m^{-2}·s^{-1} fluorescent light. Plasmids were constructed using the GeneArt Seamless Cloning and Assembly Kit (Life Technologies) and propagated in *E. coli* DH5α with antibiotics. Cyanobacterial mutants were generated by transforming with knockout vectors or corresponding PCC 7942 UGS library plasmids (41, 42). Cys → Ala substitutions were introduced into the GAF domains by site-directed mutagenesis (Pfu Turbo DNA polymerase; Agilent), and all constructs were verified by sequencing. Homogenous segregation of alleles was confirmed by PCR for all knockout mutants.

UTEX 3055 Isolation. *S. elongatus* UTEX 3055 was isolated from Waller Creek at The University of Texas at Austin at the following GPS coordinates: latitude 30°17′5.636″N, longitude 97°44′4.501″W. Samples were given three designations based on collection location: P1, collected from a small pool of water in rocks; M1, collected from soil particles and water from embankment mud; and R1, collected by scraping blue-green growth off rocks at the waterline. A 4-mL aliquot from each sample location was used to inoculate each of nine 50-mL glass culture tubes containing 40 mL of sterile Kratz–Myers C medium (KMC) (36). The tubes were placed on a temperature gradient (38.5–51 °C) under a 12-h light:12-h dark diel cycle and bubbled with 1.5% CO$_2$ in air. Cultures were screened by microscopy for the presence of putative strains of *Synechococcus* spp. Based on the results of this screening, serial dilutions (1:10–1:1,000) from the P1 isolate grown at 42 °C were plated on KMC agar supplemented with 25 μg/mL cycloheximide. Unialgal colonies were picked from plates after 17 d and were used to inoculate 10-mL volumes of KMC. All P1 isolates were maintained at 42 °C under a 12-h:12-h diel cycle. The internal transcribed spacer (ITS) regions of unialgal isolates were amplified by PCR and sequenced (59). BLAST searches against the National Center for Biotechnological Information database identified one unialgal isolate with 99% similarity to the ITS region of PCC 7942.

The unialgal strain, archived as AMC2389, was made axenic through streaking for single colonies; the lack of other bacteria was verified using BG-11 plates containing 0.04% (wt/vol) glucose and 5% (vol/vol) LB broth. The axenic strain was archived as AMC2388. To avoid domestication, UTEX 3055 culturing was alternated in liquid and on solid medium, and the strain was revived from the same frozen stock every 6 mo. We observed that continuous liquid culture promotes loss of biofilm formation, and passaging exclusively on standard hard-agar plates promotes loss of phototaxis.

Sequencing, Assembly, and Annotation of UTEX 3055. The UTEX 3055 genome was sequenced on the Illumina MiSeq platform using a paired-end library construction with a 300-bp insert size as well as with PacBio RS Single Molecule Real-Time sequencing. A 2.76-Mbp chromosome and an 89-kb plasmid were assembled from PacBio sequence data using Canu (60). Illumina MiSeq reads were mapped to the assembly, and unmapped reads were extracted and assembled into a 24-kb plasmid using SPAdes (61). The assembly was corrected for small errors obtained with Illumina data using Pilon (62). Contigs were circularized using Circlator (63). The finished assembly was annotated by JGI (64), and annotation was further refined manually.

Phototaxis Assay. BG-11 medium solidified with 0.3% agarose (wt/vol) and 10 mM sodium thiosulfate was used for phototaxis assays. Wild-type *S. elongatus* and mutants grown in liquid BG-11 were adjusted to an OD$_{750}$ of 0.6–1.0, and 2-μL samples of culture were spotted at specific positions on the surface of agarose plates. After inoculation, the plates were placed in a dark box with one side opening toward a fluorescent light. Photographs were taken after 3–5 d of incubation. Phototaxis to LED light was performed as shown in *SI Appendix*, Fig. S11*A*.

Expression and Purification of GAF-Domain Protein. *E. coli* strain LMG194, which contains plasmid pPL-PCB for synthesis of PVB, was transformed with pAM5498 for expression of PixJ$_{Se}$GAF2-His or with pBAD_Cph1 (56) as a positive control. Overnight cultures were used to inoculate 100 mL of LB medium containing 50 μg/mL ampicillin and 12 μg/mL kanamycin to an OD$_{600}$ of 0.5. After about 8 h of growth at 37 °C, this 100-mL culture was added to 900 mL of LB medium supplemented with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) to induce PCB expression. After a 1-h incubation, L-arabinose was added to a final concentration of 0.04% (wt/vol) to induce PixJ$_{Se}$GAF2-His expression (0.02% for Cph1), and incubation was continued at 37 °C for 5 h. Cells were harvested by centrifugation at 3,795 × *g* for 10 min at 4 °C, and pellets were frozen at −20 °C or were immediately resuspended in lysis buffer (50 mM Tris, 500 mM NaCl, and 20 mM imidazole, pH 8.0). Cells were lysed with a homogenizer (Emulsiflex C3; Avestin) at 15,000–20,000 psi for 10 min and then were subjected to centrifugation at 32,500 × *g* for 40 min to remove cell debris. The soluble fraction of lysates was incubated with an Ni-NTA gravity column (Qiagen). The unbound proteins were removed by washing with 50 mL lysis buffer, and bound protein was recovered with 10 mL of elution buffer (50 mM Tris, 500 mM NaCl, and 250 mM imidazole, pH 8.0).

Zinc Blots. Chromophore incorporation was assayed as previously reported (46). SDS/PAGE gels were soaked in 100 mM zinc acetate with gentle shaking for 30 min in the dark. The zinc-impregnated gel was then irradiated with 305-nm UV light, and the fluorescent bands were recorded with a FluorChem HD2 system (Alpha Innotech).

Spectroscopy. Cuvettes containing a 500-μL protein solution were exposed to blue or green LED light for 2 min, and then absorption was measured immediately with a UV-Vis spectrophotometer (NanoDrop 2000c; Thermo Scientific). All measurements were performed in the dark at room temperature.

Microscopy. An Olympus IX71 inverted microscope with an attached WeatherStation environmental chamber was used for imaging and time-lapse movie production. Images were captured using a CoolSnap HQ2 CCD camera (Photometrics).

For images of PixJ$_{Se}$-YFP localization, a 2-μL sample of culture grown to an OD$_{750}$ of 0.6 was fixed on an 1.2% agarose (wt/vol) pad in BG-11 for imaging. TRITC filters (excitation 555/28 nm and emission 617/73 nm) were used to image cyanobacterial autofluorescence. YFP filters (excitation 500/20 nm and emission 535/30 nm) were used to image PixJ$_{Se}$-YFP protein localization. Series of Z-stack images were taken and deconvolved using the softWoRx imaging program (Applied Precision). For time-lapse movies 2-μL cell samples were placed on a pad of 0.3% (wt/vol) agarose in BG-11, left to absorb into the agarose for 5 min, and then covered with a coverslip (*SI Appendix*, Fig. S4*A*). A 10× or 20× objective was used for monitoring projections of colony fingers. For tracking single-cell motility, a 40× or 100× objective was used (*SI Appendix*, Fig. S4*B*). Lateral illumination was provided by a white LED. Images were acquired at 0.5-s or 2-s intervals. To image cells in the dark, the LED light was turned off, and images were captured using the microscope's transillumination lighting system (a 100-W halogen lamp), with a limited exposure time of 0.01–0.05 s.

Lensing Simulation. The optical field distribution of blue light (460 nm) propagating through a single cell was calculated using FDTD simulations as previously described (27). A space grid of $\Delta x = \Delta y = 5$ nm was used for the simulation of a 5 × 5 μm square array. The refractive index of the cell was approximated to be 1.4, thus leading to an effective refractive index of 1.05 of the cell relative to that of water. The cell length of 3.25 μm with an area of 3.69 μm^2 was determined by averaging microscopy images of cells (62 and 77 cells, respectively).

ACKNOWLEDGMENTS. We thank Dr. J. S. Parkinson for providing the *E. coli* strain used for receptor localization, members of the James W. Golden and S.S.G. laboratories for helpful discussions, and the University of California, San Diego qBio Hacker Laboratory for instruments and equipment used in phototaxis assays. The research reported in this publication was supported by NIH Grant R35 GM118290 (to S.S.G.); National Science Foundation (NSF) and US-Israel Binational Science Foundation Grant NSF-BSF 2012823 (to S.S.G); NSF Grant 1201881 (to D.R.N.); an Excellence Initiative of the German Research Foundation Spemann Graduate School Award (to A.J.); and Chemical Sciences, Geosciences, and Biosciences Division, Office of Basic Energy Sciences, Office of Science, US Department of Energy Grant DOE DE-FG02-09ER16117 (to J.C.L.).

1. Jékely G (2009) Evolution of phototaxis. *Philos Trans R Soc Lond B Biol Sci* 364:2795–2808.
2. Armitage JP, Hellingwerf KJ (2003) Light-induced behavioral responses (;phototaxis') in prokaryotes. *Photosynth Res* 76:145–155.
3. Wilde A, Mullineaux CW (2017) Light-controlled motility in prokaryotes and the problem of directional light perception. *FEMS Microbiol Rev* 41:900–922.
4. Bhaya D (2004) Light matters: Phototaxis and signal transduction in unicellular cyanobacteria. *Mol Microbiol* 53:745–754.
5. Schuergers N, Mullineaux CW, Wilde A (2017) Cyanobacteria in motion. *Curr Opin Plant Biol* 37:109–115.
6. Wadhams GH, Armitage JP (2004) Making sense of it all: Bacterial chemotaxis. *Nat Rev Mol Cell Biol* 5:1024–1037.
7. Sourjik V, Wingreen NS (2012) Responding to chemical gradients: Bacterial chemotaxis. *Curr Opin Cell Biol* 24:262–268.
8. Parkinson JS, Hazelbauer GL, Falke JJ (2015) Signaling and sensory adaptation in *Escherichia coli* chemoreceptors: 2015 update. *Trends Microbiol* 23:257–266.
9. Bi S, Sourjik V (2018) Stimulus sensing and signal processing in bacterial chemotaxis. *Curr Opin Microbiol* 45:22–29.
10. Yoshihara S, Katayama M, Geng X, Ikeuchi M (2004) Cyanobacterial phytochrome-like PixJ1 holoprotein shows novel reversible photoconversion between blue- and green-absorbing forms. *Plant Cell Physiol* 45:1729–1737.
11. Bhaya D, Takahashi A, Grossman AR (2001) Light regulation of type IV pilus-dependent motility by chemosensor-like elements in *Synechocystis* PCC6803. *Proc Natl Acad Sci USA* 98:7540–7545.
12. Rockwell NC, Martin SS, Feoktistova K, Lagarias JC (2011) Diverse two-cysteine photocycles in phytochromes and cyanobacteriochromes. *Proc Natl Acad Sci USA* 108:11854–11859.
13. Hirose Y, Narikawa R, Katayama M, Ikeuchi M (2010) Cyanobacteriochrome CcaS regulates phycoerythrin accumulation in Nostoc punctiforme, a group II chromatic adapter. *Proc Natl Acad Sci USA* 107:8854–8859.
14. Rockwell NC, Martin SS, Gulevich AG, Lagarias JC (2012) Phycoviolobilin formation and spectral tuning in the DXCF cyanobacteriochrome subfamily. *Biochemistry* 51:1449–1463.
15. Rockwell NC, Martin SS, Lagarias JC (2012) Red/green cyanobacteriochromes: Sensors of color and power. *Biochemistry* 51:9667–9677.
16. Ikeuchi M, Ishizuka T (2008) Cyanobacteriochromes: A new superfamily of tetrapyrrole-binding photoreceptors in cyanobacteria. *Photochem Photobiol Sci* 7:1159–1167.
17. Yoshihara S, Suzuki F, Fujita H, Geng XX, Ikeuchi M (2000) Novel putative photoreceptor and regulatory genes required for the positive phototactic movement of the unicellular motile cyanobacterium *Synechocystis* sp. PCC 6803. *Plant Cell Physiol* 41:1299–1304.
18. Ishizuka T, et al. (2006) Characterization of cyanobacteriochrome TePixJ from a thermophilic cyanobacterium *Thermosynechococcus elongatus* strain BP-1. *Plant Cell Physiol* 47:1251–1261.
19. Ishizuka T, Narikawa R, Kohchi T, Katayama M, Ikeuchi M (2007) Cyanobacteriochrome TePixJ of *Thermosynechococcus elongatus* harbors phycoviolobilin as a chromophore. *Plant Cell Physiol* 48:1385–1390.
20. Okajima K, et al. (2005) Biochemical and functional characterization of BLUF-type flavin-binding proteins of two species of cyanobacteria. *J Biochem* 137:741–750.
21. Sugimoto Y, Nakamura H, Ren S, Hori K, Masuda S (2017) Genetics of the blue light-dependent signal cascade that controls phototaxis in the cyanobacterium *Synechocystis* sp. PCC6803. *Plant Cell Physiol* 58:458–465.
22. Narikawa R, et al. (2011) Novel photosensory two-component system (PixA-NixB-NixC) involved in the regulation of positive and negative phototaxis of cyanobacterium *Synechocystis* sp. PCC 6803. *Plant Cell Physiol* 52:2214–2224.
23. Song J-Y, et al. (2011) Near-UV cyanobacteriochrome signaling system elicits negative phototaxis in the cyanobacterium *Synechocystis* sp. PCC 6803. *Proc Natl Acad Sci USA* 108:10780–10785.
24. Fiedler B, Börner T, Wilde A (2005) Phototaxis in the cyanobacterium *Synechocystis* sp. PCC 6803: Role of different photoreceptors. *Photochem Photobiol* 81:1481–1488.
25. Wilde A, Fiedler B, Börner T (2002) The cyanobacterial phytochrome Cph2 inhibits phototaxis towards blue light. *Mol Microbiol* 44:981–988.
26. Savakis P, et al. (2012) Light-induced alteration of c-di-GMP level controls motility of *Synechocystis* sp. PCC 6803. *Mol Microbiol* 85:239–251.
27. Schuergers N, et al. (2016) Cyanobacteria use micro-optics to sense light direction. *eLife* 5:e12620.
28. Cohen SE, Golden SS (2015) Circadian rhythms in cyanobacteria. *Microbiol Mol Biol Rev* 79:373–385.
29. Golden SS (1995) Light-responsive gene expression in cyanobacteria. *J Bacteriol* 177:1651–1654.
30. Bustos SA, Golden SS (1992) Light-regulated expression of the psbD gene family in Synechococcus sp. strain PCC 7942: Evidence for the role of duplicated psbD genes in cyanobacteria. *Mol Gen Genet* 232:221–230.
31. Angermayr SA, Gorchs Rovira A, Hellingwerf KJ (2015) Metabolic engineering of cyanobacteria for the synthesis of commodity products. *Trends Biotechnol* 33:352–361.
32. Schatz D, et al. (2013) Self-suppression of biofilm formation in the cyanobacterium *Synechococcus elongatus*. *Environ Microbiol* 15:1786–1794.
33. Parnasa R, et al. (2016) Small secreted proteins enable biofilm development in the cyanobacterium *Synechococcus elongatus*. *Sci Rep* 6:32209.
34. Kondou Y, Nakazawa M, Higashi S, Watanabe M, Manabe K (2001) Equal-quantum action spectra indicate fluence-rate-selective action of multiple photoreceptors for photomovement of the thermophilic cyanobacterium *Synechococcus elongatus*. *Photochem Photobiol* 73:90–95.
35. Welkie DG, et al. (December 5, 2018) A hard day's night: Cyanobacteria in diel cycles. *Trends Microbiol*, 10.1016/j.tim.2018.11.002.
36. Kratz WA, Jack M (1955) Nutrition and growth of several blue-green algae. *Am J Bot* 42:282–287.
37. Herdman M, Castenholz RW, Waterbury JB, Rosmarie R (2001) Form-genus XIII. *Synechococcus. Bergey's Manual of Systematic Bacteriology*, eds Boone DR, Castenholz RW, Garrity GM (Springer, New York), 2nd Ed, pp 544–546.
38. Sugita C, et al. (2007) Complete nucleotide sequence of the freshwater unicellular cyanobacterium *Synechococcus elongatus* PCC 6301 chromosome: Gene content and organization. *Photosynth Res* 93:55–67.
39. Burriesci M, Bhaya D (2008) Tracking phototactic responses and modeling motility of *Synechocystis* sp. strain PCC6803. *J Photochem Photobiol B* 91:77–86.
40. Varuni P, Menon SN, Menon GI (2017) Phototaxis as a collective phenomenon in cyanobacterial colonies. *Sci Rep* 7:17799.
41. Holtman CK, et al. (2005) High-throughput functional analysis of the *Synechococcus elongatus* PCC 7942 genome. *DNA Res* 12:103–115.
42. Chen Y, Holtman CK, Taton A, Golden SS (2012) Functional analysis of the Synechococcus elongatus PCC 7942 Genome. *Functional Genomics and Evolution of Photosynthetic Systems, Advances in Photosynthesis and Respiration* (Springer, Dordrecht), pp 119–137.
43. Mackey SR, Ditty JL, Clerico EM, Golden SS (2007) Detection of rhythmic bioluminescence from luciferase reporters in cyanobacteria. *Methods Mol Biol* 362:115–129.
44. Ishizuka T, et al. (2011) The cyanobacteriochrome, TePixJ, isomerizes its own chromophore by converting phycocyanobilin to phycoviolobilin. *Biochemistry* 50:953–961.
45. Rockwell NC, Martin SS, Lagarias JC (2012) Mechanistic insight into the photosensory versatility of DXCF cyanobacteriochromes. *Biochemistry* 51:3576–3585.
46. Berkelman TR, Lagarias JC (1986) Visualization of bilin-linked peptides and proteins in polyacrylamide gels. *Anal Biochem* 156:194–201.
47. Gambetta GA, Lagarias JC (2001) Genetic engineering of phytochrome biosynthesis in bacteria. *Proc Natl Acad Sci USA* 98:10566–10571.
48. Fushimi K, et al. (2016) Cyanobacteriochrome photoreceptors lacking the canonical Cys residue. *Biochemistry* 55:6981–6995.
49. Rockwell NC, et al. (2008) A second conserved GAF domain cysteine is required for the blue/green photoreversibility of cyanobacteriochrome Tlr0924 from *Thermosynechococcus elongatus*. *Bioche*$^{\mathit{mistry}}$ 47:7304–7316.
50. Taton A, Ma AT, Ota M, Golden SS, Golden JW (2017) NOT gate genetic circuits to control gene expression in cyanobacteria. *ACS Synth Biol* 6:2175–2182.

51. Liu PY, et al. (2014) An optofluidic imaging system to measure the biophysical signature of single waterborne bacteria. *Lab Chip* 14:4237–4243.
52. Whitman WB (2015) Insights into the life of an oxygenic phototroph. *Proc Natl Acad Sci USA* 112:14747–14748.
53. Yang Y, Sourjik V (2012) Opposite responses by different chemoreceptors set a tunable preference point in *Escherichia coli* pH taxis. *Mol Microbiol* 86:1482–1489.
54. Paulick A, et al. (2017) Mechanism of bidirectional thermotaxis in *Escherichia coli*. *eLife* 6:e26607.
55. Ng W-O, Grossman AR, Bhaya D (2003) Multiple light inputs control phototaxis in *Synechocystis* sp. strain PCC6803. *J Bacteriol* 185:1599–1607.
56. Campbell EL, et al. (2015) Genetic analysis reveals the identity of the photoreceptor for phototaxis in hormogonium filaments of *Nostoc punctiforme*. *J Bacteriol* 197: 782–791.
57. Gestwicki JE, et al. (2000) Evolutionary conservation of methyl-accepting chemotaxis protein location in bacteria and archaea. *J Bacteriol* 182:6499–6502.
58. Kondou Y, et al. (2002) Bipolar localization of putative photoreceptor protein for phototaxis in thermophilic cyanobacterium *Synechococcus elongatus*. *Plant Cell Physiol* 43:1585–1588.
59. Boyer SL, Flechtner VR, Johansen JR (2001) Is the 16S-23S rRNA internal transcribed spacer region a good tool for use in molecular systematics and population genetics? A case study in cyanobacteria. *Mol Biol Evol* 18:1057–1069.
60. Koren S, et al. (2017) Canu: Scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. *Genome Res* 27:722–736.
61. Bankevich A, et al. (2012) SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19:455–477.
62. Walker BJ, et al. (2014) Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9:e112963.
63. Hunt M, et al. (2015) Circlator: Automated circularization of genome assemblies using long sequencing reads. *Genome Biol* 16:294.
64. Huntemann M, et al. (2015) The standard operating procedure of the DOE-JGI Microbial Genome Annotation Pipeline (MGAP v.4). *Stand Genomic Sci* 10:86.

2.2 Supplemental Material

PNAS
www.pnas.org

Supplementary Information for

Phototaxis in a wild isolate of the cyanobacterium *Synechococcus elongatus*

Yiling Yang, Vinson Lam, Marie Adomako, Ryan Simkovsky, Annik Jakob, Nathan C. Rockwell, Susan E. Cohen, Arnaud Taton, Jingtong Wang, J. Clark Lagarias, Annegret Wilde, David R. Nobles, Jerry J. Brand, and Susan S. Golden

Susan S. Golden
Email: sgolden@ucsd.edu

This PDF file includes:

Supplementary text
Figs. S1 to S15
Tables S1 to S5
Captions for movies S1 to S8
References for SI reference citations

Other supplementary materials for this manuscript include the following:

Movies S1 to S8

www.pnas.org/cgi/doi/10.1073/pnas.1812871115

SI materials and methods

Plasmid and strain construction

All plasmids and strains are listed in Tables S3, S4 and S5. Transposon insertion mutants were constructed by homologous recombination following transformation of cyanobacterial strains with gene-specific cosmids from the *S. elongatus* PCC 7942 unigene set (UGS) library (1, 2). We designed plasmids for specific gene replacement as described elsewhere (3) using the CYANO_VECTOR server (http://golden.ucsd.edu /CyanoVECTOR/). Constructs for complementation of mutants were made by amplifying the target gene and inserting the fragment into the vector pAM5431 at a SwaI site; the resulting plasmid enables gene expression from genome Neutral Site I (NS1) under control of the P*trc* promoter. A YFP-tagged PixJ fusion was constructed by inserting *yfp* into pAM5477 with sequences that encode a GSGGG linker. All DNA fragments were assembled using an Invitrogen GeneArt seamless cloning kit (Thermo Fisher, Carlsbad, CA). Plasmid cloning was carried out in *Escherichia coli* strain DH5α using standard techniques.

Biofilm formation

S. elongatus strains grown on BG-11 agar plates were collected and used to inoculate 15-ml starter cultures grown in BG-11R (BG-11 with fresh iron and HEPES, as described in (4)) in 125-ml glass culture flasks. These starter cultures were grown shaking for 3 - 4 days at 30 °C under 150 µmol photons $m^{-2}s^{-1}$ illumination from fluorescent lights. Starter cultures were then diluted to OD 0.5 with BG-11R and 5 ml of this dilution was distributed to 25-ml glass culture flasks and placed in a stationary low-light (20 - 30 µmol photons $m^{-2}s^{-1}$) box at room temperature for 7 days to allow biofilm formation to occur. To assess biofilm formation, liquid cultures were slowly decanted, gently washed with water to remove unattached cells, stained with 5 ml of a 1% crystal violet solution for 15 min, washed again with water, and then allowed to dry.

Motility at different times of day

S. elongatus UTEX 3055 cells were cultured on BG-11 agar plates and entrained for two days in a 12-h light:12-h dark cycle before being released into constant light conditions at the end of the second dark period. Cells were harvested from the plate at 0.5, 12.5, 24.5, and 36.5 h after release into constant light and resuspended in fresh BG-11 medium. The cell suspension was flowed into a small chamber made from taped coverslips and cells were allowed to settle. Suspended cells were removed by flowing in fresh BG-11 medium. Cell motility was observed on a Nikon TE300 inverted microscope at 40x magnification, with brightfield illumination provided by infrared LED at 850 nm wavelength. Cells were illuminated with oblique white-LED light. Images were acquired at 1-s intervals for 20 min. Cell tracking was performed using the Oufti software package (5). Cells that moved slower than 0.03 μm/s were regarded as non-motile and discarded from analysis.

Circadian bioluminescence monitoring

As described previously (6), *S. elongatus* strains expressing a P_{kaiBC}-*luc* reporter were grown at 30 °C for two cycles of 12-h light: 12-h dark to synchronize the population before transfer to constant-light conditions, during which bioluminescence was recorded every two hours. UTEX 3055 strains expressing the reporter gene from different neutral sites both showed bioluminescence rhythms similar to the corresponding PCC 7942 strains, with a period of 25 ± 0.4 h. Data were analyzed with the Biological Rhythms Analysis Software System (http://millar.bio.ed.ac.uk /PEBrown/BRASS/BrassPage.htm).

Immunoblot analysis

Equal amounts of total protein (5 μg) from each sample extract were separated by SDS-PAGE (AnykD, BIORAD), transferred to a polyvinylidene difluoride (PVDF) membrane, and blocked with 2.5% w/v nonfat dry milk / Tris Buffered Saline + 0.1 % Tween-20 (TBST). Membranes were incubated with α-GFP (mouse, Abgent) at 1:10,000 in 2.5% non-fat milk in TBST for 2 h, followed by five washes in TBST. Membranes were then incubated with horseradish peroxidase (HRP)-conjugated goat anti-mouse IgG secondary antibody (Thermo Scientific). Chemiluminescent detection was performed using Pierce SuperSignal West Femto detection reagents (Thermo Scientific).

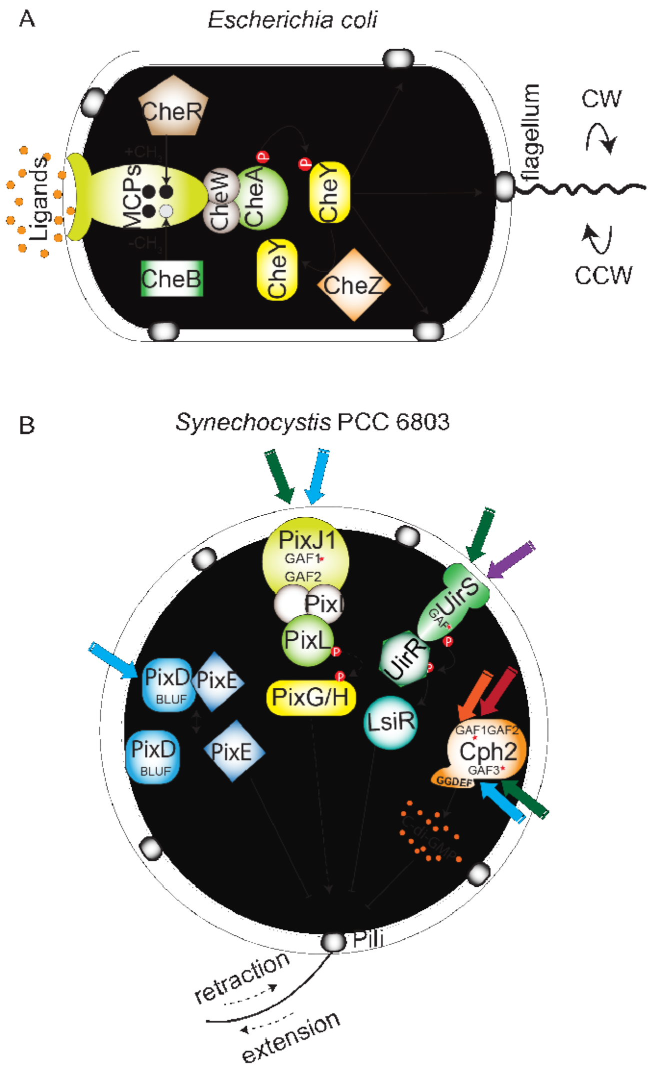


Fig. S1. Schematic illustration of chemotaxis signaling pathway in *E. coli* (A) and phototaxis pathways in *Synechocystis* PCC 6803 (B). *Synechocystis* employs homologs of *E. coli* chemotaxis proteins for sensing and regulation of phototaxis: MCP (PixJ1), CheA (PixL), CheW (PixI), and CheY (PixG/H). Adaptation proteins CheR and CheB, as well as phosphatase CheZ, are not encoded by cyanobacterial genomes. In addition to the Che-like pathway that senses blue and green light, *Synechocystis* contains other systems that control phototactic behavior. Notably, cyanobacteria move over solid surfaces through extension and retraction of type-IV pili that are distributed around the cell exterior (7, 8), whereas *E. coli* swim in liquid environments using flagella that rotate either clockwise (CW) or counterclockwise (CCW).

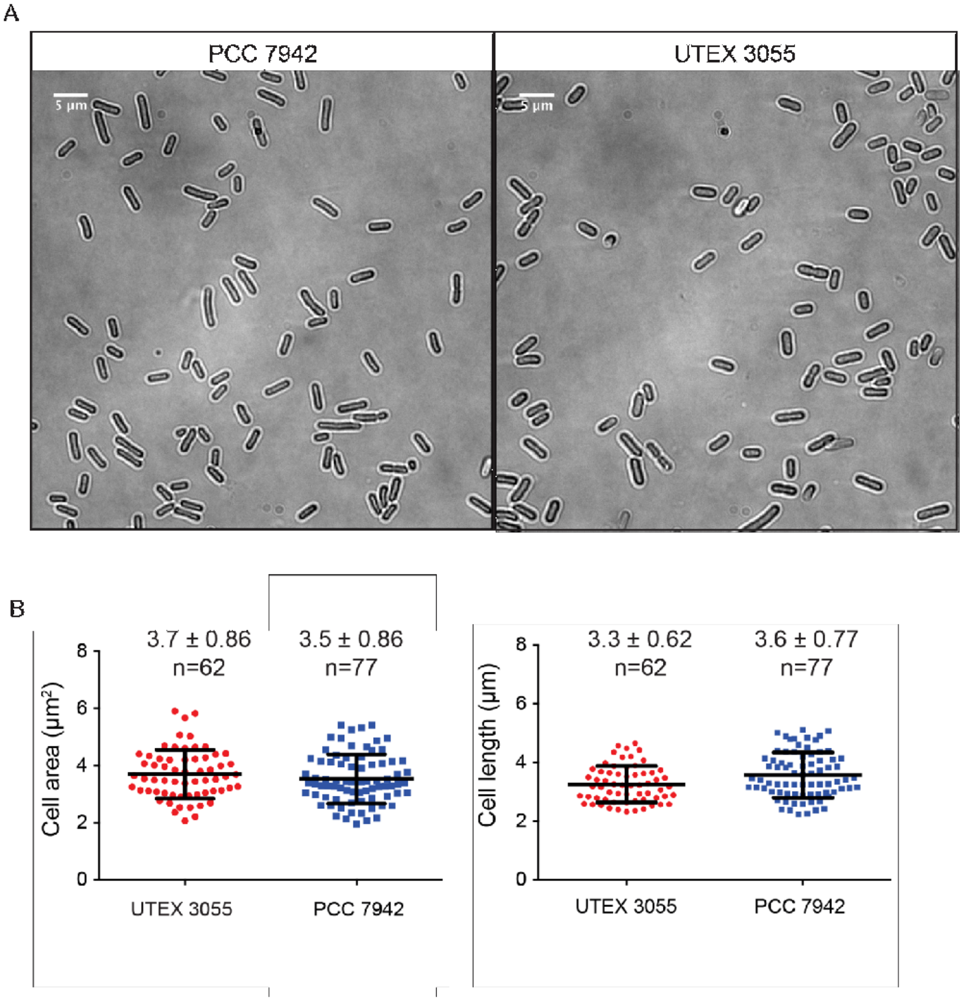


Fig. S2. Similar morphology of *S. elongatus* PCC 7942 and UTEX 3055. (A) Brightfield images of PCC 7942 and UTEX 3055. (B) Cell length and cell area of UTEX 3055 and PCC 7942 quantified from cells in (A). Bar represents mean with standard deviation (SD) and individual measurements are indicated by dots.

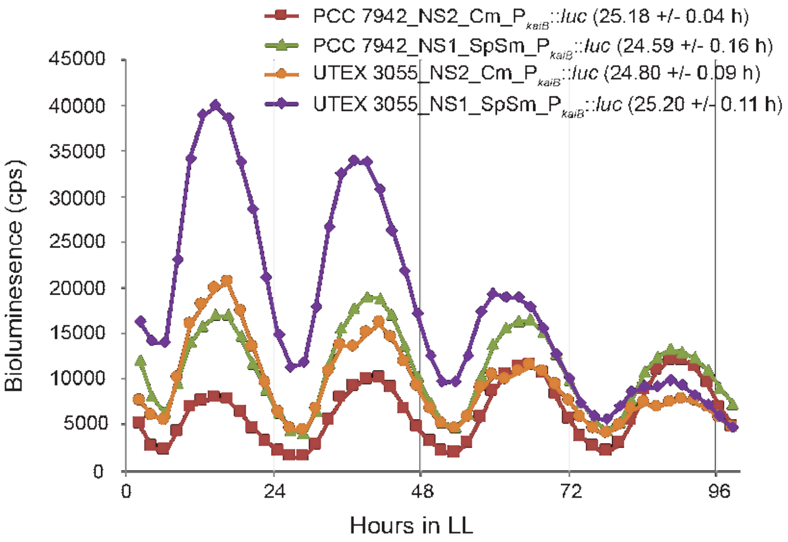


Fig. S3. Circadian rhythms of gene expression in *S. elongatus* PCC 7942 and UTEX 3055. Bioluminescence from strains carrying a P_{kaiB}-*luc* reporter at NS1 or NS2 was recorded as an assay for circadian rhythms of gene expression. The circadian period and standard error of the mean of each strain is indicated. LL, constant light after entrainment in a 12-h light:12-h dark cycle.

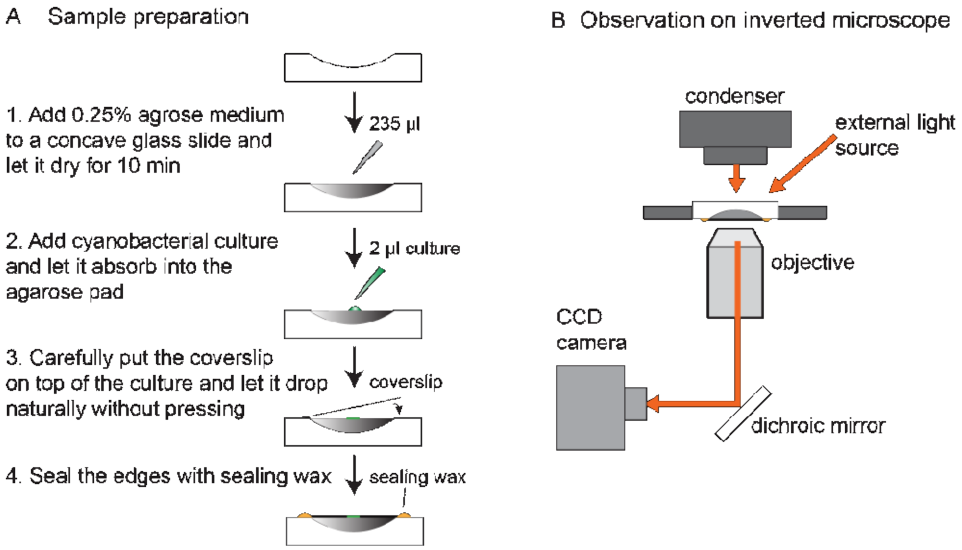

A Sample preparation
1. Add 0.25% agrose medium to a concave glass slide and let it dry for 10 min
235 µl
2. Add cyanobacterial culture and let it absorb into the agarose pad
2 µl culture
3. Carefully put the coverslip on top of the culture and let it drop naturally without pressing
coverslip
4. Seal the edges with sealing wax
sealing wax
B Observation on inverted microscope
condenser
external light source
objective
CCD camera
dichroic mirror

Fig. S4. Microscopic observation of cyanobacterial phototaxis. (A) Steps for cell sample preparation. (B) Observation of cell movement with inverted microscope under directional light provided by external light source or condenser light.

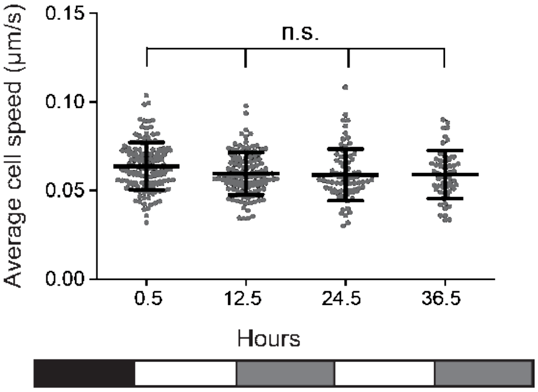


Fig. S5. Motility at different times of day. UTEX 3055 cell motility was monitored at the indicated time points in constant light following two days of entrainment in 12-h light:12-h dark cycles. The average cell speeds did not differ significantly at different time points. Dark box: dark night time; white box: subjective day; grey box: subjective night (circadian night in a light environment). 56-133 cells were measured at each time point; bar represents mean with SD. n.s.: not significant according to a one-way ANOVA test ($p > 0.05$).

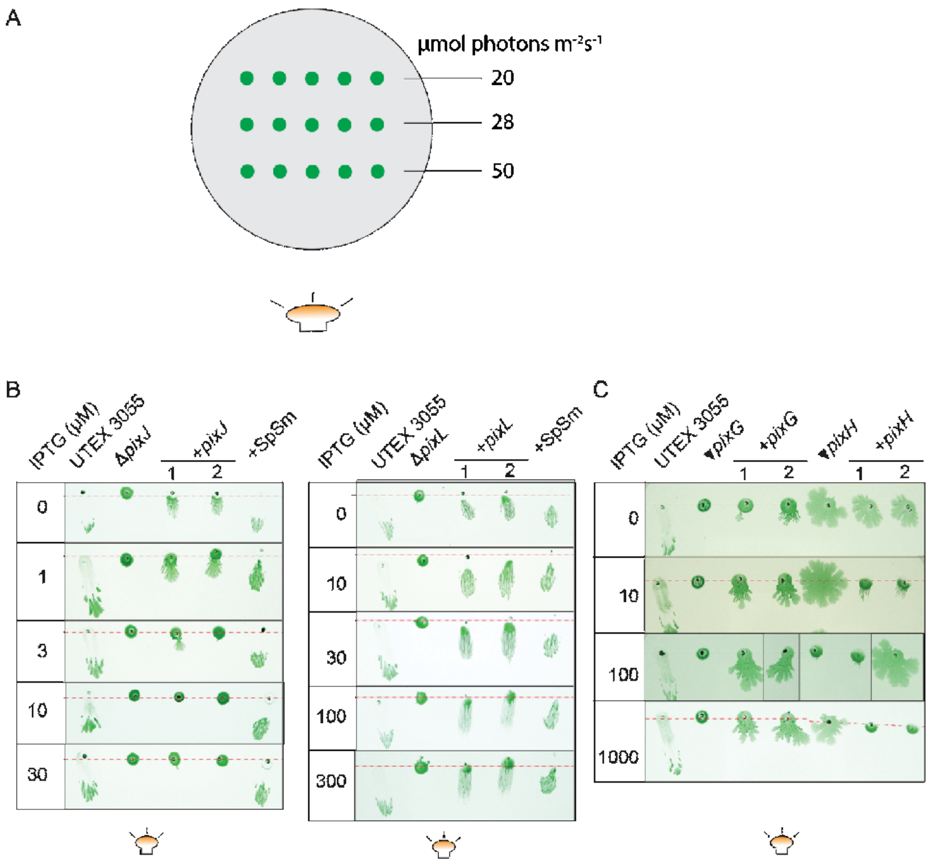


Fig. S6. Complementation of the phototaxis mutants. (A) Phototaxis assay setup. Green dots represent inoculation spots of *S. elongatus* cells, which were placed at different distances to the lateral light source. The fluence rate at each distance is noted. (B, C) Complementation of mutants *pixJ*, *pixL* (B) and *pixG*, *pixH* (C) by the respective genes under indicated level of IPTG induction. Two independent transformants were tested for each complementation assay. Empty vector expressing SpSm-resistance gene was introduced into UTEX 3055 at NS1 as a control. Red-dashed line indicates the starting position. Experiment was performed as in (A) and the representative results at 28 µmol photons m^{-2}s^{-1} were presented.

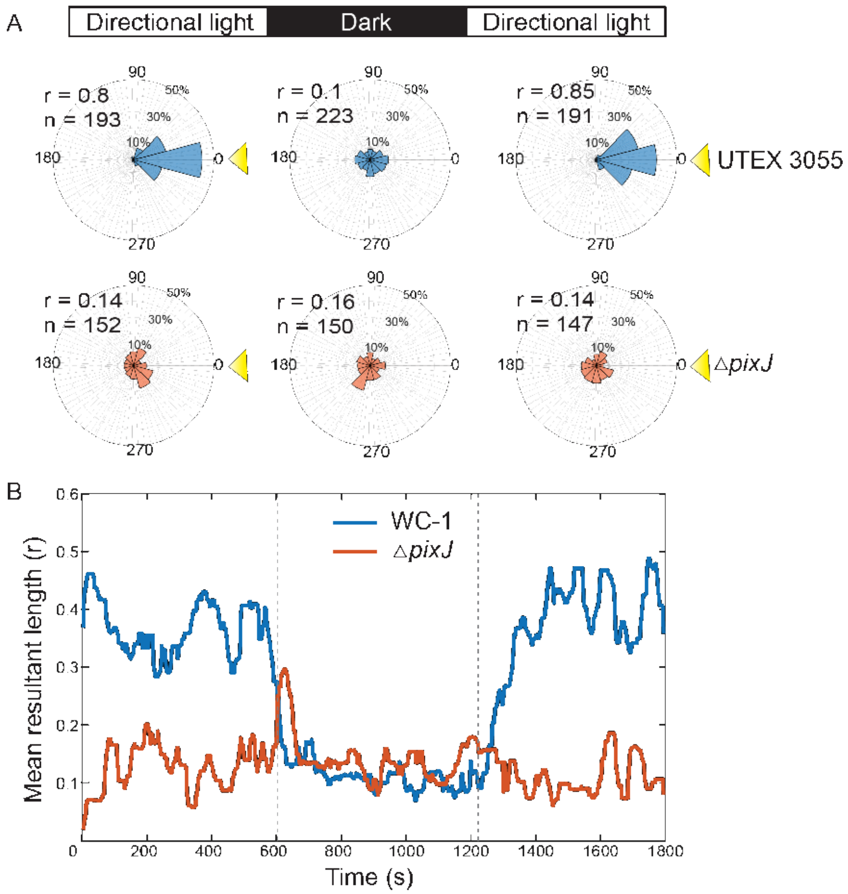


Fig. S7. Single-cell movement under lateral illumination and in the dark. Cell movement under directional illumination with a period of darkness for wild-type UTEX 3055 (blue) and Δ*pixJ* (orange). (A) Fraction of cells moving in a certain direction was quantified and plotted. (B) Mean resultant length 'r' from a Rayleigh test over time analyzed as in Fig. 1E. Dashed line indicates the time points of turning the lights off and on.

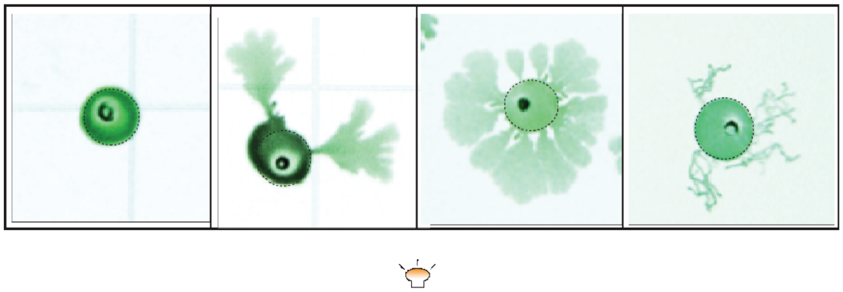


Fig. S8. Representative images of non-phototactic Δ*pixJ* cells on phototaxis plates. The appearance of the cell spots varied, even at different positions of the same assay plate. However, all samples showed loss of directional movement under a lateral light source. Light is coming in from the side indicated by a light bulb.

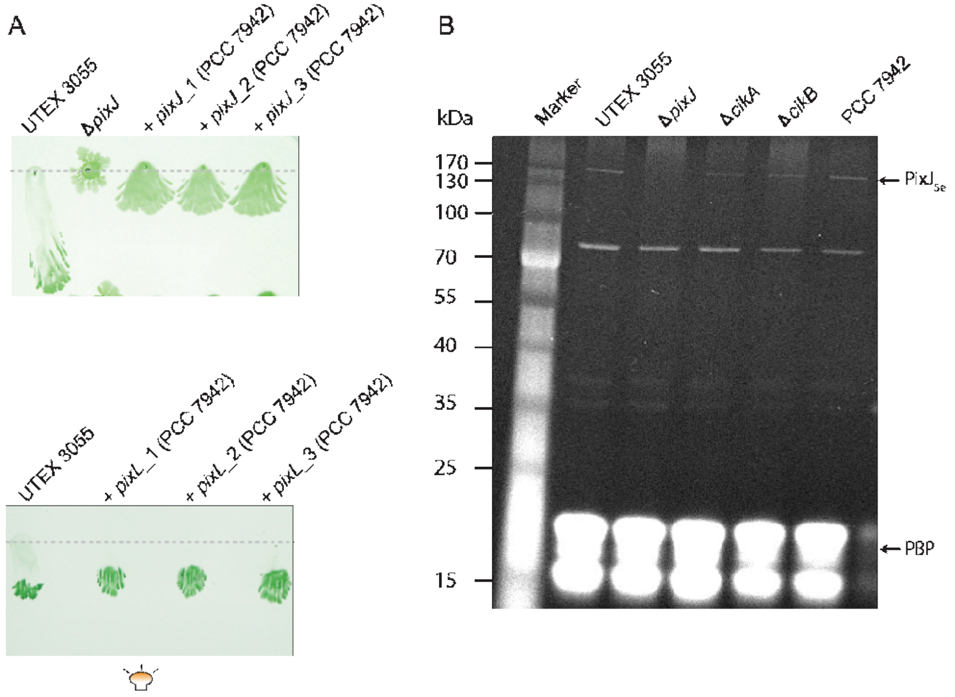


Fig. S9. Homologous proteins from PCC 7942 are functional in the respective UTEX 3055 mutants. (A) Expression of *pixJ* and *pixL* from PCC 7942 restored phototaxis to respective mutants of UTEX 3055. Three independent transformants marked as _1-3 were tested. A dashed line marks the point of inoculation. (B) Whole cell lysates of UTEX 3055, Δ*pix*J, Δ*cik*A, Δ*cik*B and PCC 7942 were separated by SDS-PAGE and tested for fluorescence in a zinc blot assay (9). The clock proteins *cikA* and *cikB* do not affect phototaxis but are included here to exclude the possibility that either is responsible for the 80 kDa zinc-reactive band. CikA and CikB are both approximately the same size as the PixJ band and each carries a GAF domain, but neither GAF is predicted to bind a bilin chromophore. Arrows indicate positions of PixJ$_{Se}$ and phycobiliproteins (PBP). This result showed that PCC 7942 expresses a bilin-binding PixJ homolog at a similar level as that of UTEX 3055.

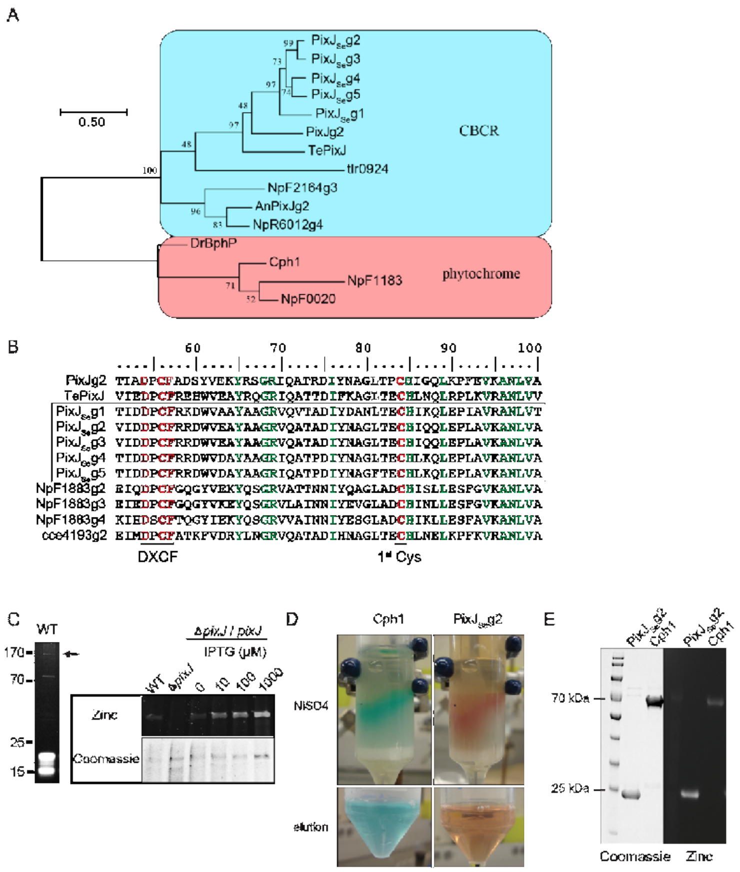
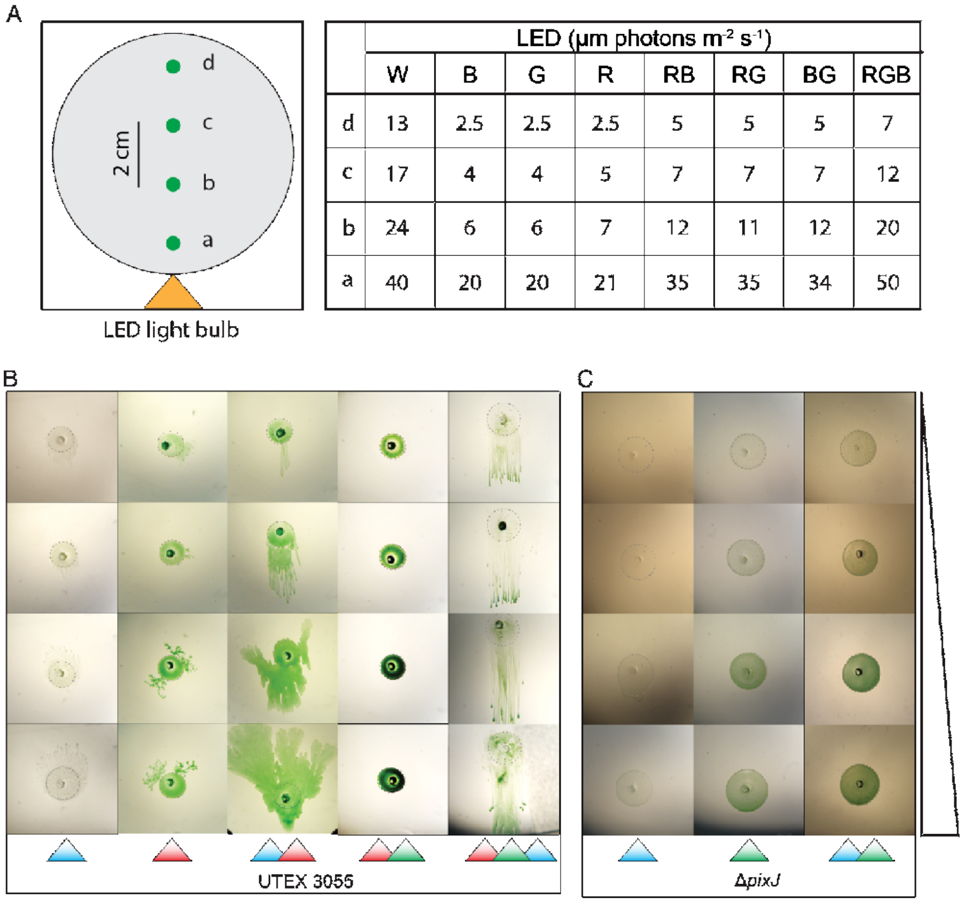

A
0.50
PixJ$_{Se}$g2
PixJ$_{Se}$g3
PixJ$_{Se}$g4
PixJ$_{Se}$g5
PixJ$_{Se}$g1
PixJg2
TePixJ
tlr0924
NpF2164g3
AnPixJg2
NpR6012g4
DrBphP
Cph1
NpF1183
NpF0020
CBCR
phytochrome
99
73
97
75
48
97
48
100
96
83
71
52
B
60
70
80
90
100
PixJg2 TIADPCFADSYVEKYRSGRIQATRDIYNAGLTPCHIGQLKPFEVKANLVA
TePixJ VIEDPCFREEWVEAYRQGRIQATTDIFKAGLTECHLNQLRPLKVRANLVV
PixJ$_{Se}$g1 TIDDPCFRKDWVAAYAAGRVQVTADIYDANLTECHIKQLEPIAVRANLVT
PixJ$_{Se}$g2 VIDDPCFRRDWVEAYAAGRVQATADIYNAGLTECHIQQLEPLAVKANLVA
PixJ$_{Se}$g3 VIDDPCFRRDWVEAYAAGRVQATADIYNAGLTECHIQQLEPLAVKANLVA
PixJ$_{Se}$g4 TIDDPCFRRDWVDAYAAGRIQATPDIYNAGLTECHLKQLEPLAVKANLVA
PixJ$_{Se}$g5 TIDDPCFRRDWVDAYAAGRIQATPDIYNAGFTECHLKQLEPLAVKANLVA
NpF1883g2 EIQDPCFGQGYVEKYQSGRVATTNNIYQAGLADCHISLLESPGVKANLVA
NpF1883g3 EIEDPCFGQGYVKEYQSGRVLAINNIYEVGLADCHINLLESPGVKANLVA
NpF1883g4 KIEDSCFTQGYIEKYQSGRVVAINNIYESGLADCHIKLLESFAVKANLVA
cce4193g2 EIMDPCFATKFVDRYLNGRVQATADIHNAGLTECHLNELKPFKVRANLVA
DXCF
1st Cys
C
WT
170
70
25
15
ΔpixJ / pixJ
IPTG (μM)
WT
ΔpixJ
0
10
100
1000
Zinc
Coomassie
D
Cph1
PixJ$_{Se}$g2
NiSO4
elution
E
PixJ$_{Se}$g2
Cph1
PixJ$_{Se}$g2
Cph1
70 kDa
25 kDa
Coomassie
Zinc

Fig. S10. Classification of PixJ$_{Se}$ and purification of the second GAF domain. (A) Phylogenetic tree of GAF domains from PixJ$_{Se}$ and bilin-binding photoreceptors from other species: *Synechocystis* (Cph1 and PixJg2); *Thermosynechococcus elongatus* BP-1 (Te); *Anabaena sp.* PCC 7120 (An); *Nostoc punctiforme* ATCC 29133 (NpF/NpR); *Deinococcus radiodurans* (Dr). Phylogenetic analysis was performed using the maximum-likelihood method. Cyanobacteriochromes are shaded in blue and phytochromes are shaded in pink. Number of bootstrap replications is 500. Bootstrap values are shown. (B) All GAF domains in PixJ$_{Se}$ show high similarity to DXCF CBCR. GAF domain sequences of PixJ$_{Se}$ (g1-g5) were aligned to known DXCF CBCRs. Conserved Cys residues and "DXCF" module are highlighted in red and identical sequences are shown in green. *Cyanothece* sp. ATCC 5114 (cce4193g2). (C) Presence of bilin-bound protein detected by zinc-blot assay in UTEX 3055, Δ*pixJ* and complemented strain under indicated level of IPTG induction. (D) Nickel-affinity chromatography. Cell lysates containing Cph1(N514)-His or PixJ$_{Se}$GAF2-His in presence of phycocyanobilin (PCB) exhibits cyan or pink color, respectively, when passing through the nickel column. (E) SDS gel and zinc blot of purified Cph1(N514)-His and PixJ$_{Se}$GAF2-His protein obtained from (D).



	LED (μm photons m^{-2} s^{-1})							
	W	B	G	R	RB	RG	BG	RGB
d	13	2.5	2.5	2.5	5	5	5	7
c	17	4	4	5	7	7	7	12
b	24	6	6	7	12	11	12	20
a	40	20	20	21	35	35	34	50

Fig. S11. Phototactic response of UTEX 3055 to different colors of light. (A) Schematic drawing of the experimental design. Grey area represents a Petri dish filled with soft BG-11 medium and the green dots represent inoculation spots of *S. elongatus* cells. The Petri dish was placed in a black box with an LED bulb (clear white or 5-mm RGB controllable from microtivity) mounted on one sidewall indicated by an orange triangle. Cells were placed at different distances (a, b, c and d) from the light bulb and the total irradiance level at each position measured with a photometer (Biospherical Instrument QSL-100) is listed on the right for each lighting condition. W, white; R, red; G, red; B, blue. red, λ=630 nm, FWHM (full width at half maximum) =25 nm; blue, λ=465 nm, FWHM =25 nm; green λ=516 nm, FWHM=30 nm. (B) Phototactic migration of wild-type UTEX 3055 toward single color or additive colors of LED light as indicated. (C) Cells of a *pixJ* mutant do not show phototaxis to either individual or a combination of blue and green light. All

experiments were performed for at least three replicates and representative results are shown. Triangle at right indicates fluence gradient. Tree-like twisted branches formed under red light in panel A indicate non-directional movement as shown in Fig. S8.

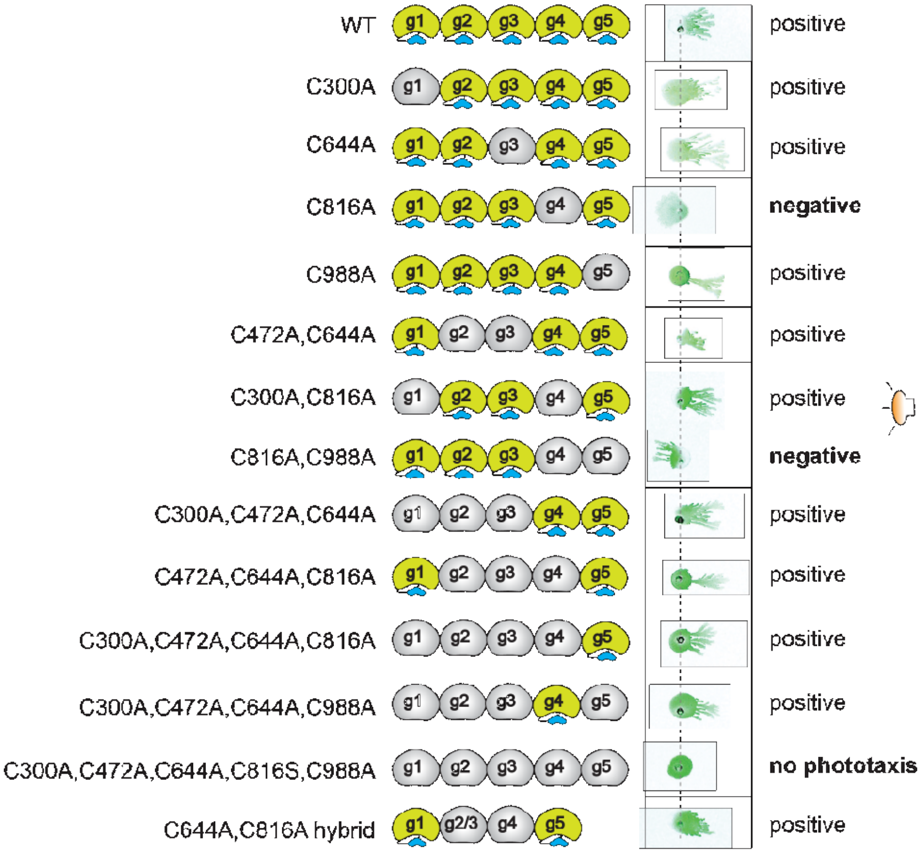


Fig. S12. Phototaxis phenotypes of all PixJ$_{Se}$ variants that carry Cys→ Ala mutations in specific GAF domains. Intact GAFs are shown in yellow and GAF domains with the first conserved Cys mutated to Ala are shown in gray. Representative image of phototaxis phenotype of each mutant is shown on the right. Images of cells taken from different plates are separated by solid black lines. Dashed line indicates the inoculation position of cells with directional white light provided from the right (See Fig. S6A for experimental setup). All assays were performed at least three times with the indicated outcomes.

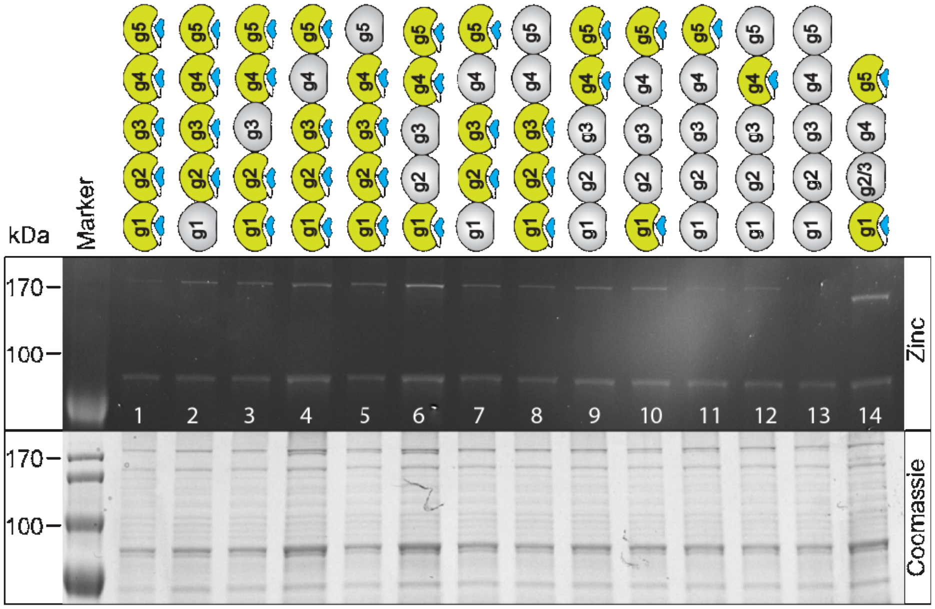


Fig. S13. Bilin-binding ability of PixJ$_{\mathrm{Se}}$ variants with substitution of Cys → Ala in different GAF domains. Lane 1: Wild-type UTEX 3055; lanes 2-14: UTEX 3055 *pixJ* mutant expressing PixJ$_{\mathrm{Se}}$ variants represented by schematic GAF domains with or without bilin. Note that the mutant in lane 14 expresses PixJ$_{\mathrm{Se}}$ with a lower molecular weight due to presence of only four GAF domains, in which the first half of GAF2 and the second half of GAF3 were fused together through a cloning artifact. Zinc stain of SDS-PAGE gel of proteins from lysates of indicated strains (upper panel). Coomassie staining of proteins in the same SDS-PAGE gel (lower panel).

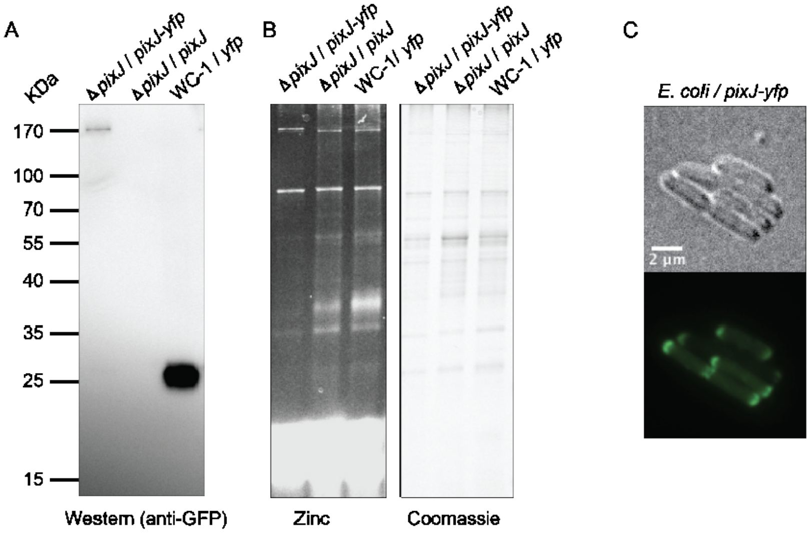


Fig. S14. Characterization of PixJ$_{Se}$-YFP. (A) Immunoblot performed with α-GFP as primary antibody. This result showed that a full-length PixJ$_{Se}$-YFP fusion (181 KDa) is expressed in UTEX 3055 and no protein degradation was detected. (B) Zinc blot shows PixJ$_{Se}$-YFP retains the ability to bind bilin. (C) Heterologously expressed PixJ$_{Se}$-YFP localized at cell poles of *E. coli* strain UU1581.

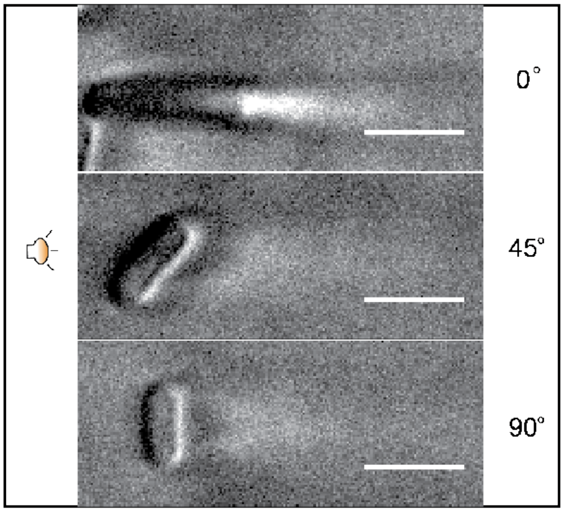


Fig. S15. Lensing effect of *S. elongatus* PCC 7942 cells at different orientations relative to the incident light direction. Cell sample was prepared as in Fig. S5 and imaged at 100x magnification. Scale bars = 3 µm.

Table S1. Genome information for *S. elongatus* PCC 7942 and UTEX 3055.

	PCC 7942			UTEX 3055		
	# base pairs	% GC	# ORFs	# base pairs	% GC	# ORFs
Chromosome	2,750,104	55	2621	2,767,524	55	2770
Large plasmid	46,366	53	50	89,249	51	97
PCC 7942 pANL	7,835	59	9			
UTEX 3055 plasmid				24,450	50	25

Table S2. Clock-related PCC 7942 genes identified in UTEX 3055.

PCC 7942	UTEX 3055 homolog	Annotation
synpcc7942_1218	UTEX 3055_1318	KaiA, circadian oscillator protein
synpcc7942_1217	UTEX 3055_1317	KaiB, circadian oscillator protein
synpcc7942_1216	UTEX 3055_1316	KaiC, circadian oscillator protein
synpcc7942_0095	UTEX 3055_0092	RpaA, two-component response regulator
synpcc7942_2114	UTEX 3055_2238	SasA, signal transduction histidine kinase
synpcc7942_0644	UTEX 3055_0779	CikA, GAF sensor signal transduction histidine kinase
synpcc7942_0480	UTEX 3055_0485	CikB, GAF sensor signal transduction histidine kinase
synpcc7942_0624	UTEX 3055_0759	LdpA, light-dependent period
synpcc7942_0677	UTEX 3055_0813	Pex, transcriptional regulator, PadR family
synpcc7942_0600	UTEX 3055_0735	PrkE, serine/threonine protein kinase
synpcc7942_1453	UTEX 3055_1555	RpaB, two-component response regulator, winged helix family
synpcc7942_1891	UTEX 3055_2009	LabA, uncharacterized conserved protein, 2C LabA/DUF88 family
synpcc7942_1168	UTEX 3055_1266	CpmA, circadian phase modifier
synpcc7942_2526	UTEX 3055_2679	ClpX, ATP-dependent Clp protease ATP-binding subunit
synpcc7942_2525	UTEX 3055_2678	ClpP, ATP-dependent Clp protease proteolytic subunit
synpcc7942_2160	UTEX 3055_2292	Nht1, alanine-glyoxylate aminotransferase apoenzyme
synpcc7942_2387	UTEX 3055_2539	IrcA, hypothetical protein / Cytochrome c
synpcc7942_1604	UTEX 3055_1596	CdpA, hypothetical protein
synpcc7942_1143	UTEX 3055_1241	LalA, hypothetical protein
Between synpcc7942_0095 and 0096	UTEX 3055_0093	Crm, circadian rhythm modulator

Table S3. Plasmids used in this study.

Plasmid	Description	Antibiotics	Source
8S23-L12	Tn*5*-insertion mutation of *pixI-2*	Km	(1)
8S23-E4	Tn*5*-insertion mutation of *pixG*	Km	(1)
8S23-H7	Tn*5*-insertion mutation of *pixH*	Km	(1)
8S42-B8	Tn*5*-insertion mutation of synpcc7942_1014	Km	(1)
8S26-H11	Tn*5*-insertion mutation of synpcc7942_1015	Km	(1)
8S26-O4	Tn*5*-insertion mutation of synpcc7942_1016	Km	(1)
pBAD-Cph1	P_{BAD} promoter, Cph1(N514)-producing plasmid	Ap	(11)
pKT271	P_{BAD} promoter, PCB-producing plasmid	Cm	(10)
pPL-PCB	$P_{lac/ara-1}$, PCB-producing plasmid	Km	(11)
pAM2105	P_{KaiB}-*luc* expressed from NS1	Cm	Lab collection
pAM2226	P_{KaiB}-*luc* expressed from NS2	SpSm	Lab collection
pAM4819	Cloning vector carrying *Aph*I cassette	Km	(3)
pAM4843	Cloning vector carrying an origin of replication for *E. coli.*	Ap	(3)
pAM4933	P_{conII}-yfp expressed from NS1	SpSm	(3)
pAM5431	P_{trc}-*SpSm-lacI-rrnB* expressed from NS1	SpSm	This work
pAM5472	PixG KO vector, derived from pAM4819 and pAM4843	Km	This work
pAM5473	PixH KO vector, derived from pAM4819 and pAM4843	Km	This work
pAM5474	Synpcc7942_2534 KO vector, derived from pAM4819 and pAM4843	Km	This work
pAM5475	P_{trc}-*pixL* (PCC 7942) expressed from NS1, derivative of pAM5431	SpSm	This work
pAM5476	P_{trc}-*pixJ* (PCC 7942) expressed from NS1, derivative of pAM5431	SpSm	This work
pAM5477	P_{trc}-*pixJ* expressed from NS1, derivative of pAM5431	SpSm	This work
pAM5478	P_{trc}-*pixL* expressed from NS1, derivative of pAM5431	SpSm	This work
pAM5479	P_{trc}-*pixG* expressed from NS1, derivative of pAM5431	SpSm	This work
pAM5480	P_{trc}-*pixH* expressed from NS1, derivative of pAM5431	SpSm	This work
pAM5481	PixJ KO vector, derived from pAM4819 and pAM4843	Km	This work
pAM5482	PixL KO vector, derived from pAM4819 and pAM4843	Km	This work
pAM 5483	PixI-1 KO vector, derived from pAM4819 and pAM4843	Km	This work
pAM5484	P_{trc}-*pixJ*^{C300A} expressed from NS1, derivative of pAM5477	SpSm	This work
pAM5485	P_{trc}-*pixJ*^{C644A} expressed from NS1, derivative of pAM5477	SpSm	This work
pAM5486	P_{trc}-*pixJ*^{C816A} expressed from NS1, derivative of pAM5477	SpSm	This work
pAM5487	P_{trc}-*pixJ*^{C988A} expressed from NS1, derivative of pAM5477	SpSm	This work
pAM5488	P_{trc}-*pixJ*$^{C472A, C644A}$ expressed from NS1, derivative of pAM5477	SpSm	This work
pAM5489	P_{trc}-*pixJ*$^{C300A, C816A}$ expressed from NS1, derivative of pAM5477	SpSm	This work
pAM5490	P_{trc}-*pixJ*$^{C816A, C988A}$ expressed from NS1, derivative of pAM5477	SpSm	This work
pAM5491	P_{trc}-*pixJ*$^{C300A, C472A, C644A}$ expressed from NS1, derivative of AM5477	SpSm	This work
pAM5492	P_{trc}-*pixJ*$^{C472A, C644A, C816A}$ expressed from NS1, derivative of AM5477	SpSm	This work
pAM5493	P_{trc}-*pixJ*$^{C300A, C472A, C644A, C816A}$ expressed from NS1, derivative of AM5477	SpSm	This work
pAM5494	P_{trc}-*pixJ*$^{C300A, C472A, C6444A, C988A}$ expressed from NS1, derivative of AM5477	SpSm	This work
pAM5495	P_{trc}-*pixJ*$^{C300A, C472A, C644A, C816A, C988A}$ expressed from NS1, derivative of AM5477	SpSm	This work
pAM5496	P_{trc}-*pixJ*$^{C644A, C816A}$ (GAF2,3 hybrid) expressed from NS1, derivative of AM5477	SpSm	This work
pAM5497	P_{trc}-*pixJ*-yfp expressed from NS1, derivative of AM5477	SpSm	This work
pAM5498	P_{BAD} promoter, PixJ$_{Se}$GAF2 producing plasmid	Ap	This work

Table S4. Cyanobacterial strains used in this study.

Strain	Genotype	Antibiotics	Source
AMC06	WT *S. elongatus* PCC 7942		Lab collection
AMC2388	WT *S. elongatus* UTEX 3055		This work
AMC2450	UTEX 3055_1014 (UGS 23G8)	Km	This work
AMC2492	*pixG*::Tn5 (UGS 22C11)	Km	This work
AMC2493	*pixH*::Tn5 (UGS 22C12)	Km	This work
AMC2494	Δ*pixI-1*::Km	Km	This work
AMC2495	*pixJ* (UGS 22D2)	Km	This work
AMC2496	Δ*pixJ*::Km	Km	This work
AMC2497	*pixL* with (UGS 22D3)	Km	This work
AMC2498	Δ*pixL*::Km	Km	This work
AMC2499	*pixI*-2 (UGS 22D4)	Km	This work
AMC2501	UTEX 3055_1015 (UGS 23G9)	Km	This work
AMC2502	AMC2496 (Δ*pixJ*::Km) and *pixJ* in NS1	Km SpSm	This work
AMC2503	AMC2496 with *pixJ*_C300A in NS1	Km SpSm	This work
AMC2504	AMC2496 with *pixJ*_C644A in NS1	Km SpSm	This work
AMC2505	AMC2496 with *pixJ*_C816A in NS1	Km SpSm	This work
AMC2506	AMC2496 with *pixJ*_C988A in NS1	Km SpSm	This work
AMC2507	AMC2496 with *pixJ*_C472A, C644A in NS1	Km SpSm	This work
AMC2508	AMC2496 with *pixJ*_C300A, C816A in NS1	Km SpSm	This work
AMC2509	AMC2496 with *pixJ*_C816A, C988A in NS1	Km SpSm	This work
AMC2510	AMC2496 with *pixJ*_C300A, C472A, C644A in NS1	Km SpSm	This work
AMC2511	AMC2496 with *pixJ*_C472A, C644A, C816A in NS1	Km SpSm	This work
AMC2512	AMC2496 with *pixJ*_C300A, C472A, C644A, C816A in NS1	Km SpSm	This work
AMC2513	AMC2496 with *pixJ*_C300A, C472A, C644A, C988A in NS1	Km SpSm	This work
AMC2514	AMC2496 with *pixJ*_C300A, C472A, C644A, C816A, C988A in NS1	Km SpSm	This work
AMC2515	AMC2496 with *pixJ*_C644A, C816A (GAF2, 3 hybrid) in NS1	Km SpSm	This work
AMC2516	AMC2496 with P$_{trc}$_*pixJ*_*yfp* in NS1	Km SpSm	This work
AMC2517	AMC2496 with *pixJ* from PCC 7942	Km SpSm	This work
AMC2518	AMC2496 with *pixL* from PCC 7942	Km SpSm	This work
AMC2519	AMC2497 with *pixL* in NS1	Km SpSm	This work
AM2520	P$_{kaiB}$::*luc* in NS1	SpSm	This work
AM2521	P$_{kaiB}$::*luc* in NS2	Cm	This work

Table S5. *E. coli* strains used in this study.

Strain	Genotype	Source
DH5α	F– Φ80*lacZ*ΔM15 Δ(*lacZYA-argF*) U169 *recA1 endA1 hsdR17* (rK–, mK+) *phoA supE44 λ– thi-1 gyrA96 relA1*	Lab collection
UU1581	(*flhD-flhA*)Δ*tr4*(*tsr*) Δ*7028 zdb::Tn5*(*trg*)Δ*100thr*(Am)-1 *leuB6 his*-4 *met*F(Am)*159 rpsL136 thi-1 ara-14 lacY1 mtl-1 xyl-1 xyl-5 tonA31 tsx-78*	Gift from J.S. Parkinson
LGM194	F- Δ*lacX74 galE thi rpsL* Δ*phoA* (Pvu II) Δ*ara714 leu::Tn10*	Lab collection
JM109	F′ *traD36 proA*$^+$*B*$^+$ *lacI*^{q} *Δ(lacZ)M15/ Δ(lac-proAB) glnV44 e14*$^-$ *gyrA96 recA1 relA1 endA1 thi hsdR17*	Lab collection

Movie S1. Formation of finger-like projections during UTEX 3055 phototaxis.

movie S2. Flow of UTEX 3055 cells moving toward light source.

Movie S3. Moving of finger tips toward light source.

Movie S4. UTEX 3055 cell movement under parallel gradient created from light source above cells.

Movie S5. UTEX 3055 cell movement under lateral illumination.

Movie S6. Movement of PCC 7942 cells under lateral illumination.

Movie S7. Lensing effect of UTEX 3055 cells during phototaxis.

Movie S8. Lensing effect of UTEX 3055Δ*pixJ* mutant during phototaxis.

References

1. Holtman CK, et al. (2005) High-throughput functional analysis of the *Synechococcus elongatus* PCC 7942 genome. *DNA Res* 12(2):103–115.

2. Chen Y, Holtman CK, Taton A, Golden SS (2012) Functional analysis of the *Synechococcus elongatus* PCC 7942 Genome. *Functional Genomics and Evolution of Photosynthetic Systems*, Advances in Photosynthesis and Respiration. (Springer, Dordrecht), pp 119–137.

3. Taton A, et al. (2014) Broad-host-range vector system for synthetic biology and biotechnology in cyanobacteria. *Nucleic Acids Res* 42(17):e136.

4. Sendersky E, Simkovsky R, Golden S, Schwarz R (2017) Quantification of Chlorophyll as a Proxy for Biofilm Formation in the Cyanobacterium *Synechococcus elongatus*. *BIO-PROTOCOL* 7(14). doi:10.21769/BioProtoc.2406.

5. Paintdakhi A, et al. (2016) Oufti: an integrated software package for high-accuracy, high-throughput quantitative microscopy analysis. *Mol Microbiol* 99(4):767–777.

6. Mackey SR, Ditty JL, Clerico EM, Golden SS (2007) Detection of rhythmic bioluminescence from luciferase reporters in cyanobacteria. *Methods Mol Biol* 362:115–129.

7. Yoshihara S, Katayama M, Geng X, Ikeuchi M (2004) Cyanobacterial phytochrome-like PixJ1 holoprotein shows novel reversible photoconversion between blue- and green-absorbing forms. *Plant Cell Physiol* 45(12):1729–1737.

8. Nagar E, et al. Type 4 pili are dispensable for biofilm development in the cyanobacterium *Synechococcus elongatus*. *Environmental Microbiology* 19(7):2862–2872.

9. Berkelman TR, Lagarias JC (1986) Visualization of bilin-linked peptides and proteins in polyacrylamide gels. *Anal Biochem* 156(1):194–201.

10. Angermayr SA, Gorchs Rovira A, Hellingwerf KJ (2015) Metabolic engineering of cyanobacteria for the synthesis of commodity products. *Trends Biotechnol* 33(6):352–361.

11. Gambetta GA, Lagarias JC (2001) Genetic engineering of phytochrome biosynthesis in bacteria. *Proc Natl Acad Sci USA* 98(19):10566–10571.

2.3 Acknowledgements

Chapter 2, in full, is a reprint of the published manuscript: Yang Y, Lam V, Adomako M, Simkovsky R, Jakob A, Rockwell NC, Cohen SE, Taton A, Wang J, Lagarias JC, Wilde A, Nobles DR, Brand JJ, Golden SS. 2018. Phototaxis in a wild isolate of the cyanobacterium Synechococcus elongatus. *Proc Natl Acad Sci* 115(52). The dissertation author was the tertiary author of this paper.

CHAPTER 3: Comparative genomics of *Synechococcus elongatus* explains the phenotypic diversity of the strains

3.1 Abstract

Strains of the freshwater cyanobacterium *Synechococcus elongatus* were first isolated approximately 60 years ago, and PCC 7942 is well established as a model for photosynthesis, circadian biology, and biotechnology research. The recent isolation of UTEX 3055 and subsequent discoveries in biofilm and phototaxis phenotypes suggest that lab strains of *S. elongatus* are highly domesticated. We performed a comprehensive genome comparison among the available genomes of *S. elongatus* and sequenced two additional laboratory strains to trace the loss of native phenotypes from the standard lab strains and determine the genetic basis of useful phenotypes. The genome comparison analysis provides a pangenome description of *S. elongatus*, as well as correction of extensive errors in the published sequence for the type strain PCC 6301. The comparison of gene sets and single nucleotide polymorphisms (SNPs) among strains clarifies strain isolation histories, and together with large-scale genome differences supports a hypothesis of laboratory domestication. Prophage genes in laboratory strains but not UTEX 3055 affect pigmentation, while unique genes in UTEX 3055 are necessary for phototaxis. The genomic differences identified in this study include previously reported SNPs that are in reality sequencing errors as well as SNPs and genome differences that have phenotypic consequences. One SNP in the circadian response regulator *rpaA* that has caused confusion is clarified here as belonging to an aberrant clone of PCC 7942, used for the published genome sequence, that has confounded the interpretation of circadian fitness research.

Importance. *Synechococcus elongatus* is a versatile and robust model cyanobacterium for photosynthetic metabolism and circadian biology research, with utility as a biological production platform. We compared the genomes of closely related *S. elongatus* strains to create a pangenome annotation to aid gene discovery for novel phenotypes. The comparative genomic analysis revealed the need for a new sequence of the species type strain PCC 6301 and includes two new sequences for *S. elongatus* strains PCC 6311 and PCC 7943. The genomic comparison revealed a pattern of early laboratory domestication of strains and clarifies the relationship between the strains PCC 6301 and UTEX 2973, and showed that differences of large prophage regions, operons, and even single nucleotides have effects on phenotypes as wide-ranging as pigmentation, phototaxis, and circadian gene expression.

3.2 Introduction

Cyanobacteria are important on a global scale as widespread primary producers in environments as diverse as the world’s oceans, rivers, freshwater lakes, and deserts (1–3). In addition to their roles in natural environments, cyanobacteria have attracted interest for their use as biotechnology production platforms (4). *Synechococcus elongatus* PCC 7942 is a well-studied freshwater cyanobacterium long established as a cyanobacterial model organism used for research in prokaryotic photosynthesis and circadian rhythms as well as one of a few cyanobacterial model strains adopted for biotechnology purposes (5–7). Its model status accrues from its facile genetic manipulation based on natural transformability and robust homologous recombination machinery (8), along with a small genome, planktonic growth habit, and formation of distinct colonies on plates. In addition to PCC 7942, there are four other strains of *S. elongatus* with nearly identical genomes that did not reach the same status due to either their loss of natural competence or historical quirks of fate (9).

Recent discoveries raised the likelihood that lab strains of *S. elongatus* are highly domesticated. For example, under laboratory culturing conditions PCC 7942 exhibits a persistent suspended planktonic phenotype, even in the absence of agitation or bubbling, with no evidence of biofilm formation on the culture vessel. Schatz et al. identified and characterized a biofilming mutant of PCC 7942 (10). Studies using conditioned media showed that the wild-type (WT) lab strain secretes an unknown repressor of biofilm formation, supporting a model of constitutive repression of the biofilm genetic program in PCC 7942. This model, coupled with a 40-year history of lab-adaptation for the strain that may have favored planktonic growth, led to a hypothesis that an environmental isolate of *S. elongatus* would readily form biofilms.

As a test of this hypothesis, *S. elongatus* UTEX 3055 was isolated from Waller Creek, TX, USA in 2014 and was found to share 98.5% nucleotide identity with PCC 7942. Although clearly the same species as PCC 7942, the genome of UTEX 3055 is notably distinct from that of PCC 7942. Moreover, UTEX 3055 forms biofilms in laboratory conditions and is phototactic, with an unusual photoreceptor that controls bidirectional phototaxis (11). Although PCC 7942 is not phototactic, genetic transplantation of the genes for the photoreceptor and other components of the phototaxis pathway from PCC 7942 to UTEX 3055 showed that the photoreceptor genes of PCC 7942 are functional, and phototaxis may be an intrinsic property of PCC 7942 that was lost during laboratory propagation (11). We hypothesized the loss of phenotypes like biofilm formation and phototaxis from the standard lab strains of *S. elongatus* through domestication during laboratory cultivation might be traceable using comparative genomics.

The isolation history of *S. elongatus* strains is the context for understanding the connection between their phenotypes and genotypes. The legacy strains of *S. elongatus* include the earliest isolations from freshwater sources in Texas (PCC 6301, alias UTEX 625) and Southern California (PCC 6311) (12), and strains later isolated from freshwater near San Francisco, California (9). As the earliest isolate and entry in cyanobacterial culture collections, PCC 6301 became the type strain for *S. elongatus*. One of the San Francisco strains was found to be highly transformable and genetically very similar to another transformable strain of unknown isolation history in a collection in Russia (13) (14), and these strains were deposited in the collection of Pasteur Cultures of Cyanobacteria as PCC 7942 and PCC 7943, respectively. PCC 6311 and PCC 7943 were sequenced for this study. The last legacy strain, UTEX 2973, was isolated recently from a frozen archive of UTEX 625 (PCC 6301) (15). In 2015, UTEX 3055 was isolated from Waller Creek, Texas about 60 years after PCC 6301 was sampled from the same source (11).

We undertook a comprehensive genome comparison among UTEX 3055 and the previously characterized *S. elongatus* isolates PCC 6301, PCC 6311, PCC 7942, PCC 7943, and UTEX 2973, hereafter referred to as “legacy strains.” The first results of this analysis are sequence and annotation refinement through a re-sequencing of the type strain PCC 6301 and sequencing of PCC 6311 and PCC 7943, as well as the creation of a curated pangenome annotation for all *S. elongatus* strains. Examination of the genome differences at successively narrowing scales reveals large genome regions that control pigmentation phenotypes, a putative operon of UTEX 3055 necessary for phototaxis, and patterns of SNPs in legacy strains that led to a re-evaluation of the relationships among PCC 6301, PCC 7942, and UTEX 2973, and an explanation of a perplexing SNP in *rpaA*, the master regulator output of the circadian clock, that has previously caused confusion in the literature.

3.3 Results and Discussion

A pangenome analysis approach refines genome sequences and annotations. A pangenome compilation strategy using whole genome alignments and ortholog comparisons was adopted to facilitate comparisons among *S. elongatus* strains, since some strains have DNA segments that are unique or absent relative to others. The pangenome of *S. elongatus* contains 3079 genes, with a shared core genome of 2632 genes. There is high sequence conservation among core genome genes, and yet ~15% of the annotations varied among genomes (Table 3-S1). These annotation variations were adjusted using available published transcriptomics (16, 17), gene essentiality (18), and RNA-Seq (16) data for PCC 7942 to create a universal *S. elongatus* pangenome annotation (Fig. 3-S1). The pangenome annotation adjusts 178 gene annotations, removes pseudogenes and hypothetical annotations that lack transcriptional evidence, and adds non-coding RNAs with transcriptional and essentiality evidence in PCC 7942. The pangenome annotations and associated metadata are available as a supplementary file (Supplementary File 3-S1).

The type strain PCC 6301 was one of the first cyanobacterial genomes sequenced, before the advent of next-generation sequencing (19). When the published sequence of

PCC 6301 is compared with the other legacy strains, there appear to be more than 1000 SNPs and insertion-deletion events (indels) in PCC 6301 (15). However, close examination showed that many apparent SNPs in PCC 6301 result in frameshift or nonsense mutations in genes that are essential for viability in PCC 7942 (18). A sample of PCC 6301 archived cryogenically in 1988 in the Golden lab was re-sequenced and this updated sequence contains none of the previously observed SNPs in essential genes. This new sequence (Genbank: CP085785-CP085787) was used in the subsequent analyses in this paper, and is recommended for any future genomic comparison analysis that uses PCC 6301 as the type strain of *S. elongatus*.

S. elongatus strains share an average nucleotide identity of 98.5%, and yet they have distinct phenotypes in natural competence, light tolerance, phototaxis, and biofilm formation (Fig. 3-1A and 1C). The legacy strains share an even higher average nucleotide identity of 99.9%, and yet previous genome comparison studies have found SNPs that contribute to the high-light tolerance phenotype of UTEX 2973 (20) and the loss of natural competence in PCC 6301 and UTEX 2973 (21). There are reports of transformation of PCC 6301 in the early literature (22–24) at the same time as reports of the superior transformability of PCC 7942 (25). PCC 6301 contains a SNP resulting in a frameshift mutation that inactivates the Type IV pilus component *pilN* necessary for transformation, a mutation shared with UTEX 2973 and PCC 6311*.* Considering the tens of genes required for natural competence in *S. elongatus* (21, 26), it is unsurprising that natural competence would be lost in laboratory strains that are not actively propagated for the trait, and that this difference in transformability paved the way for PCC 7942 to enter labs around the world as a genetically tractable cyanobacterial model.

S. elongatus forms a monophyletic group in the *Synechococcus-Prochlorococcus* clade, with the members of this species clustering separately from other *Synechococcus* species (Fig. 3-1B, 3-S2). Within this monophyletic group there are two groups of strains that have been published as *S. elongatus*: those strains isolated from California and Texas, and two *Synechococcus* spp. recently isolated from Powai Lake in India (27, 28). Although the Indian isolates were named *S. elongatus* in publication, they share a sequence identity of only ~83% with PCC 7942, well below the 95% threshold for species relatedness (29). These isolates broaden the phylogenetic branch of this unique group of freshwater *Synechococcus* but were not included in our analysis because of the narrow species-level focus of this study.

Large-scale genome differences suggest a pattern of laboratory domestication. There are three types of large-scale genomic differences among *S. elongatus* strains: a chromosomal inversion region, plasmids, and prophage regions (Fig. 3-2A). A known 188.6 kb inversion is present in PCC 7942 relative to the other strains (19, 25). The sequence of PCC 7943 also contains this inversion (Fig. 3-S3), occurring in the early N-terminal coding region of two porin genes, *somB* and *somB2*, before the predicted conserved porin-domain coding region. Both *somB* and *somB2* contain HIP1 (highly iterated palindrome) sequences ahead of the inversion. HIP1 sequences are hyper-abundant in cyanobacterial genomes and are implicated in site-specific recombination (30). The inversion does not have any known phenotypic effect, but does correlate with a close relationship between PCC 7942 and PCC 7943 that is consistent with the known history of the strains.

The legacy strains of *S. elongatus* carry a 46.3 kb plasmid (large, pANL), and a 7.8 kb plasmid (small, pANS) that can be cured from the strains (31, 32) (Fig. 3-2B and 3-2D). There is a long history of constructing cyanobacterial shuttle vectors from the backbone of pANS (33), including a self-replicating shuttle vector (34). The large legacy strain plasmid, pANL, has four regions characterized by functions in replication, signal transduction, plasmid maintenance, and sulfur metabolism (35). UTEX 3055 lacks both pANS and pANL, but has two plasmids not seen in the legacy strains, here named pMAS and pMAL. The large plasmid of UTEX 3055, pMAL, is 89.2 kb, and shares ~35 kb of homologous content with pANL of the legacy strains (Fig. 3-2B). The homologous regions of pANL and pMAL include a plasmid maintenance region and the sulfur metabolism cluster of pANL. The experimentally determined replication origin of pANL contains 149-bp direct repeats and overlapping pairs of paralogous ORFs hypothesized to be the result of duplication or transposition events (35). A region homologous to the replication origin of pANL is found in UTEX 3055 pMAL, followed by a 49.5 kb region with 49 ORFs not homologous to pANL in gene clusters related to sulfonate and heavy-metal metabolism as well as a putative plasmid maintenance region. This expanded region of pMAL is flanked on either side by a duplicated pair of genes (UTEX3055_pgB029/B030; UTEX3055_pgB080/B081) homologous to a pair of genes in pANL (Synpcc7942_B2632/B2633) (Fig. 3-2B). The homology and synteny with pANL, duplicated genes flanking the expanded region, and the presence of plasmid maintenance genes within the expanded region point to a possible fusion of pANL and another plasmid as the origin of pMAL in UTEX 3055. Site-specific recombination between plasmids at HIP1 sequences has been documented in *Synechococcus* (30), supporting this hypothesis.

The small 24.4 kb plasmid of UTEX 3055, pMAS, contains a plasmid maintenance region, a putative signal-transduction region, and a Type I-C CRISPR-Cas system (36–38) (Fig 2C) similar to that of *Synechococcus* sp. PCC 7002, including a similar direct repeat sequence in the CRISPR array (Table S2). The spacer sequences were used to search the NCBI nucleotide database and the UTEX 3055 genome for self-targeting spacers, with no significant matches. This outcome is not unexpected, as only a tiny fraction of spacers found in genomic CRISPR arrays can be matched confidently to a protospacer sequence (39). In the pMAS CRISPR-Cas system, Cas4 is fused with Cas1, a common arrangement in several Type I systems, but also contains Cas6, which is typically absent from Type I-C systems (37). The system may be under the transcriptional control of a WYL-domain containing protein gene directly upstream of the first gene of the system, as a similar transcriptional regulator in *Synechocystis* sp. PCC 6803 negatively regulates a CRISPR-Cas system in that strain (40).

The legacy prophage controls pigmentation in PCC 7942. The two largest regions of difference between the legacy strains and UTEX 3055 are prophage regions (Fig. 3-1A). The legacy strains have a 49 kb insertion not present in UTEX 3055 that was previously described in PCC 7942 and PCC 6301 as encoding a 25 kb cryptic prophage with similarity to marine cyanosiphoviruses (41, 42). Further investigation of this insertion from 711,254 – 759,991 (Synpcc7942_0716 – Synpcc7942_0767) in PCC 7942 confirms a prediction made by Phage_Finder (43) that this insertion encodes a 49 kb prophage, which inserted into a tRNA-Leu gene (Synpcc7942_R0040/UTEX3055_pg0872) so that the prophage is flanked by phage attachment (*attL*/*attR*) sites composed of an exact duplication of the last 60 bp of the tRNA-Leu gene. This prophage is completely missing from UTEX 3055, which has a different 89 kb prophage inserted in tRNA-Gly (Synpcc7942_R0032/UTEX3055_pg0587) (Fig. 3-1A). Similar to the prophage found in legacy strains, the UTEX 3055 prophage was identified through predicted phage genes and flanking duplicate *att* sites. Completely dissimilar to that of the legacy strains, the UTEX 3055 prophage is most similar to the freshwater cyanophage S-EIV1 (44).

We recognized in the literature a strain of PCC 7942 described by Watanabe et al. to be lacking a ~50 kb region covering the majority of the prophage region as a potential prophage excision strain (45). After obtaining this strain from the Yoshikawa lab, named Δ50kb in this work, we verified through PCR and Sanger sequencing that this strain lacks the prophage and possesses only one copy of the *att* site, as would be expected if the prophage excised or had never integrated into this strain. Given the presence of the complete prophage in the other legacy strains, we hypothesized that the prophage may not be cryptic and could excise from the genome. The prophage in the legacy strains encodes a putative Cro/C1-type lytic-lysogenic switch between two divergent operons that each encodes putative DNA-binding proteins (Fig. 3-S4). According to published transcriptomic and proteomic data, the lysogenic control operon beginning with Synpcc7942_0764 is actively transcribed and translated under standard laboratory conditions, while the lytic activation operon that includes Synpcc7942_0766, encoding a putative DNA damage inducible antirepressor, is not (16, 46). Because efforts to induce phage excision through DNA damaging treatments including UV irradiation, mitomycin C, and metal toxicity were not successful, we tested the ability of the prophage to excise through overexpression of Synpcc7942_0766 regulated by a theophylline-inducible riboswitch [49]. Ectopic induction of Synpcc7942_0766 overexpression resulted in cell lysis after three days, while theophylline-treated WT PCC 7942 cultures and uninduced cultures continued to grow (Fig. 3-3A and 3-3B). A PCR amplification strategy to detect excision and circularization of the phage genome showed circularized phage genomes and prophage-excised chromosomes following induction of Synpcc7942_0766 (Fig. 3-3C). Although examination of lysed cultures did not reveal phage particles nor did the lysate enable subsequent rounds of infection in WT PCC 7942, the capacity of the prophage to excise suggests that it is not completely cryptic, though it may require environmental conditions not yet tested in the laboratory (47) or a helper phage for mobilization (48). Prophages often undergo “domestication” by the host genome, losing structural or lytic components while retaining those that are beneficial to the host (49, 50). An example of this type of domestication in the laboratory is exemplified by the deletion of the second *att* site in the prophage region of UTEX 2973 (Fig. 3-S3) that would preclude excision of the prophage from this strain.

Although the Yoshikawa lab reported no impact of the lack of the prophage on the growth of ∆50kb as compared to WT PCC 7942, we observed that ∆50kb displays a darker appearance upon long-term growth under high light on solid media. Because resequencing of the ∆50kb strain demonstrated that it possesses five SNPs in addition to the phage deletion, we created a clean deletion of the prophage region in our laboratory’s WT PCC 7942. This strain, designated D1K3, has the same dark pigmentation phenotype as Δ50kb (Fig. 3-4A and 3-4B), which time-course data indicate is due to the lack of chlorosis that is otherwise observed as decreasing concentrations of phycocyanin and chlorophyll in WT PCC 7942 (Fig. 3-4A). UTEX 3055, which does not contain the prophage region of the legacy strains, also has a dark pigmentation phenotype like Δ50kb and D1K3 (Fig. 3-4B). We hypothesized that the legacy strain prophage encodes genes that regulate the concentration of the photosystem pigments of *S. elongatus.*

Regions of the prophage with functional similarity and transcriptional orientation were identified and deleted region by region and tested for pigmentation phenotype. Deletion of section 7 (S7), which contains genes encoding a lysozyme and DNA-binding proteins, resulted in the dark pigmentation phenotype of the phageless strains (Fig. 3-4B). Integration of an S7 amplicon into neutral site I of the *S. elongatus* chromosome (NS1) resulted in the recovery of the chlorosis phenotype in both the D1K3 phageless strain and the S7 deletion strain, indicating that one or more of the genes present in this region of the prophage is necessary for this phenotype. Genes within S7 were then individually deleted and analyzed, revealing that only deletion of either of the co-transcribed Synpcc7942_0759 or Synpcc7942_0760 genes resulted in a dark pigmentation strain (Fig. 3-4C). Synpcc7942_0759 and Synpcc7942_0760 respectively encode a hypothetical protein and a putative restriction endonuclease with high transcription levels under normal laboratory conditions (16). As observed with the S7 complementation, neutral site integration of the Synpcc7942_0759-0760 operon under control of its native promoter recovered the WT light pigmentation phenotype, though an analogous addition of only Synpcc7942_0759 failed to complement the dark phenotype (Fig. 3-4B). Attempts to generate a vector that expresses only Synpcc7942_0760 were consistently unsuccessful, and may be due to expression of the putative restriction endonuclease selecting against clones in *Escherichia coli*. Nonetheless, these data demonstrate that either Synpcc7942_0760 alone or in combination with Synpcc7942_0759 regulates the concentration of the photosystem pigments, likely through the degradation of the light-harvesting phycobilisomes and the photosystem complexes. Degradation of phycobilisomes and the subsequent bleaching of cells is mediated in PCC 7942 in response to nutrient limitation (51) by non-bleaching protein A (NblA). Some marine and freshwater cyanophages carry *nblA* genes, presumably favoring the metabolic needs of the phage during a lytic infection (52–54). Synpcc7942_0759 and Synpcc7942_0760 may represent a similar phage strategy of dismantling light harvesting complexes through a pathway independent of a phage-encoded *nblA*.

Unique genes in UTEX 3055 are necessary for phototaxis and support a domestication hypothesis for legacy strains. A gene set enrichment analysis (GSEA) of the set of genes in UTEX 3055 that lack homologs in the legacy strains indicates the genome of UTEX 3055 is enriched in mobilome (mobile genetic elements), defense-mechanism, motility, and cell-cycle COG category genes (Fig. 3-S5, Table 3-S3). Many genes in the defense mechanism COG category are toxin-antitoxin systems (TAS), which are associated with phage inhibition (55) as well as exposure to diverse environmental stresses (56) where they may be beneficial as stress-response elements for bacteria living in varying environments (57). UTEX 3055, as a new environmental isolate, has a more recent history of environmental stress than legacy strains that have been cultivated in controlled laboratory environments for decades. UTEX 3055 has 9 novel TAS not found in the legacy strains, and shares 8 of the 11 TAS found in legacy strains. In four of the shared TAS, UTEX 3055 has a deletion or frameshift mutation in the toxin gene of the

TAS (Table S4), suggesting that these TAS are in the process of being lost. Prokaryotic genomes are shaped by the flux of gene addition via horizontal gene transfer and gene loss, which is a more common mechanism (58). The stasis of TAS in legacy strains compared to the addition and loss of TAS in UTEX 3055 is an indication of how the forces of laboratory domestication do not always lead to loss, as in the case of the prophage *attR* site in UTEX 2973, but can instead stabilize some types of genome elements.

The enrichment of motility genes in the unique gene set of UTEX 3055 is not unexpected, considering its phototactic phenotype. The enrichment in cell cycle genes in UTEX 3055 largely reflects the plasmid maintenance genes of the two plasmids, but further investigation of hypothetical genes listed in this category found two genes, UTEX3055_pg2477/pg2478, that are homologous to an operon of *Synechocystis* sp. PCC 6803 necessary for optimal motility and photosystem function (59). In PCC 6803, their gene products may interact with pilus assembly proteins like the Type II transport protein GspH. UTEX 3055 has a homolog of GspH (UTEX3055_pg2265) encoded within a four-gene operon (UTEX3055_pg2263-pg2266) that is included in the motility category of the enriched unique gene set. This region nestles within a putative operon for synthesizing nucleotide sugars (60), and contains a homolog of *gspH* and hypothetical genes in motility and extracellular structure COG categories. A protein homology search with Phyre2 (61) predicts that each of the four genes in this cluster encodes similarity to pilin or adhesin domains.

In the course of screening a transposon insertion mutant library of UTEX 3055 for phototaxis mutants, we isolated a non-phototactic mutant with an insertion in UTEX3055_pg2266. Because all four genes in the region were hypothesized to have motility functions, the entire region was investigated. We first deleted and replaced the region with a kanamycin resistance cassette through homologous recombination with a mutagenic shuttle vector, and, as expected, this deletion mutant is no longer phototactic (Fig. 3-5). Four complementation vectors for introduction into a genome neutral site (Fig. 3-5) were created to test which of the genes is necessary to restore phototaxis to the deletion mutant. Only addition of the complete novel region restored phototaxis to the deletion mutant (Fig. 3-5). This suggests that all four genes in the region are necessary for phototaxis in UTEX 3055. The addition of this novel four-gene region in the same neutral genome site of PCC 7942 did not confer phototaxis in PCC 7942 (Fig. 3-5). In addition to previous findings that PCC 7942 contains a functional photoreceptor and phototaxis operon that is necessary for phototaxis in UTEX 3055 (11), there are likely additional genes necessary for phototaxis in *S. elongatus* yet to be discovered, and a combination of genes and operons is responsible for the phototaxis phenotype of UTEX 3055.

SNPs in the pangenome of S. elongatus contextualize strain histories and phenotypes. The pangenome analysis revealed more than 40,000 SNPs and ~350 indels among the homologous regions of all *S. elongatus* strains (Supplemental File S2), but only 20% of those SNPs result in amino acid sequence changes in proteins. A GSEA analysis of the homologous regions of the pangenome with high sequence conservation between UTEX 3055 and the legacy strains was used to assess what gene categories are fundamental to the fitness of *S. elongatus* in either environmental or laboratory growth conditions. Homologs with 100% nucleotide sequence conservation are enriched in circadian machinery genes (Fig. 3-S6) while homologs with 95% amino acid conservation are enriched in type IV pili machinery, transcription machinery, and energy-production genes (Fig. 3-S7). These enriched categories are in addition to an enrichment of genes that are conserved across all cyanobacteria and genes that are known essential genes in PCC 7942, underscoring the importance of the circadian and natural competence to *S. elongatus,* two traits that have made the strain such an attractive model organism.

The standardized laboratory culturing conditions that facilitate reproducibility in experiments also present a suite of selective pressures, perhaps unintended, that may shape the genome of *S. elongatus*, and we hypothesized that examination of the differences among legacy strains could provide insight into these selective pressures. We compared the sequence of a currently propagated culture of PCC 7942 in our lab, a revived culture cryogenically archived in our lab in 1988, our resequence of PCC 6301, recent resequencing data available for PCC 7942 and PCC 6301 archived at the Freshwater Algae Culture Collection at the Institute of Hydrobiology, Wuhan, China (62), the sequences of PCC 6311 and PCC 7943 that are presented in this work, and the previously published genomes of PCC 7942 (NC_007604) and UTEX 2973 (NZ_CP006471) (Table S5). In contrast to the tens of thousands of SNPs present between UTEX 3055 and these strains, there are only 120 SNPs and other differences among all available legacy-strain genome data (Table S6). The pattern of shared SNPs across legacy strains correlates with the known isolation and archival history of the strains, with the “Texas” strains isolated from Waller Creek (PCC 6301, UTEX 2973, UTEX 3055) sharing many of the same SNPs (Fig. 3-6), and clarifies a confusing conclusion about the relationship of UTEX 2973 to PCC 6301 and PCC 7942 (15). *Synechococcus* UTEX 2973 was isolated from an archived sample of UTEX 625 (alias

PCC 6301), and was introduced in the literature with a genomic comparison to PCC 6301 and PCC 7942. Yu et al. found ~1600 SNPs and indels between UTEX 2973 and PCC 6301 but only 55 nucleotide differences with PCC 7942 and concluded that UTEX 2973 is more closely related to PCC 7942 than to PCC 6301, acknowledging that the finding was unexpected considering the history of the strains. Our resequencing of PCC 6301 as well as previously published resequencing of the small plasmid of PCC 7942 pANS (alias pUH24) (34) shows that the majority of these reported SNPs were unfortunate sequencing errors. The comparative genome analysis with updated sequence information indicates a greater similarity between UTEX 2973 and PCC 6301 that agrees with the strains' common isolation and cultivation history.

The UTEX 2973 alleles of three genes, ATP synthase subunit alpha *atpA*, NAD$^+$ kinase *ppnK*, and the master regulator output of the circadian clock *rpaA*, have been reported to contribute to the fast-growth (or high-light tolerant) phenotype of UTEX 2973 (20). Of these alleles, *atpA* and *ppnK* are the common allele among *S. elongatus* strains with PCC 7942 as the sole outlier, and Ungerer et al. hypothesize that PCC 7942 has adapted to a low-light lifestyle with these mutations. When the UTEX 2973 sequence of *rpaA* is compared to the published sequence of PCC 7942, there are three differences: an 8 bp deletion in the region upstream of the gene and R121Q and K134E substitutions in the encoded protein. Resequencing of our lab strain PCC 7942 consistently finds four SNPs relative to the published PCC 7942 genome (Table S5), one of which is the same R121Q substitution in RpaA reported in UTEX 2973. However, we have found that the WT allele encodes RpaA-R121 in all cyanobacterial strains, and the RpaA-Q121-encoding allele in the published genome of PCC 7942 is present only in the clone used for sequencing. The expectation that the RpaA-Q121-encoding allele is the WT confounded the interpretation of Ungerer et al. on the contribution of the UTEX 2973 allele of RpaA to the fast-growth phenotype, and has also caused confusion in previous work in our lab on the genetic network of the circadian clock of *S. elongatus.*

One SNP in the circadian response regulator rpaA results in an arrhythmic phenotype. Cyanobacteria are currently the only prokaryotic system with a molecularly described circadian clock, and PCC 7942 is the premier model organism for its study. In cyanobacteria, rhythmic phosphorylation and dephosphorylation of KaiC, a component of the circadian core oscillator, regulates global patterns of gene expression through phosphorylation of the clock output response regulator RpaA. Previously in our lab, a mutant strain of PCC 7942 that lacks rhythmic clock-controlled gene expression was isolated from a transposon mutagenesis screen (63), and the Tn*5* insertion was mapped to a putative open reading frame 358 bp upstream of *rpaA* named *crm (64)*. This insertion in *crm* did not phenocopy an *rpaA*-null mutant and had no impact on *rpaA* transcript or protein accumulation, but KaiC abundance and rhythmic phosphorylation were diminished. These results suggested that the *crm1* mutation had no cis-regulatory impact on *rpaA*, and instead perturbed clock-controlled gene expression through an unknown mechanism (64). However, we recently discovered that the phenotypes ascribed to crm derive from an unusual allele of rpaA (65).

In an effort to understand the role of *crm* in clock-controlled gene expression, the *crm1* insertion allele was reconstructed in a WT background using the mutagenesis cosmid from the original transposon-mutagenesis screen. Of six randomly selected *crm1* clones, three showed WT rhythms of circadian gene expression and three showed the expected arrhythmic phenotype (Fig. 3-6A). Sequencing of *rpaA* from these clones showed that the arrhythmic subpopulation contained the RpaA-Q121 allele matching the published genome of PCC 7942. The rhythmic subpopulation contained the apparent mutant allele RpaA-R121; however, BLAST results of one hundred cyanobacterial RpaA homologs show universal conservation of arginine at this position (Fig. 3-6B) and WT PCC 7942 strains resequenced in our lab also encode the conserved arginine residue. The conservation of the RpaA-R121 allele among cyanobacteria and the arrhythmic phenotype of RpaA-121Q show that RpaA-R121 is the true WT allele of PCC 7942.

From this evidence, we discovered that the single colony of PCC 7942 from our lab that was sequenced by JGI and and published in Genbank, and also used to construct a uni-gene set (UGS) mutant library (63), carried the arrhythmic RpaA-Q121 allele. The UGS cosmid containing the *crm1* transposon insertion also carries the complete coding sequence of *rpaA*-Q121, and so mutants constructed using this cosmid may encode either the arrhythmic RpaA-Q121 or the WT RpaA-R121, depending on where crossovers occur during homologous recombination (Fig. 3-S8). This variable resulted in the two subpopulations observed when reconstructing *crm1* mutants, and explains the arrhythmic phenotype of the original *crm1* mutant. The *rpaA*-Q121 allele was not recognized in the original *crm1* study because of apparent complementation of the arrhythmic *crm1* (i.e. RpaA-Q121) phenotype to rhythmicity attributed to ectopic expression of full-length *crm (64)*. Follow up experiments indicated that the RpaA-Q121 (*crm1*) mutant was poorly transformable, and selection for a transformant during complementation also selected for reversion that restored RpaA function. The sequence of the *crm*/*crm1* complemented strain revealed a second-site suppressor mutation in *rpaA*, RpaA-L4-Q121.

The phenotypic relationship between the RpaA-Q121 arrhythmic and RpaA-L4 suppressor mutations was confirmed by reconstructing *rpaA* mutation combinations via CRISPR/Cas12a engineering (65). In addition to its role as the transcriptional response regulator of the circadian clock, RpaA plays a critical role in redox management, and *rpaA*-null mutants become inviable in darkness (66). The RpaA-Q121 mutant strain is arrhythmic and sensitive to light-dark (LD) cycles, but RpaA-L4 has WT rhythms and LD survival (Fig. 3-7C and 3-7D). The combination of the two substitutions, RpaA-L4-Q121, restored rhythms and improved LD tolerance, confirming the RpaA-L4 mutation as a suppressor of the arrhythmic phenotype of RpaA-Q121. *In vitro* studies of the circadian oscillator show that as RpaA is rhythmically phosphorylated and dephosphorylated in its role as the output of the circadian clock, it rhythmically binds to DNA. The RpaA-Q121 mutant binds DNA poorly despite having a WT phosphorylation pattern; the Q121 substitution prevents the phosphorylated sensor domain of the protein from regulating the DNA-binding domain (65). It is possible that the RpaA-L4-Q121 suppressor mutant restores rhythmic gene expression patterns by restoring the ability of the sensor domain to regulate the DNA-binding domain, subsequently restoring rhythmic gene expression and LD tolerance.

We propose that in the previous *crm1* study, repeated exposure to LD transition events and selection for transformation as part of complementation tests provided a selection for the RpaA-L4-Q121 suppressor mutant. We leveraged the LD sensitivity of RpaA-Q121 in a selection strategy to search for more suppressors in search of new genes associated with the circadian system. The *rpaA*-Q121 mutation was introduced into a PCC 7942 containing a reporter of circadian gene expression, and diluted cultures were plated and grown in LD cycles. Colonies that emerged from the LD selection were then screened for rhythmic gene expression. Ten mutants that showed both LD tolerance and improved rhythmic gene expression were chosen for whole genome sequencing (Fig. 3-S9A and 3-S9B). Of them, one mutant contained a second-site SNP mutation in the promoter region of *rpaA*, and the eight mutants had second-site mutations in either *clpX* or *labA* (Fig. 3-S10, Table 3-S6), genes that have been shown previously to have roles that are not well understood in the mechanism of the circadian clock. *LabA* is required for negative feedback regulation of the core oscillator component KaiC and has been shown to modulate *rpaA* function (67). RpaA directly regulates *clpX* expression, and the protein degradation action of the ClpXP protease fine-tunes the circadian clock (68). The tenth mutant from the LD selection was sequenced but did not actually have improved rhythmic gene expression. This arrhythmic but LD tolerant mutant had a second-site mutation in leucyl peptidase, part of the pathway that recycles glutathione, an important antioxidant that helps maintain the redox balance in cyanobacteria (69). This suppressor screen did not find additional genes in the circadian clock network but reinforced the roles of components that fine-tune the clock mechanism, especially as it relates to maintaining the redox balance of the cell.

The results of this screen may also help explain some SNPs in PCC 7942 and UTEX 2973. In addition to the RpaA-Q121 allele, the published sequence of PCC 7942 contains a SNP in the long-chain-fatty-acid CoA ligase gene (*aas*), which plays a critical role in fatty acid recycling (70–72), resulting in the allele *aas*-L295. Like the mutation in *rpaA*, Aas-L95 is present only in the published sequence of PCC 7942; other cyanobacterial homologs encode a proline residue at that position. We propose that Aas-

L295 is also a second-site repressor of RpaAQ121. Fatty acid accumulation is seen in *rpaA* null mutants of PCC 7942 (66), possibly as a result of redox crisis; the Aas-L295 mutation may mitigate the effects of this accumulation. In UTEX 2973, there are two unique differences in RpaA that have not been fully investigated, due to the erroneous expectation of RpaA-Q121 as the WT allele: a deletion 107 bp upstream of the start codon of *rpaA*, and a K134E substitution. These differences are similar to the -30 *rpaA* G>A suppressor of RpaA-Q121 pair of mutations resulting from the suppressor screen and may represent an inactivating mutant of RpaA with its suppressor mutation.

3.4 Conclusions

This work paves the way for improved future genomic analysis in *S. elongatus* by correcting the PCC 6301 genome sequence and bringing it closer to the sequences of the legacy strains, specifically to UTEX 2793 that is presumably derived from it. It also explains the genetic basis of the *crm1* arrhythmic mutant of PCC 7942, previously attributed to an ORF upstream of *rpaA*, but in fact deriving from an allele that is neither WT nor a sequencing error, but deriving from a rare mutant clone used for the published reference sequence for PCC 7942.

The comparative genomics analysis identified specific loci that explain a difference in pigmentation and phototaxis phenotypes between UTEX 3055 and the legacy strains. The patterns of shared and unique SNPs and genes between UTEX 3055 and the legacy strains are compatible with a domestication hypothesis; the repeated passage of laboratory cultures by pipetting or pouring would favor planktonic cells that do not form biofilms, and may have led to an early selection among the legacy strains resulting in a planktonic phenotype. In the absence of biofilms, there would be no selection for phototaxis, a phenotype also missing from legacy strains. These patterns of differences will aid the future discovery of additional genes responsible for the phenotypic differences between strains. For example, the model strain PCC 7942 has been actively curated for its facile genetic manipulation in the lab, and has highly efficient transformation; in direct comparison experiments testing, UTEX 3055 is ~100X less efficient (unpublished data) than PCC 7942. One suspected difference in UTEX 3055 that may be responsible for this reduced transformation efficiency is its CRISPR-Cas system not present in the legacy strains, and further investigation of this system could lead to more efficient genetic manipulation of UTEX 3055 in the lab. Another pattern of difference between UTEX 3055 and the legacy strains is an enrichment in unique motility COG category genes, including genes related to exopolysaccharide synthesis. This enrichment is consistent with the biofilming and phototactic phenotype of UTEX 3055, and investigation of this gene set may reveal the genetic basis of biofilm formation and phototaxis in *S. elongatus.*

3.5 Materials and Methods

Whole genome alignment and pangenome annotation analysis. The following genomes of *S. elongatus* PCC 7942 [NC_007604; NC_007595; KT751091], PCC 6301 [NC_006576 (previous);], PCC 6311[], PCC 7943[], UTEX 2973 [CP006471-3], UTEX 3055 [NZ_CP033061-3] were used for whole-genome alignment. Chromosomes and plasmids were separately aligned using Mauve (73), and the alignment was manually inspected and adjusted using the Mauve plugin in Geneious Prime 2020.1.2 (https://geneious.com). SNPs, gap locations, and ortholog groups were exported from Mauve and further analyzed with a custom R script, and are available in Supplemental File S2. Core and pangenome elements were determined using full genome alignment and ortholog analysis. Hypothetical proteins in regions of interest were examined further using PSI-BLAST (74) searches for homologs and Phyre2 to search for homologous protein domain architectures (61). Ortholog assignments from Mauve were further refined using Pfam and COG category analysis in eggNOG-mapper (75, 76). These ortholog assignments were checked against homology groups created through reciprocal nucleotide BLAST search using Vespa (77). Homology groups were translated and realigned using MAFFT (78) in Geneious Prime, and the percent identical residues for the nucleotide and amino acid alignments were reported. Annotations were adjusted using annotation consensus agreement, RNA-seq data (16, 17)], and PCC 7942 essentiality data (18) using custom R scripts. Gene metadata, including previously used locus tags, the pangenome annotations, and gene categorical information such as: known biofilm genes from previous S. elongatus PCC 7942 biofilm publications (79–83), pili genes described by Taton, et al. (26), essentiality and conservation data from Rubin, et al.(18), and functional categories from Pfam and COG category analysis were amassed for all genes in the *S. elongatus* pangenome and are provided in Supplemental File S1.

Average Nucleotide Identity, Average Amino Acid Identity, and phylogenetic tree analysis. Average nucleotide identity (ANI) and average amino acid identity (AAI) of whole genomes was obtained using online tools (http://enve-omics.ce.gatech.edu/) (84, 85). A phylogenetic tree was built using 29 conserved housekeeping genes previously defined for bacterial multilocus sequence analysis (MLSA) (86). In addition to the six *S. elongatus* strains, 32 additional cyanobacteria were used to build the tree; *Prochlorothrix hollandica* PCC 9006 was used as the outgroup. Each of the 29-gene sets was aligned using the MAFFT (78) algorithm in Geneious and trimmed using trimAl version 1.2 (87) using the “automated1” option optimized for maximum-likelihood tree construction. The resulting trimmed, aligned files were concatenated and processed using FastTree version 2.1.11 (88) in Geneious. The tree was visualized with FigTree version 1.4.4.

Bacterial Strains and Growth Conditions. The strains used in this study are described in Table. *S. elongatus* PCC 7942 and UTEX 3055 and their derivative strains were grown in BG-11 medium (89) as liquid cultures with continuous shaking (125 rpm) or on agar plates (40 ml, 1.5% agarose) at 30°C under continuous illumination of 100-150 μmol photons m^{-2} s^{-1} from fluorescent cool white bulbs. Culture media for recombinant cyanobacterial strains were supplemented as needed with 2 μg ml^{-1} spectinomycin (Sp) plus 2 μg ml^{-1} streptomycin (Sm), 2 μg ml^{-1} gentamycin (Gm), 7.5 μg ml^{-1} chloramphenicol (Cm), and 5 μg ml^{-1} kanamycin (Km). *S. elongatus* PCC 6301, PCC 6311 and PCC 7943 were grown at the Pasteur Culture Collection (PCC) at 6 μmol photon m^{-2} s^{-1} at 25°C in liquid BG11 medium. PCC 6301 was revived from a 1988 frozen archive in the Susan Golden lab. PCC 6311 and PCC 7943 were used from alive and axenic cultures at the PCC.

Full Genome Sequencing. In preparation for full-genome sequencing *S. elongatus* PCC 6301, PCC 6311 and PCC 7943 cultures were centrifuged and the cell pellets rinsed twice with sterile water and then freeze-dried and lyophilized prior to DNA extraction. For PCC 6301, genomic DNA was extracted with the NucleoBond Genomic DNA purification kit (Macherey-Nagel) as previously used for various pure cyanobacteria (90). For PCC 6301 sequencing a DNA library was prepared using the NextFlex PCR-free DNA sequencing kit (Bioo Scientific, USA) following the manufacturer's recommendations. The library was sequenced on HiSeq 2000 platform (Illumina, San

Diego, CA, USA) in paired-end reads of 101 bases. Sequence files were generated using Illumina Analysis Pipeline version 1.8 (CASAVA; Illumina). After quality filtering with Institut Pasteur-in-house bioinformatic tools, 25,191,450 reads were analyzed using clc_assembly_cell 4.4.0 and CLC Genomics Workbench 7.5.1 (CLC Bio, Qiagen); 20,073,488 paired reads were mapped on the genome sequence of the strain PCC 6301 with clc_assembly_cell 4.4.0 with an average coverage of 917x. The two plasmid sequences were assembled using Genomics Workbench 7.5.1. For PCC 6311 and PCC 7943 the Illumina sequencing service at GATC (GATC Biotech SARL, Mulhouse) was used to generate genome sequences for PCC 6311 (10 scaffolds) and PCC 7943 (9 scaffolds). The genome scaffolds were further assembled into complete chromosome and plasmid sequences using full genome alignment comparison to PCC 7942 in Mauve. The PCC 6301 sequence is deposited in Genbank under accession numbers CP085785-CP085787. The PCC 6311 (ID: SUB10542407) and PCC 7943 (ID:SUB10542407) sequences are deposited in Genbank (pending accessions).

Gene set enrichment analysis. Categorical meta-data available from multiple sources for PCC 7942 was curated for all genes in the pangenome, such as: essentiality and conservation data (18), pili and competence genes (26), and known biofilm genes (10, 79–82), along with the functional categories (COGs) determined in the pangenome analysis. Gene sets of interest in the pangenome were identified, and significant enrichments in meta-data categories for these gene sets were determined using custom R scripts. Briefly, enrichment values were determined using two-sided Fisher's exact tests, with FDR-adjusted p-values ≤ 0.05 being designated as significant. Fold enrichment (F) was calculated as the number of genes in the pangenome interest group that are also in the meta-data category (N_{gc}) divided by the number of genes expected in the group and category (E_{gc}). This expected number was calculated by multiplying the number of total genes in the pangenome interest group (N_g) by the frequency of all genes in the genome that are found in the meta-data category (f_c), which was determined as the number of genes in the category (N_c) divided by the number of genes in the genome (N).

$$F = N_{gc}/E_{gc};\ E_{gc} = N_g * f_c;\ f_c = N_c/N$$

Construction of knockout and complementation strains. Recombinant strains of *S. elongatus* were constructed by natural transformation using standard protocols (91). Excepting RpaA point mutations, cyanobacterial mutants were generated by transforming with knockout vectors engineered with the CYANO-VECTOR assembly system (92) or transposon insertion vectors from the PCC 7942 Unigene set (UGS) library (63, 93). To generate D1K3 a multistep approach was used: the prophage was first tagged at neutral site 3 (NS3), located within the prophage, with a counter-screenable antibiotic-resistance cassette (SpSm) and a counter-selectable marker, *sacB*, which results in cell death in the presence of sucrose. This tagged strain was then transformed with a prophage deletion vector and selected on plates containing sucrose. Complementation and riboswitch expression strains were constructed by expressing gene(s) ectopically in *S. elongatus* chromosomal neutral sites: NS1 and NS2 (42, 91). Complete segregation of the mutant loci was PCR verified. Unless described otherwise, plasmids were constructed using the GeneArt Seamless Cloning and Assembly Kit (Life Technologies) and propagated in *Escherichia coli* DH5α or DB3.1 with appropriate antibiotics. *E. coli* strains were grown at 37 °C in lysogeny broth (LB, Lennox) liquid culture or on agar plates, supplemented as needed with: 100 μg mL^{-1} ampicillin (Ap), 20 μg mL^{-1} Sp plus 20 μg mL^{-1} Sm, 15 μg mL^{-1}

Gm, 17 µg mL^{-1} Cm, and 50 µg mL^{-1} Km. The plasmids used in this study are described in Table

Phage Lysis and Pigmentation Assays. Strains were grown in liquid with continuous shaking as previously described to OD_{750} of approximately 0.8-0.95. Strains were induced with 2 mM theophylline (200 mM stock dissolved in 100% DMSO) or 1% DMSO as a control. For three days following induction, OD_{750} was measured daily and colony PCR was used to determine excision of phage genome. Primers used are in Table 3-S7. For pigmentation assays, strains were grown in liquid BG-11 for 3 - 4 days to an OD_{750} of approximately 0.5, and 4-µL samples of culture were spotted on BG-11 agar and grown at 30 °C under continuous illumination of 300 µmol photons m^{-2} s^{-1} for 30 days. Strains were spotted on the plates in grids to facilitate visual comparison of pairs of strains. For measurement of chlorophyll and phycocyanin content, spots were scraped and resuspended in 200 µL of BG-11. Samples were measured using a Tecan plate reader at OD_{625} for phycocyanin and OD_{675} for chlorophyll, and normalized for cell density by dividing by OD_{750}. In cases where resuspensions were too dense to accurately measure, resuspensions were diluted 1:1 or 1:2 with BG-11, as necessary, and then measured again.

UTEX 3055 Mutant Library Screening. A small Tn*5* mutant library in *S. elongatus* UTEX 3055 was constructed in a similar manner as previously described for PCC 7942 (18). Briefly, UTEX 3055 was grown in liquid culture as previously described until it reached OD_{750}=0.5. A diaminopimelic acid (DAP) auxotrophic *E. coli* donor strain carrying a library of barcoded Tn5 elements (pKMW7) (94) was grown in LB broth with 60 µg/mL DAP and 50 µg/mL kanamycin to an OD_{600}=1.0. Both *E. coli* and *S. elongatus* cells were washed twice and resuspended in BG-11 supplemented with 5% LB at a 1:1 donor cell:recipient cell ratio and spotted on BG-11 w/ 5% LB agar plates with 60 µg/mL DAP. The conjugation reaction was performed for 12 h under 40 µmol photons$\cdot$m^{-2} $\cdot$s^{-1} of illumination and then resuspended in BG-11 and plated onto BG-11 Km agar plates for selection of exconjugants. After 10 d of growth under 100–140 µmol photons$\cdot$m^{-2} $\cdot$s^{-1}, colonies were patched onto BG-11 Km agar plates. To screen for phototaxis mutants, strains were struck onto BG-11 medium with 10 mM sodium thiosulfate solidified with 0.3% agarose (wt/vol) and the plates were placed in a dark box with one side opening toward a fluorescent light and phototactic movement was assessed after 3 days. Strains were screened twice and confirmation of the phototaxis phenotype was performed on 2-µL samples of culture adjusted to OD_{750}= 0.6–1.0 spotted at specific positions on the surface of agarose plates and grown as described to assess phototaxis. The insertion location of the transposon in mutant selected strains was determined by colony PCR with arbitrary primers (Table 3-S7) and Sanger sequencing.

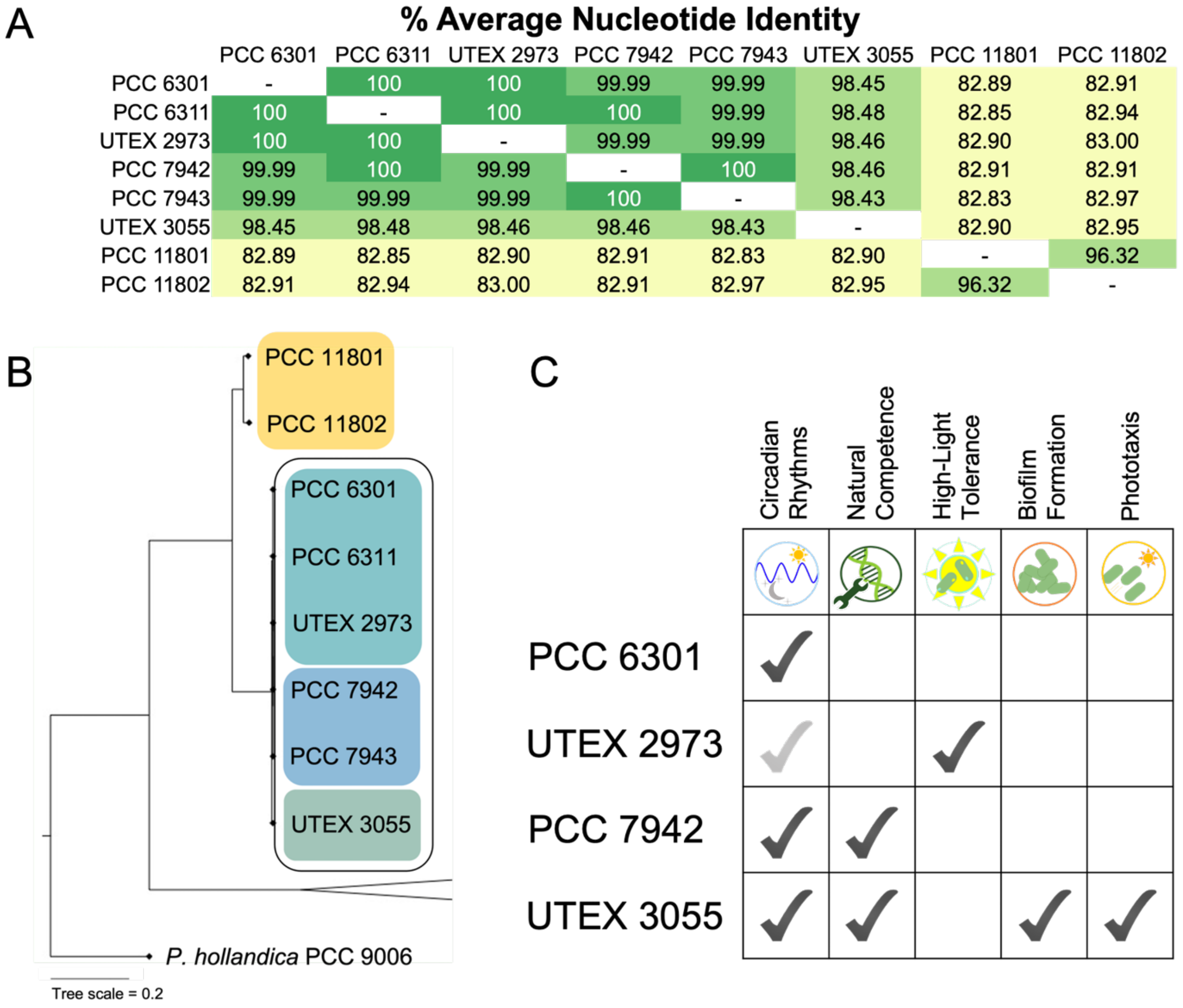
Construction of RpaA mutant strains. Introduction of point mutations into the *S. elongatus* chromosome was accomplished using a previously described CRISPR-editing approach (95). Briefly, the pSL2680 (KmR) plasmid used for CRISPR-Cas12a (formerly Cpf1) editing was purchased from Addgene (Plasmid #85581). Forward and reverse primers upstream and downstream of the desired mutation were annealed together and ligated into AarI-cut pSL2680 to serve as the gRNA template. The resulting construct was purified and digested with KpnI to facilitate insertion of the homology directed repair (HDR) template. The HDR template was generated by amplifying overlapping upstream and downstream fragments containing the desired point mutations. The upstream and downstream HDR fragments were assembled into KpnI-cut pSL2680+gRNA using the GeneArt Seamless Assembly Kit (Thermo Fisher Scientific). The plasmids used in this study are described in Table 3-S7.

Editing plasmids were electroporated into *E. coli* DH10B containing helper plasmid pRL623 and conjugal plasmid pRL443 (92). The resulting strain was grown overnight in LB medium containing antibiotics, washed 3x with fresh LB, and mixed in a 1:2 ratio with an *S. elongatus* reporter-strain aliquot. The cell mixture was plated onto BG-11 agar with added LB (5% vol/vol), incubated under 100 µmol m^{-2} s^{-1} light for 36 h, then underlaid with Km (10 µg/ml final concentration) to select for *S. elongatus* cells that contain the editing plasmid. Colonies that emerged after 6-8 days were passaged three times on BG-11 agar containing Km to allow editing to occur. Successful editing of chromosomal rpaA was verified by sequencing. Plasmids were cured from the edited strains by inoculating cells into non-selective BG-11 medium, growing the culture to OD750 = 0.6, then dilution plating on non-selective BG-11 plates. Fifty colonies were picked and replica patched to selective (Km) and non-selective medium to identify and isolate clones that had lost the editing plasmid.

Suppressor screens and circadian bioluminescence monitoring. A flask of DEC45 (*rpaA*-Q121) was grown to OD750 ~0.8, diluted 1:100 and plated (100 ul per plate) across ten BG-11 agar plates. Plates were grown in LD 12:12 (200 uE) for 7-10 days. 384/~600 colonies that emerged from the LD selection were inoculated into 200 ul of BG-11 media in 96 well plates. Bioluminescence was monitored using a P*kaiBC::luc* firefly luciferase fusion reporter inserted into a neutral site of the S. elongatus chromosome as previously described (96). Strains to be monitored were grown in liquid culture to OD750 = 0.4- 0.7, diluted to OD750 = 0.2, and added as 20 µL aliquots to 280 µL of BG-11 agar containing 3.5 mM firefly luciferin arrayed in 96-well plates. Plates were covered with a gas-permeable seal and cells were entrained under 12-h light-dark cycles (80 µmol m^{-2} s $^{-1}$ light) to synchronize clock phases. After 48 h of entrainment, cells were released into continuous light (30 µmol m^{-2} s $^{-1}$) and bioluminescence was monitored every 2 h using a Tecan Infinite Pro M200 Bioluminescence Plate Reader. Data were collected and plotted using GraphPad Prism 8, with each plot representing the average of six biological replicates. Data were analyzed for rhythmicity using the JTK_CYCLE method provided by BioDare 2 (97, 98) (https://biodare2.ed.ac.uk). LD tolerant suppressors displaying periodic bioluminescence production were scaled-up in flasks for follow-up studies and WGS.

3.6 Acknowledgments

Chapter 3, in full, is in preparation for publication: Adomako M, Ernst D, Simkovsky R, Chao Y-Y, Wang J, Fang M, Bouchier C, Lopez-Igual R, Mazel D, Gugger M, Golden SS. Comparative genomics of *Synechococcus elongatus* explains the phenotypic diversity of the strains. (*In preparation*). The dissertation author is the primary investigator and author of this paper.



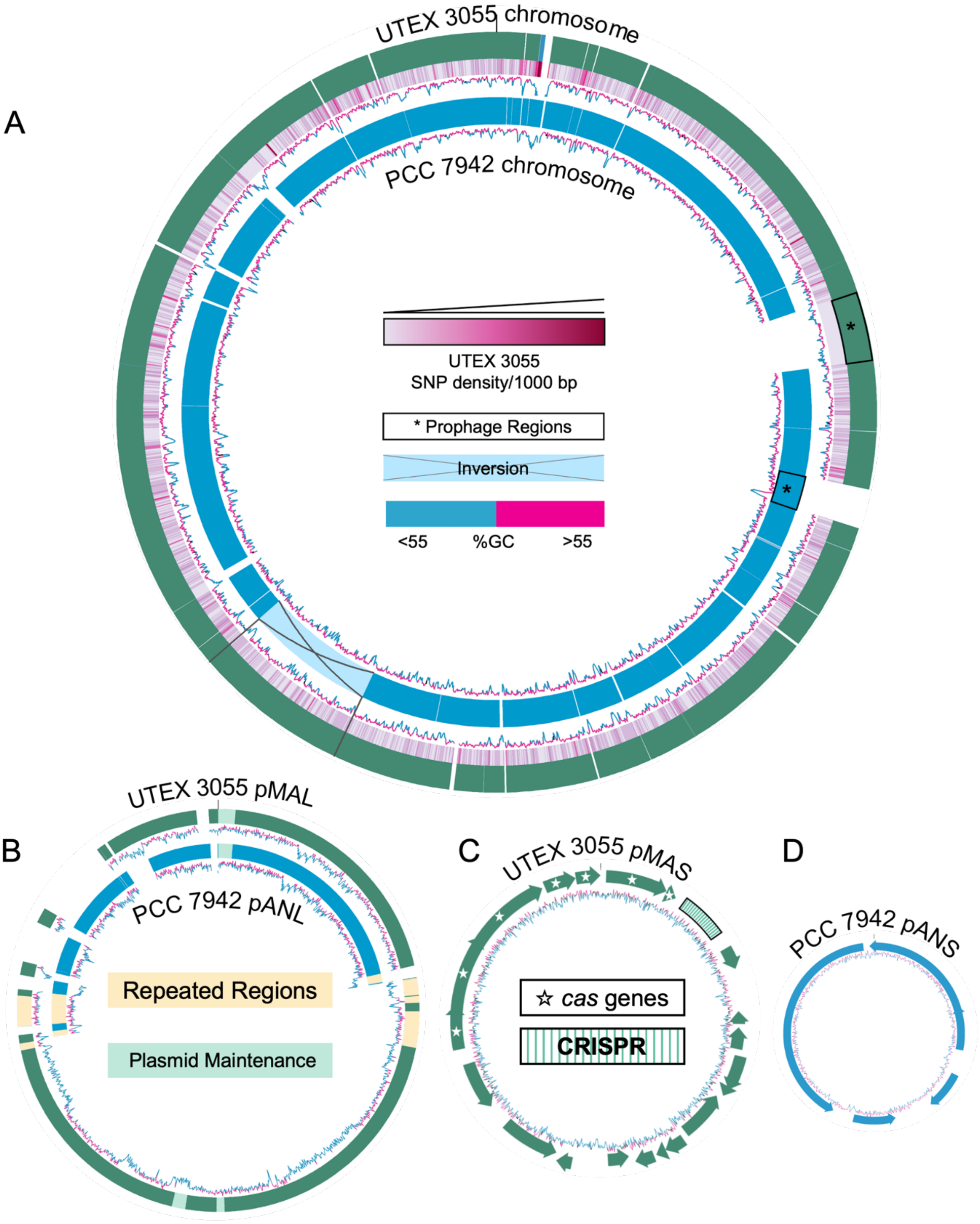
% Average Nucleotide Identity

	PCC 6301	PCC 6311	UTEX 2973	PCC 7942	PCC 7943	UTEX 3055	PCC 11801	PCC 11802
PCC 6301	-	100	100	99.99	99.99	98.45	82.89	82.91
PCC 6311	100	-	100	100	99.99	98.48	82.85	82.94
UTEX 2973	100	100	-	99.99	99.99	98.46	82.90	83.00
PCC 7942	99.99	100	99.99	-	100	98.46	82.91	82.91
PCC 7943	99.99	99.99	99.99	100	-	98.43	82.83	82.97
UTEX 3055	98.45	98.48	98.46	98.46	98.43	-	82.90	82.95
PCC 11801	82.89	82.85	82.90	82.91	82.83	82.90	-	96.32
PCC 11802	82.91	82.94	83.00	82.91	82.97	82.95	96.32	-

	Circadian Rhythms	Natural Competence	High-Light Tolerance	Biofilm Formation	Phototaxis
PCC 6301	✓				
UTEX 2973	✓		✓		
PCC 7942	✓	✓			
UTEX 3055	✓	✓		✓	✓

Figure 3-1: *S. elongatus* genotype and phenotype comparison. The legacy strains and UTEX 3055 share (A) high Average Nucleotide Identity (ANI) percentage and (B) cluster in a monophyletic group separately from other *Prochlorococcus* and *Synechococcus* species (see expanded phylogenetic tree in (Fig. 3-S2), and (C) have unique combinations of phenotypes (grey checkmark represents presumed circadian rhythms in UTEX 2973; the strain has never been explicitly tested for rhythmic gene expression).

Figure 3-2: Comparative genome alignments. (A) Alignment of PCC 7942 (representing legacy strains) and UTEX 3055 chromosomes showing gaps, inversions, and regions of high variability illustrated by SNP density per 1000 bp. (B) Alignment of UTEX 3055 pMAL and pANL of legacy strains; repeated regions of homology and plasmid maintenance genes are highlighted. Gene maps of the small plasmids UTEX 3055 pMAS (C) and legacy strain pNAS (D).


A
UTEX 3055 chromosome
PCC 7942 chromosome
UTEX 3055
SNP density/1000 bp
* Prophage Regions
Inversion
<55
%GC
>55
*
*
B
UTEX 3055 pMAL
PCC 7942 pANL
Repeated Regions
Plasmid Maintenance
C
UTEX 3055 pMAS
☆ cas genes
CRISPR
D
PCC 7942 pANS

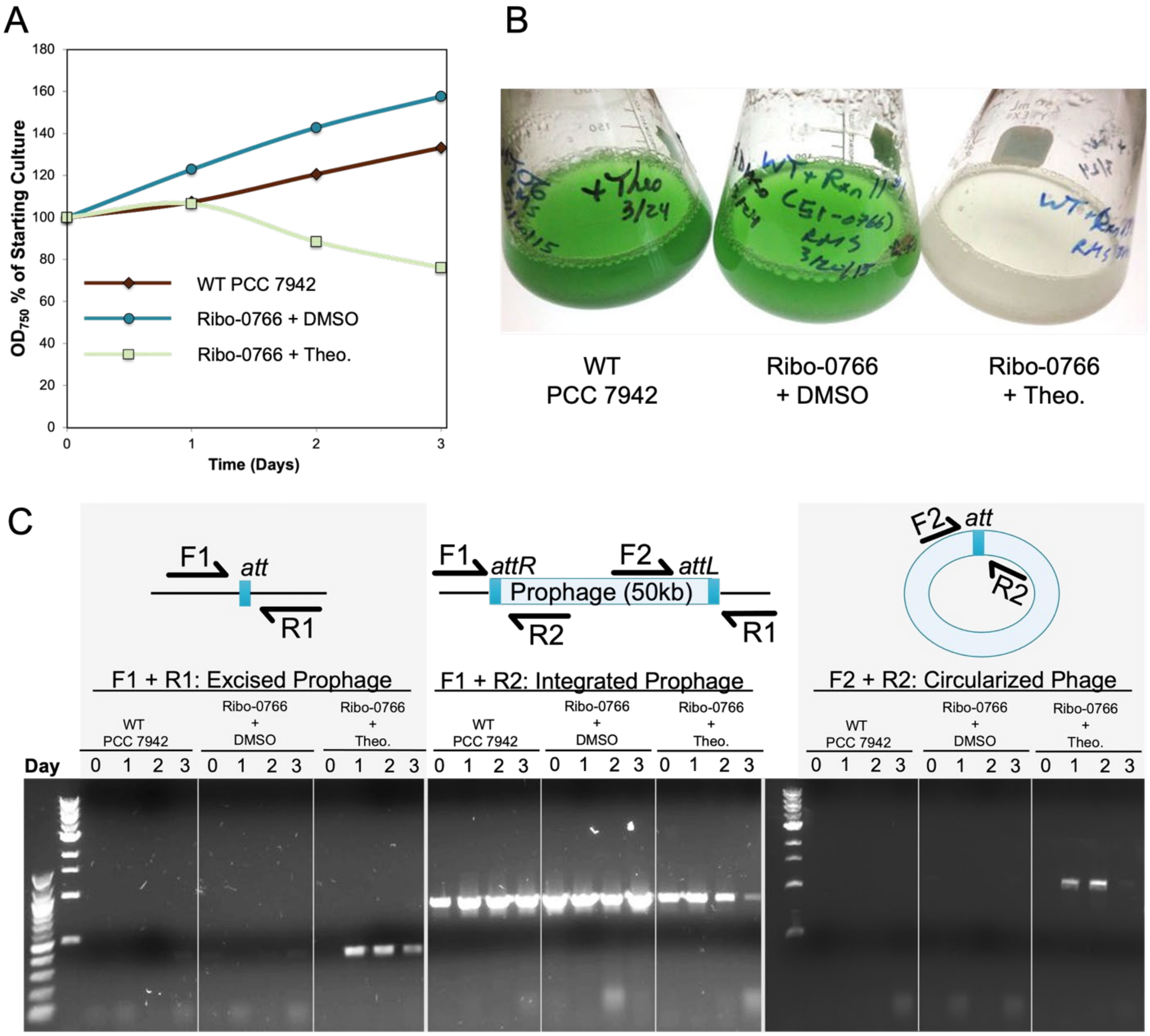


Figure 3-3: Activation of a phage lysis switch in PCC 7942. Activation of a theophylline riboswitch driving overexpression of Synpcc7942_0766 induces a loss of culture density (A) and visible lysis (B) by day 3 post-induction. (C) Excision and circularization of the prophage genome following theophylline induction was observed through whole-cell PCR using primer sets with annealing sites just within and without the prophage region. Cell lysis in induced cultures resulted in faint PCR bands from day 3 samples.

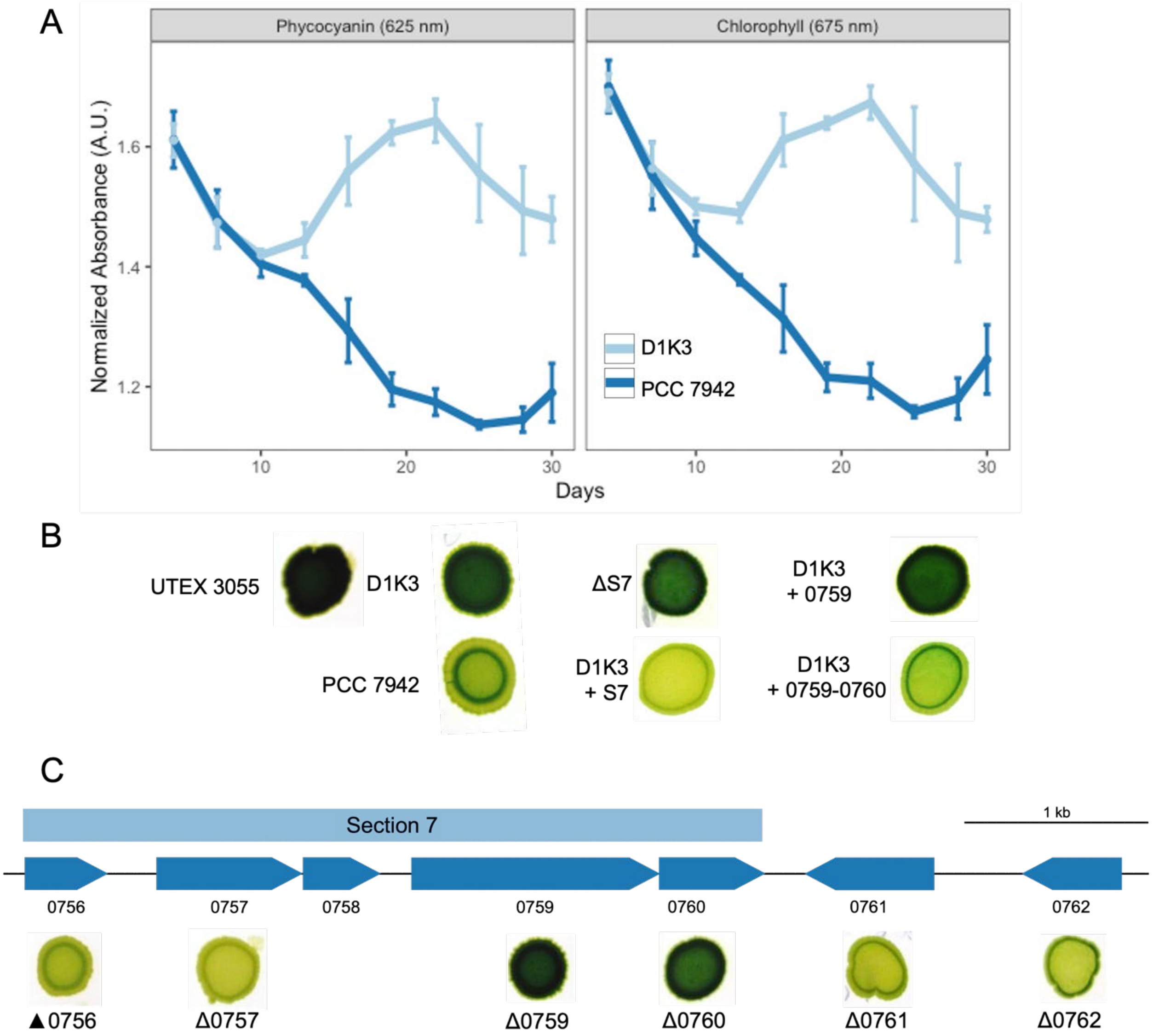


Figure 3-4: Prophage genes control pigmentation in PCC 7942. Prophage genes control pigmentation in PCC 7942. A) Absorbance readings of spot cultures show that PCC 7942 loses phycocyanin and chlorophyll pigments over time compared to the phageless strain D1K3 (average and standard deviation of n=10). B) Loss of either Synpcc7942_0759 or _0760 in Section 7 of the prophage leads to a dark pigmentation phenotype similar to the phageless strain. C) Expression of both _0759 and _0760 together but not _0759 alone in the phageless strain led to WT light pigmentation.

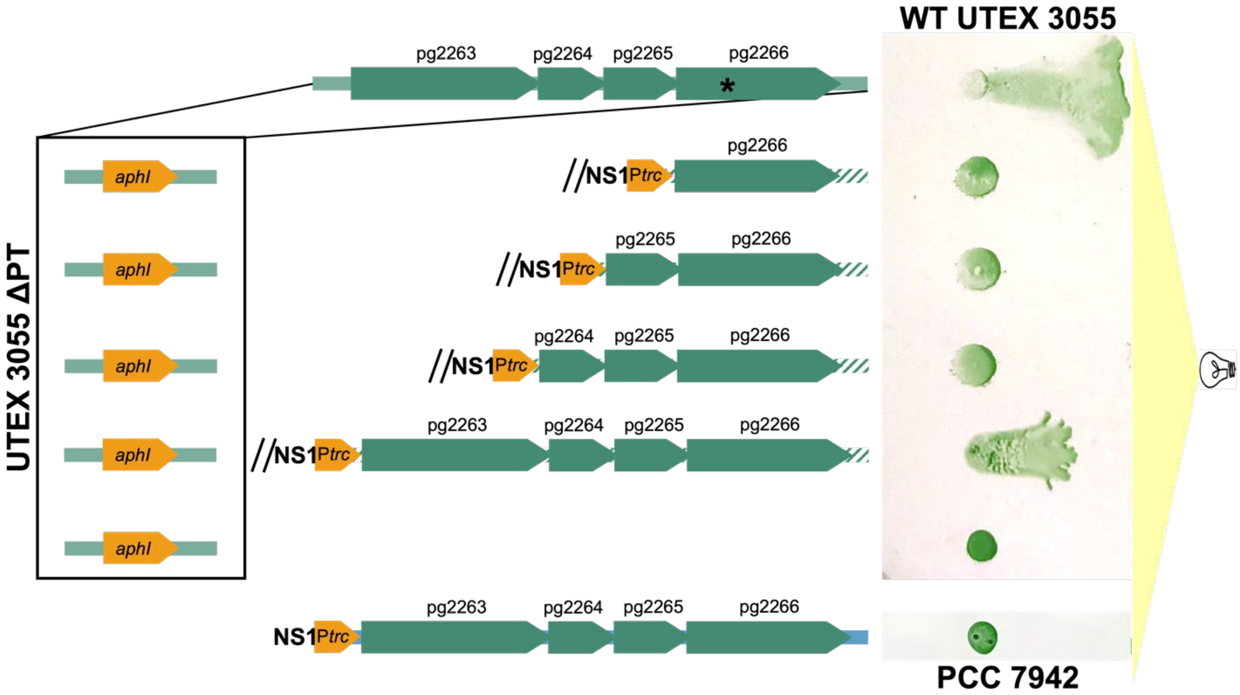


Figure 3-5: A unique operon in UTEX 3055 is necessary for phototaxis. Replacement of UTEX3055_pg2263-pg2266 with a kanamycin resistance cassette (*aphI*) leads to a loss of phototaxis. Site of Tn*5* insertion in the initial phototaxis mutant is marked by *. Complementation with the full four-gene operon restores phototaxis. Introduction of UTEX3055_pg2263-2266 is not sufficient to restore a phototaxis phenotype to PCC 7942.

Table 3-1: SNPs in legacy strains that result in mutations (see Supplementary File 2 for all SNPs and indels); SNPs shared among genomes are in green, SNPs in only one strain are yellow. Two SNPs of the published PCC 7942 sequence (in bold) result in a mutant allele (*rpaA*) and a likely suppressor mutation (*aas*).

Position in PCC 7942	PCC 7942 Locus	Mutation	PCC 7942	AMC06	FACHB-805	FACHB-1061	PCC 7942	PCC 6311	PCC 6301	FACHB-242	UTEX 2973	UTEX 3055
42617	Synpcc7942_0044	G375S	g	g	g	g	a	g	g	g	g	g
49830	Synpcc7942_0049	D52A	t	t	t	t	t	g	g	g	g	g
70982	Synpcc7942_0071	A53V	g	g	g	g	a	g	g	g	g	g
92940	Synpcc7942_0095	E134K	c	c	c	c	c	c	c	c	t	c
92978	**Synpcc7942_0095**	**Q121R**	**t**	c	c	c	c	c	c	c	c	c
233519	Synpcc7942_0238	L269W	a	a	a	a	a	a	c	a	a	a
256284	Synpcc7942_0260	R718C	g	g	g	g	g	a	g	g	g	g
331632	Synpcc7942_0336	C252Y	g	g	g	g	a	a	a	a	a	a
441400	Synpcc7942_0452	L110*	a	a	a	a	a	t	t	t	t	t
649893	Synpcc7942_0654	L39F	c	c	c	c	c	a	a	a	a	a
727190	Synpcc7942_0731	V573I	g	g	g	g	a	g	g	g	g	
753965	Synpcc7942_0760	M1V	g	g	g	g	g	g	a	g	g	
809461	Synpcc7942_0815	G33C	g	g	t	g	g	g	g	g	g	g
810725	Synpcc7942_0816	A501V	g	g	g	g	a	g	g	g	g	g
835019	Synpcc7942_0840	A152V	g	g	g	g	a	g	g	g	g	g
866096	Synpcc7942_0863	R24G	g	g	g	a	g	g	g	g	g	g
893344	Synpcc7942_0886	V80G	a	a	a	a	a	c	a	a	a	a
899800	Synpcc7942_0890	S1018L	c	c	c	c	c	t	t	t	t	c
910005	Synpcc7942_0901	P62S	c	c	c	c	t	c	c	c	c	c
924301	Synpcc7942_0918	G369R	g	g	g	g	g	c	g	g	g	g
924725	Synpcc7942_0918	G75R	g	g	g	g	g	g	c	g	g	g
924962	**Synpcc7942_0918**	**L295P**	**t**	c	c	c	c	c	c	c	c	c
925183	Synpcc7942_0918	P216R	c	c	c	c	c	g	g	g	g	g
925952	Synpcc7942_0918	T625I	c	t	c	c	c	c	c	c	c	c
1015419	Synpcc7942_1003	G329V	g	g	g	g	g	t	t	t	t	g
1060729	Synpcc7942_1047	A138S	g	g	t	g	g	g	g	g	g	g
1062907	Synpcc7942_1050	G184A	g	g	c	g	g	g	g	g	g	g
1067226	Synpcc7942_1056	I174S	t	t	g	t	t	t	t	t	t	t
1163361	Synpcc7942_1139	R292L	c	c	a	c	c	c	c	c	c	c
1192976	Synpcc7942_1159	+9nt									+9nt	
1320845	Synpcc7942_1295	S305I	g	t	g	g	g	g	g	g	g	g
1352729	Synpcc7942_1323	P261S	c	c	c	c	t	c	c	c	c	c

Table 3-1 (cont.) SNPs in legacy strains that result in mutations

Position in PCC 7942	PCC 7942 Locus	Mutation	PCC 7942	AMC06	FACHB-805	FACHB-1061	PCC 7942	PCC 6311	PCC 6301	FACHB-242	UTEX 2973	UTEX 3055
1446920	Synpcc7942_1397	E259G	a	a	a	a	g	g	a	a	a	
1446924	Synpcc7942_1397	G260E	g	g	g	g	a	a	g	g	g	
1446927	Synpcc7942_1397	G260E	g	g	g	g	a	a	g	g	g	
1446928	Synpcc7942_1397	G263E	c	c	c	c	a	a	c	c	c	
1446932	Synpcc7942_1397	G263E	g	g	g	g	a	a	g	g	g	
1446934	Synpcc7942_1397	G270S	g	g	g	g	a	a	g	g	g	
1446936	Synpcc7942_1397	G270S	g	g	g	g	a	a	g	g	g	
1446937	Synpcc7942_1397	L279F	c	c	c	c	t	t	c	c	c	
1446956	Synpcc7942_1397	R262G	c	c	c	c	g	g	c	c	c	
1446972	Synpcc7942_1397	R262G	t	t	t	t	c	c	t	t	t	
1446983	Synpcc7942_1397	V258F	g	g	g	g	t	t	g	g	g	
1451380	Synpcc7942_1400	D333N	c	c	c	c	t	c	c	c	c	c
1458762	Synpcc7942_1407	A236P	g	g	g	g	g	g	c	g	g	g
1515569	Synpcc7942_1463	L67F	g	g	g	g	g	c	c	g	c	c
1574776	Synpcc7942_1522	P350S	c	c	c	c	c	c	t	c	c	c
1574861	Synpcc7942_1522	Q378P	a	a	a	a	a	a	a	c	c	a
1614143	Synpcc7942_1557	R85H	c	c	t	c	c	c	c	c	c	c
1782040	Synpcc7942_1713	A242T	g	g	g	g	a	g	g	g	g	g
1801352	Synpcc7942_1729	P169S	c	c	c	c	t	c	c	c	c	c
1833994	Synpcc7942_1766	W12C	c	c	a	c	c	c	c	c	c	c
2001411	Synpcc7942_1925	G58D	c	c	c	c	t	c	c	c	c	c
2026361	Synpcc7942_1954	H488R	t	t	t	t		c	c	c	c	c
2026943	Synpcc7942_1954	V294A	a	a	a	g	a	g	g	g	g	g
2041501	Synpcc7942_1971	G242R	c	c	c	c	c	c	c	c	g	c
2041732	Synpcc7942_1971	V165F	c	c	c	c	c	c	c	c	a	c
2071861	Synpcc7942_2002	W176*	g	g	g	g	g	g	g	a	g	g
2150741	Synpcc7942_2073	R71L	g	g	g	g	t	g	g	g	g	g
2151115	Synpcc7942_2073	S196G	a	a	a	a	g	g	g	g	g	g
2373132	Synpcc7942_2304	E260D	g	g	g	g	g	c	c	c	c	c
2436301	Synpcc7942_2369	A421S	g	t	g	g	g	g	g	g	g	g
2440583	Synpcc7942_2373	G264V	c	t	c	c	c	c	c	c	c	c
2530540	Synpcc7942_2452	Q113*	g	g	g	g	g	a	a	a	a	a

Table 3-1 (cont.) SNPs in legacy strains that result in mutations

Position in PCC 7942	PCC 7942 Locus	Mutation	PCC 7942	AMC06	FACHB-805	FACHB-1061	PCC 7942	PCC 6311	PCC 6301	FACHB-242	UTEX 2973	UTEX 3055
2555105	Synpcc7942_2473	R35C	c	c	c	c	c	c	c	t	t	c
2564830	Synpcc7942_2483	S84*	g	g	t	g	g	g	g	g	g	g
2607826	Synpcc7942_2526	V100F	g	g	g	g	g	g	t	g	g	g
2653430	Synpcc7942_2574	H225D	g	g	g	g	g	g	g	c	c	g

Figure 3-6: A single SNP in *rpaA* causes arrhythmic gene expression and light-dark sensitivity. A) Reconstruction of the *crm1* mutant leads to arrhythmic and rhythmic populations due to the presence or absence of the RpaA-Q121 allele, respectively. B) The conserved amino acid at position 121 of RpaA in cyanobacteria is arginine. An RpaA-L4-Q121 suppressor mutation restores (C) LD fitness and (D) rhythmic gene expression. Dilution series of strains were grown in constant light (LL) or in 12 hour light/dark cycles (12:12LD) for 48 hours to assess LD fitness. Bioluminescence from strains carrying a P_{kaiBC}-*luc* reporter at NS2 was recorded as an assay for circadian rhythms of gene expression. LL, constant light after entrainment in a 12-h light:12-h dark cycle.

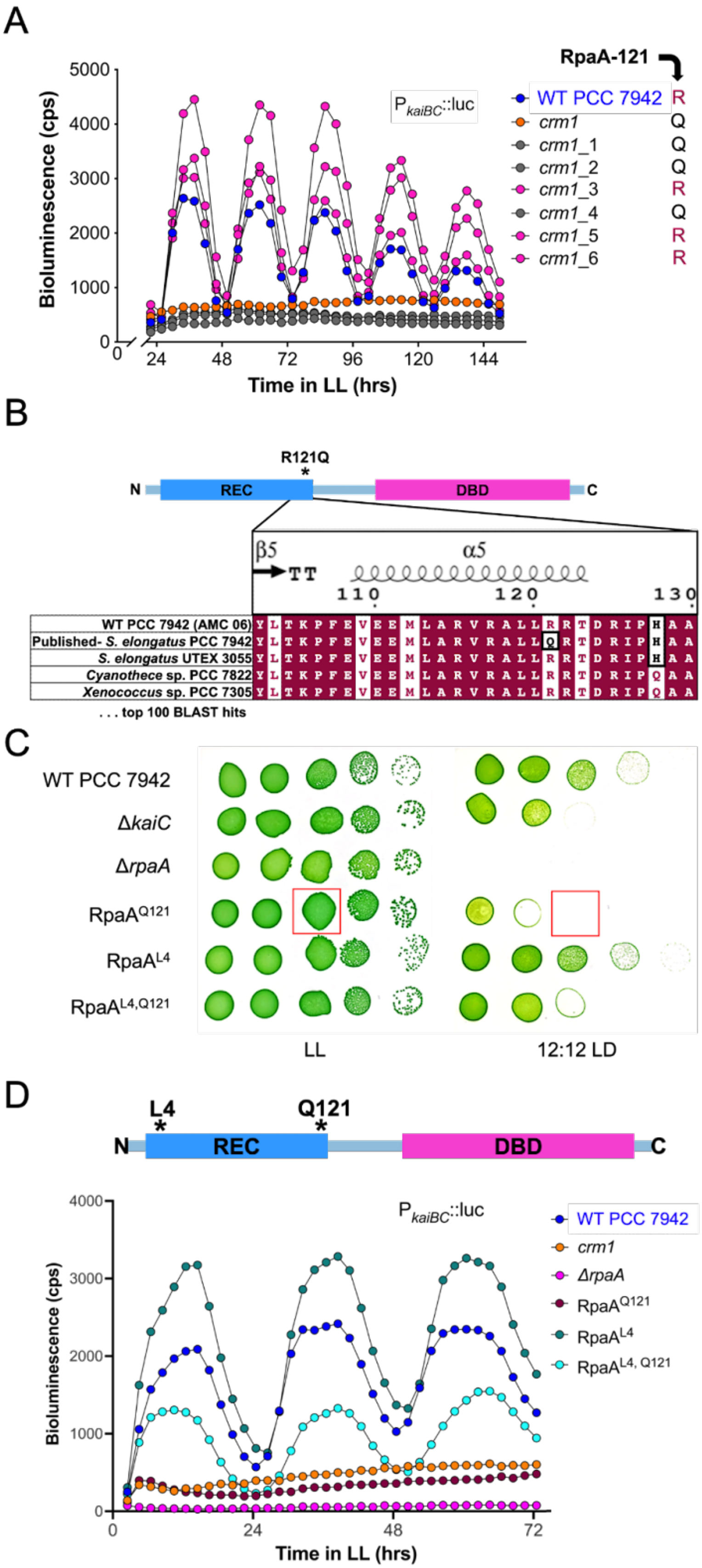

A
RpaA-121
Bioluminescence (cps)
5000
4000
3000
2000
1000
0
0
24
48
72
96
120
144
Time in LL (hrs)
P_kaiBC::luc
WT PCC 7942 R
crm1 Q
crm1_1 Q
crm1_2 Q
crm1_3 R
crm1_4 Q
crm1_5 R
crm1_6 R
B
R121Q
*
N
REC
DBD
C
β5
TT
α5
110
120
130
WT PCC 7942 (AMC 06) YLTKPFEVEEMLARVRALLRRTDRIPHAA
Published- S. elongatus PCC 7942 YLTKPFEVEEMLARVRALLQRTDRIPHAA
S. elongatus UTEX 3055 YLTKPFEVEEMLARVRALLRRTDRIPHAA
Cyanothece sp. PCC 7822 YLTKPFEVEEMLARVRALLRRTDRIPQAA
Xenococcus sp. PCC 7305 YLTKPFEVEEMLARVRALLRRTDRIPQAA
. . . top 100 BLAST hits
C
WT PCC 7942
ΔkaiC
ΔrpaA
RpaA^Q121
RpaA^L4
RpaA^L4,Q121
LL
12:12 LD
D
L4
*
Q121
*
N
REC
DBD
C
Bioluminescence (cps)
4000
3000
2000
1000
0
0
24
48
72
Time in LL (hrs)
P_kaiBC::luc
WT PCC 7942
crm1
ΔrpaA
RpaA^Q121
RpaA^L4
RpaA^L4, Q121

Table 3-S1: Comparison of Pangenome and NCBI Prokaryotic Genome Annotation Pipeline (PGAP) annotations. The Pangenome annotation adds annotations previously missing from legacy strain annotations and adjusts pseudo gene annotations to complete open reading frames. The *attR* sites of prophages are annotated as tRNA genes in the PGAP annotations of PCC 7942 and UTEX 3055 and are removed in the pangenome annotation. NCBI PGAP annotations accessed April 2021.

	PCC 7942		UTEX 2973		UTEX 3055	
	NCBI PGAP	PanGenome	NCBI PGAP	PanGenome	NCBI PGAP	PanGenome
Genes	2758	2730	2702	2715	2807	2832
Protein Coding	2703	2662	2645	2657	2739	2773
tRNA	45	44	44	44	46	45
rRNA	6	6	6	6	6	6
ncRNA	4	11	1	1		1
other RNA		4		4	4	4
Pseudo	7	0	6	0	12	5

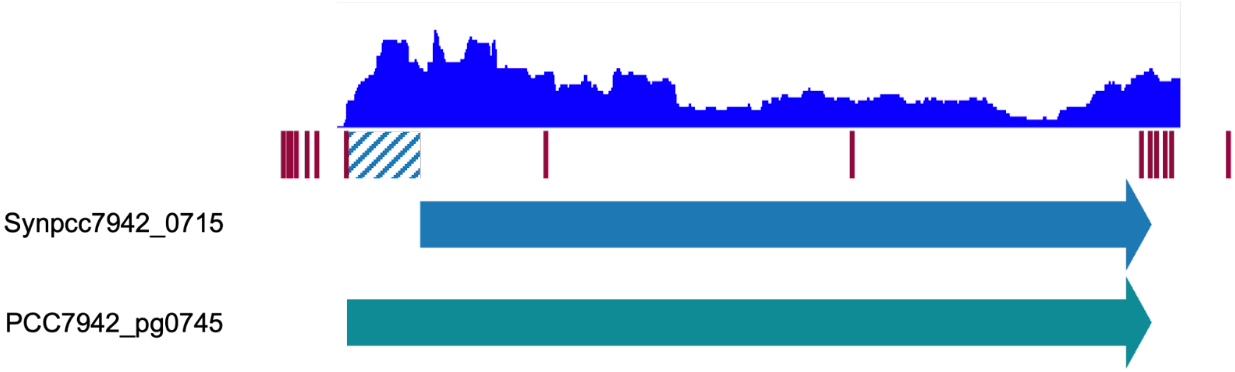


Figure 3-S1: Pangenome annotation adjustments. The annotation of *murZ* (PCC7942_pg0745) was adjusted in the pangenome using RNASeq/transcriptome (16) and gene essentiality (18) data. The NCBI PGAP annotation (Synpcc7942_0715) has an annotated start after the transcript start observed in the RNASeq data (top graph in blue). Essentiality data for PCC 7942 indicates the region between the observed transcript start site and the PGAP annotation start is essential (blue hash box; red lines indicate Tn*5* insertions in the essentiality library of PCC 7942). The pangenome annotation was adjusted to the start codon present at the transcript and essentiality region start.

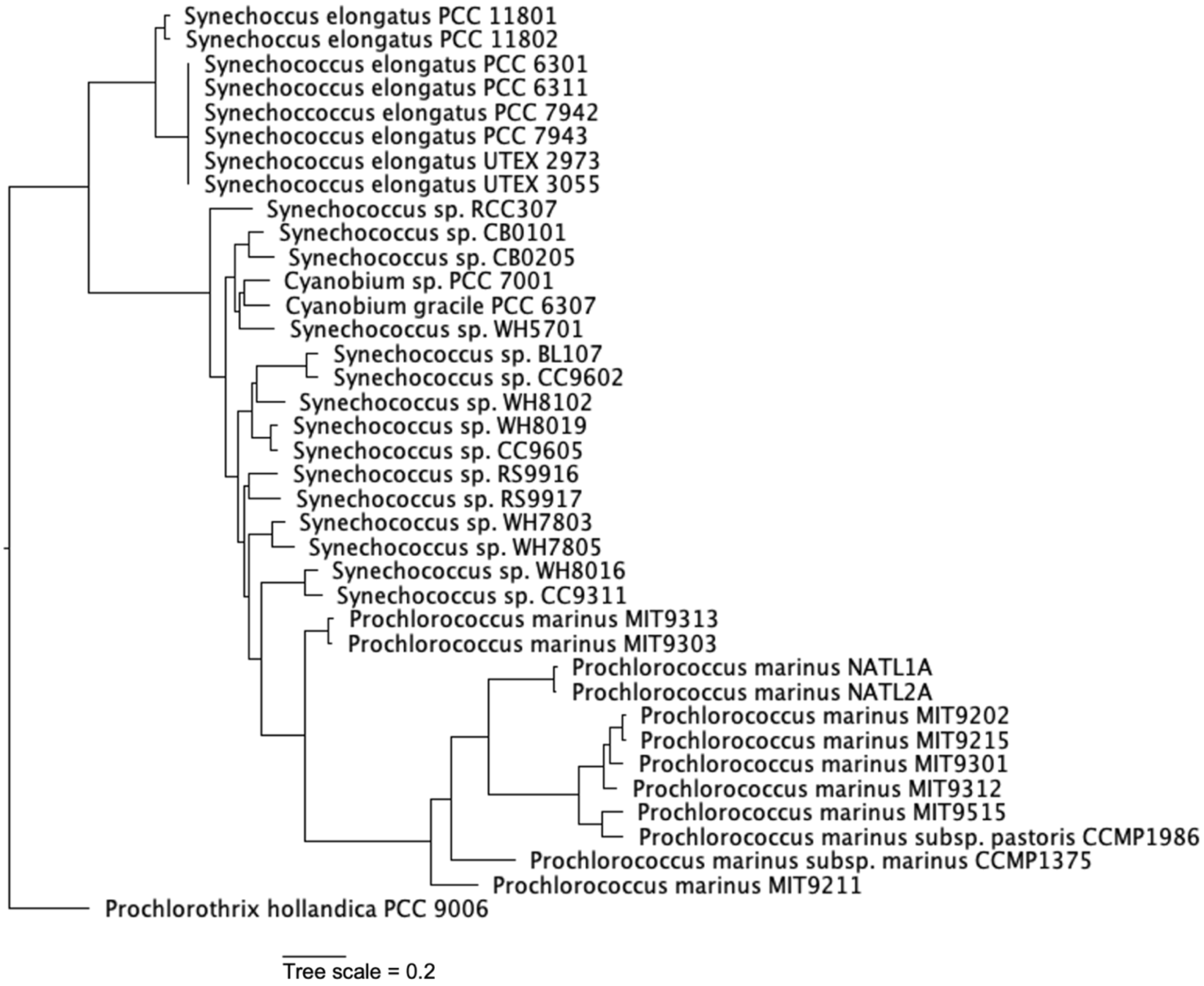

Synechoccus elongatus PCC 11801
Synechoccus elongatus PCC 11802
Synechococcus elongatus PCC 6301
Synechococcus elongatus PCC 6311
Synechoccoccus elongatus PCC 7942
Synechococcus elongatus PCC 7943
Synechococcus elongatus UTEX 2973
Synechococcus elongatus UTEX 3055
Synechococcus sp. RCC307
Synechococcus sp. CB0101
Synechococcus sp. CB0205
Cyanobium sp. PCC 7001
Cyanobium gracile PCC 6307
Synechococcus sp. WH5701
Synechococcus sp. BL107
Synechococcus sp. CC9602
Synechococcus sp. WH8102
Synechococcus sp. WH8019
Synechococcus sp. CC9605
Synechococcus sp. RS9916
Synechococcus sp. RS9917
Synechococcus sp. WH7803
Synechococcus sp. WH7805
Synechococcus sp. WH8016
Synechococcus sp. CC9311
Prochlorococcus marinus MIT9313
Prochlorococcus marinus MIT9303
Prochlorococcus marinus NATL1A
Prochlorococcus marinus NATL2A
Prochlorococcus marinus MIT9202
Prochlorococcus marinus MIT9215
Prochlorococcus marinus MIT9301
Prochlorococcus marinus MIT9312
Prochlorococcus marinus MIT9515
Prochlorococcus marinus subsp. pastoris CCMP1986
Prochlorococcus marinus subsp. marinus CCMP1375
Prochlorococcus marinus MIT9211
Prochlorothrix hollandica PCC 9006
Tree scale = 0.2

Figure 3-S2: Expanded phylogenetic tree of *S. elongatus* strains. Phylogenetic tree of *S. elongatus* strains with representative species from the *Prochlorococcus* and *Synechococcus* clade.

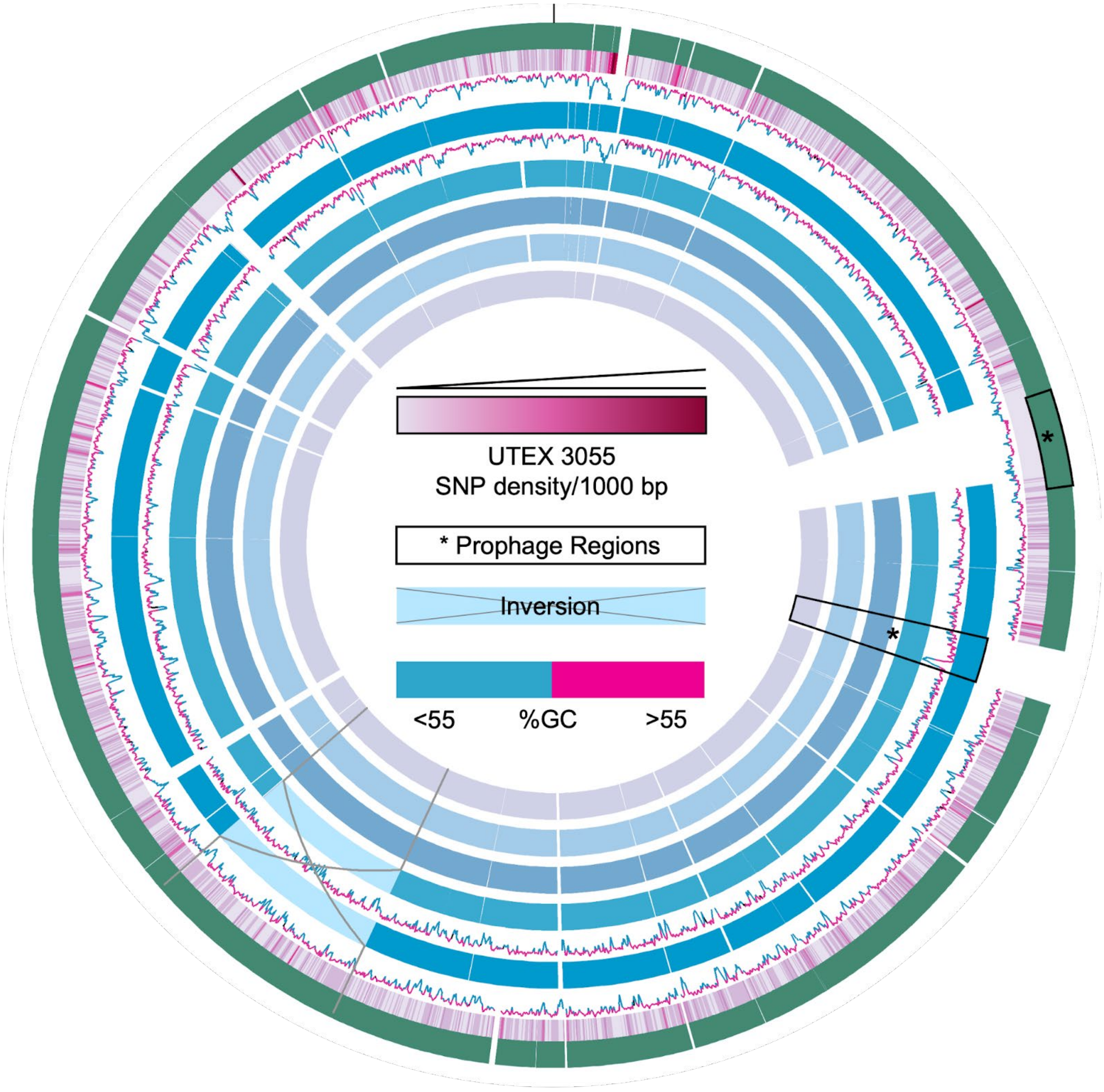
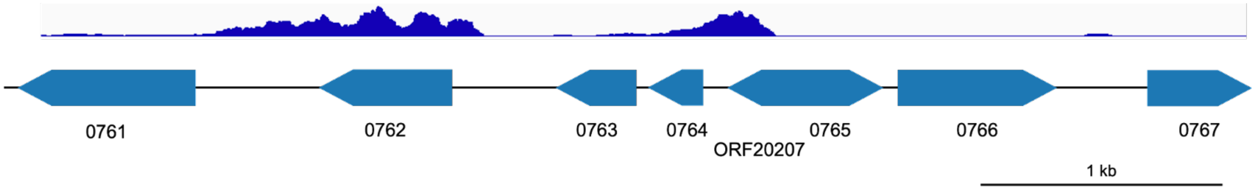


Figure 3-S3: Alignment of all *S.elongatus* chromosomes. Alignment representation of all UTEX 3055 and legacy strain chromosomes showing gaps, inversions, and regions of high variability in UTEX 3055 illustrated by SNP density per 1000 bp. The outer solid ring represents the chromosome alignment of UTEX 3055, the inner solid rings in progression to the center are chromosome alignments of PCC 7942, PCC 7943, PCC 6301, PCC 6311 and UTEX 2973.

Table 3-S2: *S. elongatus* UTEX 3055 CRISPR spacer sequences. Direct repeat sequence of the CRISPR array and nucleotide position and sequence of spacers in UTEX 3055 pMAS.

Direct Repeat: GTGTAATTACCCTTGATGCCATTAGGCGTTGATCAC	
Position	**Spacer**
11101	TTACCGGGTCAAGGTTTGGCGCTGGTGGATTTCCA
11172	AGAGACATCTGGGAAGAGAACGGCTGGGACAAAAAAT
11245	GAGAGAGACTTGCTAGAGATTGCAGATAATCTCTA
11316	AGAGGCGGGCCACCAGGGAGGCCAGGAAGGTTAGGG
11388	GCATTCAAGCTAGGAAAAGGCCCCCGCGTCTCAG
11458	CTCAAGACTTGGGCAGTGGATATCGGTGGCCTTTC
11529	GCAAACACAGGCCGTGGCCCTCGCGTGACAGTCCAGT
11603	ACTCAACTTCTCAGTCAAACCCCTACTTTATAGGA
11673	GGGAAGATGACATCGACATTGACGAGTTTGTCAGCC
11745	GAACCCCTAGGACTGCGGTTTAAGCTTCTCAAGTCTC
11818	TTCCAGGCGGTTTGTGCGGCAAGACCGGTGAGAT
11888	TAGCCCTGCATCAGTTCCTCACCCGCAGCTAGGCC
11959	TTATAGAAATGGTAGGCAAATCCTGCCCAAGCTAA
12030	ACTCTCAAGACACGAGGGCCTCCTTATGGAGTTGCT
12102	TTGTTGGGGTCTTTGGGCTTGTGCAGCCCGATCGT
12173	ACGGAATCGCAACTATTCGGTACTAATCGCCGGTG
12244	CCCCAGAAAAGAACACAGGGGGTCGCTCTCCCTCTG
12316	CGCTCAATACCGGCTGCGGTTTGCGGGAAAAACTCAC



Locus	Description
Synpcc7942_0761	conserved hypothetical protein DUF2971
Synpcc7942_0762	DUF4065 phage-associated HTH antitoxin XRE family
Synpcc7942_0763	XRE family antitoxin, phage immunity repressor protein C
Synpcc7942_0764	XRE family antitoxin, phage immunity repressor protein C
ORF02027	XRE family HTH antidote protein HigA, Cro/C1 type
Synpcc7942_0765	tetraacyldisaccharide-1-P-4-kinase, tRNA methylation domain
Synpcc7942_0766	Bro-N domain, phage antirepressor protein
Synpcc7942_0767	transcription termination factor Rho

Figure 3-S4: Lytic-lysogenic switch region in PCC 7942 prophage. PCC 7942 contains a putative Cro/C1-type lytic/lysogenic switch operon at the right side of the prophage region. The lytic repressor region is actively transcribed in WT PCC 7942 as seen in RNASeq data (top graph).

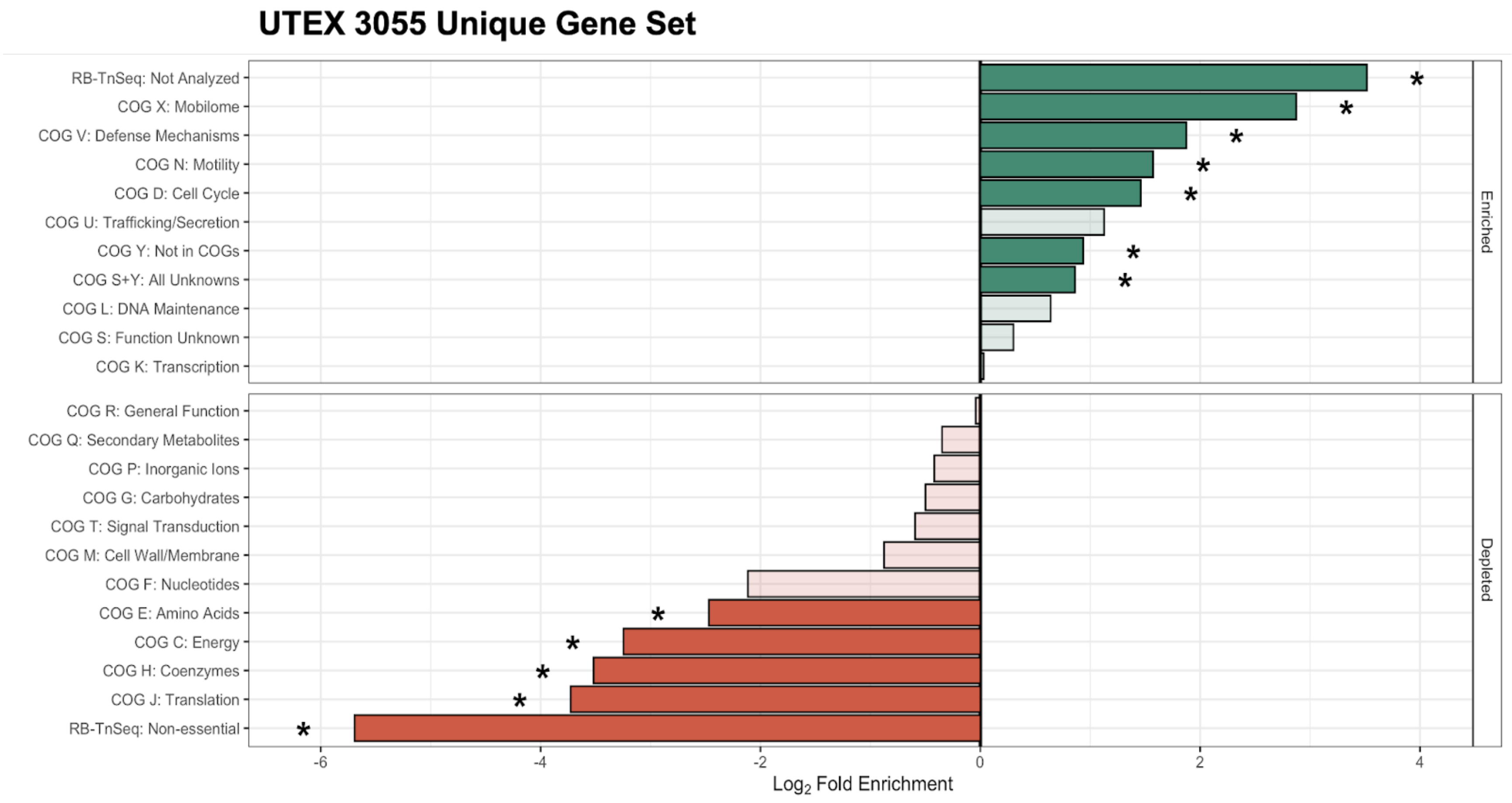


Figure 3-S5: Gene set enrichment analysis of the unique gene set of UTEX 3055. GSEA of the unique gene set of UTEX 3055 shows an enrichment in mobilome, defense mechanism, motility and cell cycle COG category genes. Categories that are significantly enriched or depleted are marked with darker category colors and asterisks.

Table 3-S3: Genes unique to each genome set of legacy strains and UTEX 3055. Genes are arranged by genome position to illustrate unique genes in the same relative genome regions. Highlighted genes indicate genes that are not homologs but perform similar gene functions in the same relative genome region.

Pangenome Locus	Description	Pangenome Locus	Description
UTEX3055_pg0017	DDE_Tnp_1-associated		
UTEX3055_pg0018	Transposase DDE domain protein		
UTEX3055_pg0024	hypothetical protein		
UTEX3055_pg0025	type II toxin-antitoxin system VapC family toxin		
UTEX3055_pg0037	hypothetical protein	PCC7942_pg0033	DUF86 domain protein
		PCC7942_pg0034	nucleotidyltransferase family protein
		PCC7942_pg0035	N-acetylornithine aminotransferase
		PCC7942_pg0036	hypothetical protein
UTEX3055_pg0048	putative transcriptional regulator		
UTEX3055_pg0049	BrnT family toxin		
		PCC7942_pg0050	type IV pilus assembly protein PilA
UTEX3055_pg0056	hypothetical protein		
UTEX3055_pg0057	class I SAM-dependent methyltransferase	PCC7942_pg0054	methyltransferase domain protein
UTEX3055_pg0058	glycosyltransferase family 2 protein	PCC7942_pg0055	glycosyltransferase family 2 protein
		PCC7942_pg0056	cephalosporin hydroxylase
		PCC7942_pg0057	GNAT family N-acetyltransferase
		PCC7942_pg0058	DegT/DnrJ/EryC1/StrS aminotransferase family protein
		PCC7942_pg0059	Class I SAM-dependent methyltransferase
		PCC7942_pg0060	dTDP-4-dehydrorhamnose 3,5-epimerase
		PCC7942_pg0061	hypothetical protein

Table 3-S3 (cont.): Genes unique to each genome set of legacy strains and UTEX 3055.

Pangenome Locus	Description	Pangenome Locus	Description
		PCC7942_pg0062	hypothetical protein
		PCC7942_pg0063	CDP-glucose 4,6-dehydratase
		PCC7942_pg0064	glucose-1-phosphate cytidylyltransferase
UTEX3055_pg0100	hypothetical protein	PCC7942_pg0106	ankyrin-repeat containing protein
UTEX3055_pg0101	Protein of unknown function DUF2283	PCC7942_pg0107	hypothetical protein
UTEX3055_pg0102	hypothetical protein		
UTEX3055_pg0115	SMI1-KNR4 cell-wall	PCC7942_pg0120	hypothetical protein
UTEX3055_pg0181	DUF4433 domain protein	PCC7942_pg0186	DUF4435 domain protein
UTEX3055_pg0182	O-acetyl-ADP-ribose deacetylase regulator of RNase III, contains Macro domain	PCC7942_pg0187	AAA family ATPase
UTEX3055_pg0544	protein of unknown function DUF4433	PCC7942_pg0549	hypothetical protein
UTEX3055_pg0545	hypothetical protein		
UTEX3055_pg0801	hypothetical protein	PCC7942_pg0676	trypsin-like peptidase domain protein
UTEX3055_pg0802	hypothetical protein		
		PCC7942_pg0822	hypothetical protein
		PCC7942_pg0831	PspA/IM30 family protein
		PCC7942_pg0911	addiction module toxin HicA family
UTEX3055_pg0929	hypothetical protein		
UTEX3055_pg0930	hypothetical protein		
UTEX3055_pg0931	AbrB family transcriptional regulator		
UTEX3055_pg0934	Protein of unknown function DUF2281		

Table 3-S3 (cont.): Genes unique to each genome set of legacy strains and UTEX 3055.

Pangenome Locus	Description	Pangenome Locus	Description
UTEX3055_pg0935	component of toxin-antitoxin system PIN domain		
UTEX3055_pg0963	transposase		
UTEX3055_pg0985	KTSC domain protein	PCC7942_pg0914	hypothetical protein
UTEX3055_pg0986	hypothetical protein DUF2188	PCC7942_pg0914	hypothetical protein
UTEX3055_pg1019	hypothetical protein	PCC7942_pg0947	hypothetical protein
UTEX3055_pg1020	hypothetical protein	PCC7942_pg0948	hypothetical protein
UTEX3055_pg1021	hypothetical protein	PCC7942_pg0949	hypothetical protein
UTEX3055_pg1022	hypothetical protein	PCC7942_pg0950	hypothetical protein
UTEX3055_pg1023	hypothetical protein	PCC7942_pg0951	hypothetical protein
		PCC7942_pg1019	UDPglucose 6-dehydrogenase
UTEX3055_pg1147	Protein of unknown function DUF1524		
		PCC7942_pg1113	DUF4351 domain protein
UTEX3055_pg1240	hypothetical protein	PCC7942_pg1169	hypothetical protein
UTEX3055_pg1241	hypothetical protein		
UTEX3055_pg1242	hypothetical protein		
UTEX3055_pg1294	Phage integrase family protein	PCC7942_pg1221	hypothetical protein
UTEX3055_pg1295	Response regulator receiver domain protein	PCC7942_pg1222	hypothetical protein
UTEX3055_pg1296	Phage integrase, N-terminal SAM-like domain	PCC7942_pg1223	phosphorylase
UTEX3055_pg1421	hypothetical protein		
UTEX3055_pg1422	VCBS repeat protein	PCC7942_pg1348	VCBS repeat protein

Table 3-S3 (cont.): Genes unique to each genome set of legacy strains and UTEX 3055.

Pangenome Locus	Description	Pangenome Locus	Description
UTEX3055_pg1423	Hemolysin-type calcium-binding repeat protein		
		PCC7942_pg1365	hypothetical protein
		PCC7942_pg1397	TIGR03032 family protein
		PCC7942_pg1398	Integrins alpha chain
UTEX3055_pg1499	hypothetical protein		
UTEX3055_pg1681	hypothetical protein		
UTEX3055_pg1682	hypothetical protein		
UTEX3055_pg1804	T5orf172 domain protein	PCC7942_pg1729	hypothetical protein
		PCC7942_pg1776	hypothetical protein
UTEX3055_pg1857	calcium-binding protein		
UTEX3055_pg1858	Glycosyltransferase		
UTEX3055_pg1859	hypothetical protein		
UTEX3055_pg1860	hypothetical protein		
UTEX3055_pg1861	hypothetical protein		
UTEX3055_pg1862	Ca2 -binding protein, RTX toxin-related		
UTEX3055_pg1863	SemiSWEET family sugar transporter		
UTEX3055_pg1864	nucleotidyltransferase family protein		
UTEX3055_pg1865	DUF86 domain protein		
UTEX3055_pg2116	tellurium resistance protein TerD		
		PCC7942_pg2051	Peptidase M20D, amidohydrolase
UTEX3055_pg2263	hypothetical protein		
UTEX3055_pg2264	hypothetical protein		
UTEX3055_pg2265	Type II transport protein GspH		

Table 3-S3 (cont.): Genes unique to each genome set of legacy strains and UTEX 3055.

Pangenome Locus	Description	Pangenome Locus	Description
UTEX3055_pg2266	hypothetical protein		
UTEX3055_pg2309	hypothetical protein	PCC7942_pg2221	hypothetical protein
UTEX3055_pg2310	hypothetical protein	PCC7942_pg2221 PCC7942_pg2222	hypothetical protein Type II toxin-antitoxin system HicA family toxin
UTEX3055_pg2311	Rieske [2Fe-2S] domain protein		
UTEX3055_pg2312	transcriptional regulator, TetR family		
UTEX3055_pg2313	type II TA system Phd/YefM family antitoxin	PCC7942_pg2222 PCC7942_pg2223	Type II toxin-antitoxin system HicA family toxin glutathione S-transferase
UTEX3055_pg2314	Txe/YoeB family addiction module toxin		
UTEX3055_pg2315	gamma-glutamylcyclotransferase		
UTEX3055_pg2316	MFS transporter	PCC7942_pg2223 PCC7942_pg2224	glutathione S-transferase cytidine deaminase
UTEX3055_pg2317	Ubiquinone biosynthesis protein Coq4		
UTEX3055_pg2318	hypothetical protein	PCC7942_pg2224	cytidine deaminase
UTEX3055_pg2319	transcriptional regulator, TetR family		
UTEX3055_pg2452	hypothetical protein		
UTEX3055_pg2453	hypothetical protein		
UTEX3055_pg2464	hypothetical protein		
UTEX3055_pg2466	hypothetical protein		
UTEX3055_pg2467	PhnD/SsuA/transferrin family substrate-binding protein		
UTEX3055_pg2469	Endonuclease YncB, thermonuclease family		
UTEX3055_pg2470	hypothetical protein		
UTEX3055_pg2471	OmpA family protein		
UTEX3055_pg2472	hypothetical protein		

Table 3-S3 (cont.): Genes unique to each genome set of legacy strains and UTEX 3055.

Pangenome Locus	Description	Pangenome Locus	Description
UTEX3055_pg2473	Uncharacterized membrane protein YeaQ/YmgE		
UTEX3055_pg2474	Peptidoglycan/xylan/chitin deacetylase, PgdA/CDA1		
UTEX3055_pg2475	hypothetical protein		
UTEX3055_pg2476	hypothetical protein		
UTEX3055_pg2477	hypothetical protein		
UTEX3055_pg2478	hypothetical protein		
UTEX3055_pg2479	hypothetical protein		
UTEX3055_pg2480	hypothetical protein		
UTEX3055_pg2481	hypothetical protein		
UTEX3055_pg2599	DUF4935 domain protein	PCC7942_pg2485	oxidoreductase
		PCC7942_pg2486	ATP-binding protein
		PCC7942_pg2555	hypothetical protein
		PCC7942_pg2556	DNA-cytosine methyltransferase
UTEX3055_pg2687	Uncharacterized conserved protein, contains HEPNdomain		
UTEX3055_pg2688	nucleotidyltransferase domain protein		
UTEX3055_pgB028	hypothetical protein	PCC7942_pgB028	DUF4062 domain-containing protein
UTEX3055_pgB030	DUF3854 domain-containing protein		
UTEX3055_pgB031	cation diffusion facilitator family transporter		
UTEX3055_pgB032	MBL fold metallo-hydrolase		
UTEX3055_pgB033	aliphatic sulfonate ABC transporter substrate-binding protein		

Table 3-S3 (cont.): Genes unique to each genome set of legacy strains and UTEX 3055.

Pangenome Locus	Description	Pangenome Locus	Description
UTEX3055_pgB034	FMNH2-dependent alkanesulfonate monooxygenase		
UTEX3055_pgB035	ABC transporter permease subunit		
UTEX3055_pgB036	ATP-binding cassette domain-containing protein		
UTEX3055_pgB037	arylsulfatase		
UTEX3055_pgB038	methyltransferase domain-containing protein		
UTEX3055_pgB039	anaerobic sulfatase maturase		
UTEX3055_pgB040	aliphatic sulfonate ABC transporter substrate-binding protein		
UTEX3055_pgB041	aliphatic sulfonate ABC transporter substrate-binding protein		
UTEX3055_pgB042	ABC transporter permease subunit		
UTEX3055_pgB043	ATP-binding cassette domain-containing protein		
UTEX3055_pgB044	methyltransferase domain-containing protein		
UTEX3055_pgB045	monooxygenase		
UTEX3055_pgB046	hypothetical protein		
UTEX3055_pgB047	hypothetical protein		
UTEX3055_pgB048	hypothetical protein		
UTEX3055_pgB049	hypothetical protein		
UTEX3055_pgB050	DUF1738 domain-containing protein		
UTEX3055_pgB051	hypothetical protein		

Table 3-S3 (cont.): Genes unique to each genome set of legacy strains and UTEX 3055.

Pangenome Locus	Description	Pangenome Locus	Description
UTEX3055_pgB052	type II toxin-antitoxin system RelE/ParE family toxin		
UTEX3055_pgB053	DNA-binding transcriptional regulator		
UTEX3055_pgB054	hypothetical protein		
UTEX3055_pgB055	ParA family protein		
UTEX3055_pgB056	tyrosine-type recombinase/integrase		
UTEX3055_pgB057	Abi family protein		
UTEX3055_pgB058	ParA family protein		
UTEX3055_pgB059	hypothetical protein		
UTEX3055_pgB060	restriction endonuclease		
UTEX3055_pgB061	DUF3387 domain-containing protein		
UTEX3055_pgB062	hypothetical protein		
UTEX3055_pgB063	ATP-binding protein		
UTEX3055_pgB064	recombinase family protein		
UTEX3055_pgB065	DUF1353 domain-containing protein		
UTEX3055_pgB066	hypothetical protein		
UTEX3055_pgB067	thermonuclease family protein		
UTEX3055_pgB068	hypothetical protein		
UTEX3055_pgB069	hypothetical protein		
UTEX3055_pgB070	ATP-binding protein		
UTEX3055_pgB071	metal-dependent hydrolase		
UTEX3055_pgB072	DUF87 domain-containing protein		
UTEX3055_pgB073	hypothetical protein		
UTEX3055_pgB074	BrnT family toxin		
UTEX3055_pgB075	BrnA antitoxin family protein		
UTEX3055_pgB076	metal ABC transporter permease		

Table 3-S3 (cont.): Genes unique to each genome set of legacy strains and UTEX 3055.

Pangenome Locus	Description	Pangenome Locus	Description
UTEX3055_pgB077	metal ABC transporter ATP-binding protein		
UTEX3055_pgB078	metal ABC transporter substrate-binding protein		
UTEX3055_pgB079	cadmium-translocating P-type ATPase		
UTEX3055_pgB080	HEAT repeat domain-containing protein		
UTEX3055_pgB082	hypothetical protein	PCC7942_pgB031	hypothetical protein
		PCC7942_pgB033	DUF1254 domain-containing protein
		PCC7942_pgB034	hypothetical protein
		PCC7942_pgB035	hypothetical protein
		PCC7942_pgB036	hypothetical protein
		PCC7942_pgB037	hypothetical protein
		PCC7942_pgB038	BrnT family toxin
		PCC7942_pgB039	hypothetical protein
		PCC7942_pgB040	DUF2442 domain-containing protein
		PCC7942_pgB041	DUF4351 domain-containing protein
		PCC7942_pgB042	NAD(P)-dependent alcohol dehydrogenase
		PCC7942_pgB043	DJ-1/PfpI family protein
		PCC7942_pgB044	GAF domain-containing protein
UTEX3055_pgB086	DUF305 domain-containing protein		
UTEX3055_pgB087	MFS transporter		
UTEX3055_pgB095	HEPN domain-containing protein	PCC7942_pgB053	type II toxin-antitoxin system PemK/MazF family toxin

Table 3-S3 (cont.): Genes unique to each genome set of legacy strains and UTEX 3055.

Pangenome Locus	Description	Pangenome Locus	Description
UTEX3055_pgB096	nucleotidyltransferase domain-containing protein	PCC7942_pgB054	hypothetical protein
		PCC7942_pgB055	hypothetical protein
UTEX3055_pgD001	cas1		
UTEX3055_pgD002	cas2		
UTEX3055_pgD003	CRISPR array		
UTEX3055_pgD004	recombinase family protein		
UTEX3055_pgD005	hypothetical protein		
UTEX3055_pgD006	type II toxin-antitoxin system VapC family toxin		
UTEX3055_pgD007	hypothetical protein		
UTEX3055_pgD008	DUF4351 domain-containing protein		
UTEX3055_pgD009	DUF4351 domain-containing protein		
UTEX3055_pgD010	hypothetical protein		
UTEX3055_pgD011	translesion error-prone DNA polymerase V autoproteolytic subunit		
UTEX3055_pgD012	DUF433 domain-containing protein		
UTEX3055_pgD013	DUF2281 domain-containing protein		
UTEX3055_pgD014	type II toxin-antitoxin system PemK/MazF family toxin		
UTEX3055_pgD015	recombinase family protein		
UTEX3055_pgD016	hypothetical protein		
UTEX3055_pgD017	hypothetical protein		
UTEX3055_pgD018	chromosome partitioning protein ParB		
UTEX3055_pgD019	ParA family protein		
UTEX3055_pgD020	WYL domain-containing protein		
UTEX3055_pgD021	cas6		
UTEX3055_pgD022	cas3		
UTEX3055_pgD023	cas8a1		
UTEX3055_pgD024	cas7i		
UTEX3055_pgD025	cas5		

Table 3-S4: Toxin-Antitoxin Systems (TAS) in *S. elongatus*. Toxin components are underlined and orphan genes italicized. Toxin components with alterations in UTEX 3055 are highlighted.

Pangenome ID	Description	Found in
24	hypothetical protein	UTEX 3055
25	hypothetical protein	UTEX 3055
37	DUF86 domain-containing protein	Legacy Strains
38	nucleotidyltransferase family protein	Legacy Strains
52	putative transcriptional regulator	UTEX 3055
53	BrnT family toxin	UTEX 3055
135	*SMI1-KNR4 cell-wall*	UTEX 3055
202	DUF4435 domain-containing protein	Legacy Strains
203	AAA family ATPase	Legacy Strains
300	nucleotidyl transferase	All
301	DUF86 domain-containing protein	All
332	addiction module component	All
333	type II toxin-antitoxin system RelE/ParE family toxin	All
441	hypothetical protein	All
442	Ribonuclease toxin, BrnT, of type II toxin-antitoxin system	All
533	*hypothetical protein*	All
571	hypothetical protein	UTEX 3055
572	Uncharacterized conserved protein, DUF433 family	UTEX 3055
1024	Protein of unknown function DUF2281	UTEX 3055
1025	PIN domain nuclease, a component of toxin-antitoxin system PIN domain	UTEX 3055
1073	addiction module toxin HicA family	Legacy Strains
1074	type II toxin-antitoxin system HicB family antitoxin	All
1443	PIN domain-containing protein	All
1444	prevent-host-death family protein	All
1447	toxin YoeB	All
1448	antitoxin YefM	All
1454	*tRNAfMet-specific endonuclease VapC*	All
1987	hypothetical protein	UTEX 3055
1988	Uncharacterized conserved protein, contains HEPNdomain	UTEX 3055
2436	*Type II toxin-antitoxin system HicA family toxin*	All
2443	prevent-host-death family protein	UTEX 3055
2444	toxin-antitoxin system, toxin component, Txe/YoeB family	UTEX 3055
2656	*addiction module antidote protein, HigA family*	All

Table 3-S4 (cont.): Toxin-Antitoxin Systems (TAS) in *S. elongatus*.

Pangenome ID	Description	Found in
2731	*hypothetical protein*	UTEX 3055
2973	hypothetical protein	All
2974	type II toxin-antitoxin system VapC family toxin	All
2995	BrnT family toxin	All
2996	conserved hypothetical protein	All
3023	type II toxin-antitoxin system RelE/ParE family toxin	UTEX 3055
3024	DNA-binding transcriptional regulator	UTEX 3055
3045	BrnT family toxin	UTEX 3055
3046	BrnA antitoxin family protein	UTEX 3055
3062	BrnT family toxin	Legacy Strains
3063	hypothetical protein	Legacy Strains
3079	*type II toxin-antitoxin system PemK/MazF family toxin*	Legacy Strains
3082	HEPN domain-containing protein	UTEX 3055
3083	nucleotidyltransferase domain-containing protein	UTEX 3055
3094	*type II toxin-antitoxin system VapC family toxin*	UTEX 3055
3100	*DUF433 domain-containing protein*	UTEX 3055
3102	*type II toxin-antitoxin system PemK/MazF family toxin*	UTEX 3055

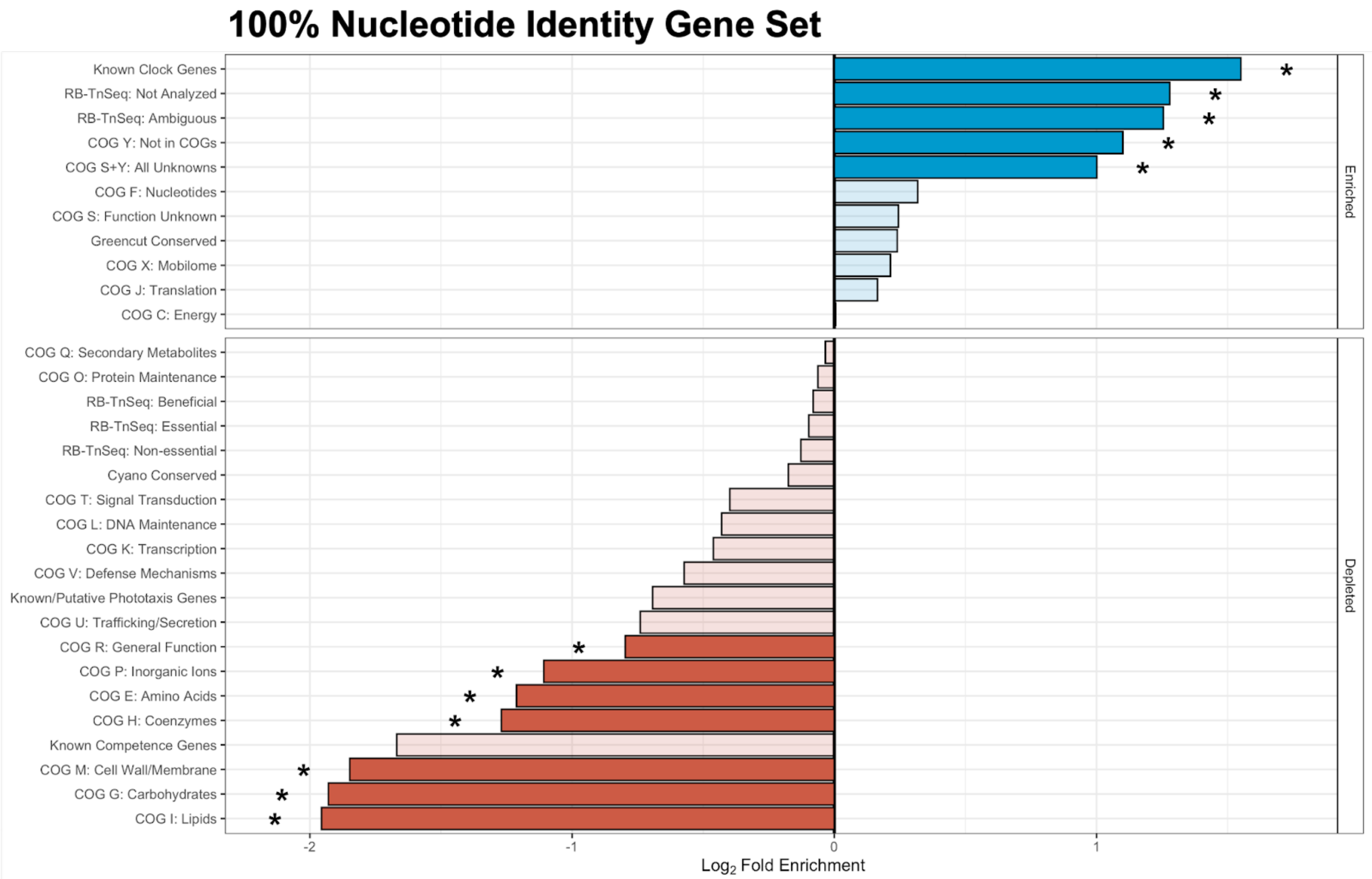


Figure 3-S6: Gene set enrichment analysis of the 100% shared nucleotide identity pangeome homolog gene set. Categories that are significantly enriched or depleted are marked with darker category colors and asterisks.

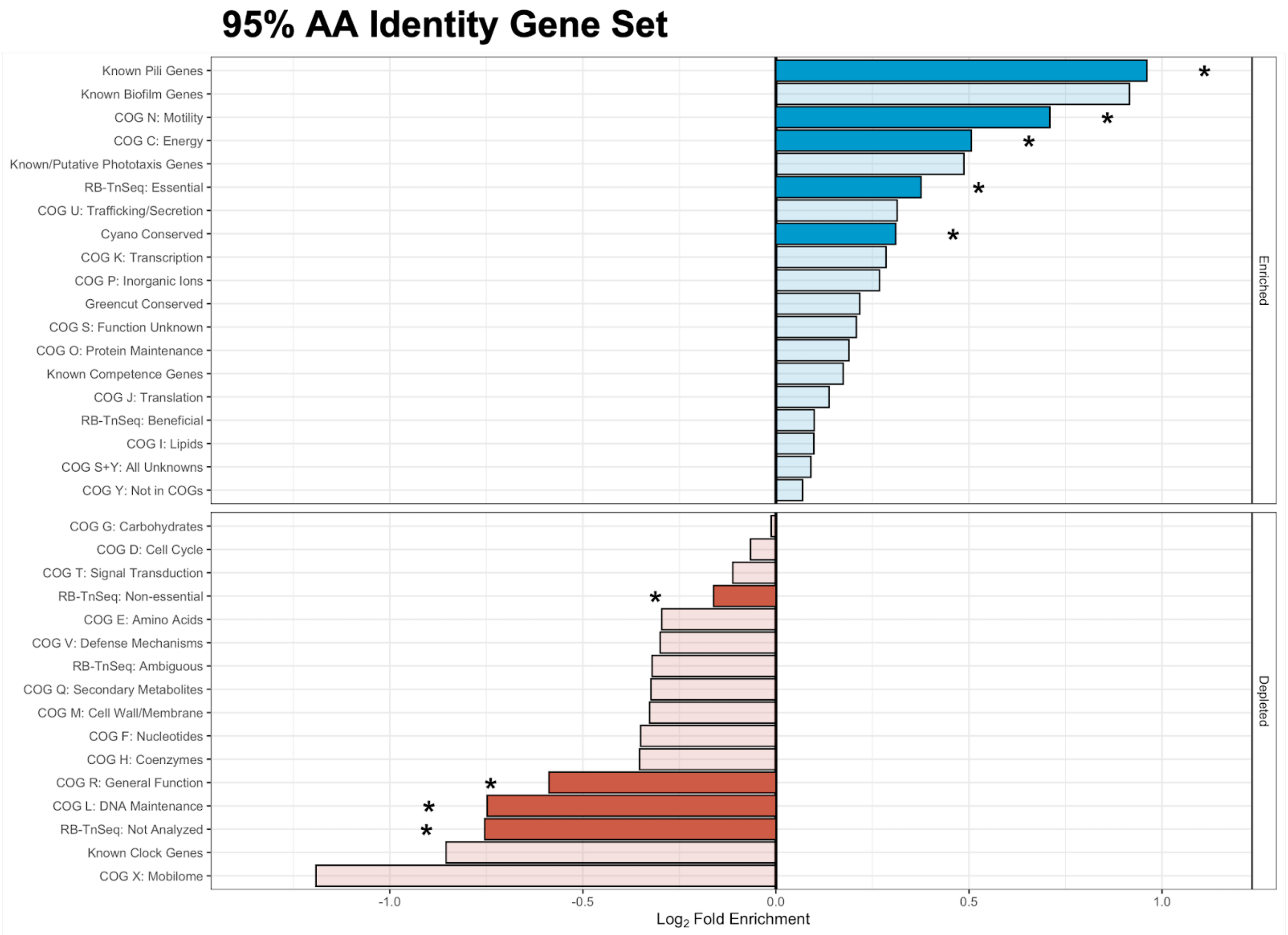


Figure 3-S7: Gene set enrichment analysis of the 95% shared amino acid identity pangeome homolog gene set. The GSEA shows enrichment in pili genes, motility and energy COG category genes as well as conserved and essential gene categories. Categories that are significantly enriched or depleted are marked with darker category colors and asterisks.

Table 3-S5: Isolation and sequencing history of *S. elongatus* strains. The table shows the documented year of isolation, year the strain entered a culture collection, and the year the sequence was published in GenBank. Versions of the same strain in different culture collections are highlighted. Strain designators for culture collections: PCC = Pasteur Culture Collection, Paris, France; FACHB = Freshwater Algae Culture Collection at the Institute for Hydrobiology, Wuhan, China; AMC = Golden Lab Strain Collection; UTEX = Culture Collection of Algae at the University of Texas at Austin, USA.

*The sequence of AMC06 has not been submitted to GenBank, see Table 3-1 and Supplementary File 3-2 for differences between PCC 7942 and AMC06.

Strain	Isolation Location	Isolation Year	Archive Year	Sequence Submission
PCC 6301	Waller Creek, Texas, USA	1952	1963	2004 & 2021
FACHB-242			-	2020
PCC 6311	California, USA	1963	1963	2021
PCC 7942	California, USA	1973	1979	2005
AMC06	California, USA		1988	2019*
FACHB-805	California, USA		-	2020
PCC 7943	Unknown	Unknown	1979	2021
FACHB-1061	China	2007	-	2020
UTEX 2973	UTEX 625 (PCC 6301)	2011	2012	2013
UTEX 3055	Waller Creek, Texas, USA	2013	2018	2018

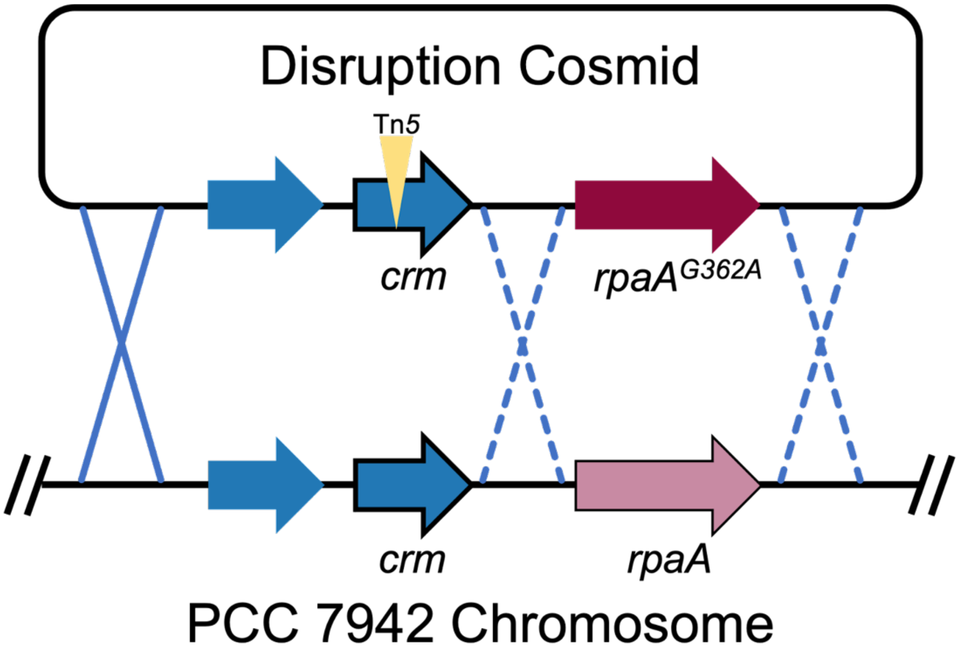
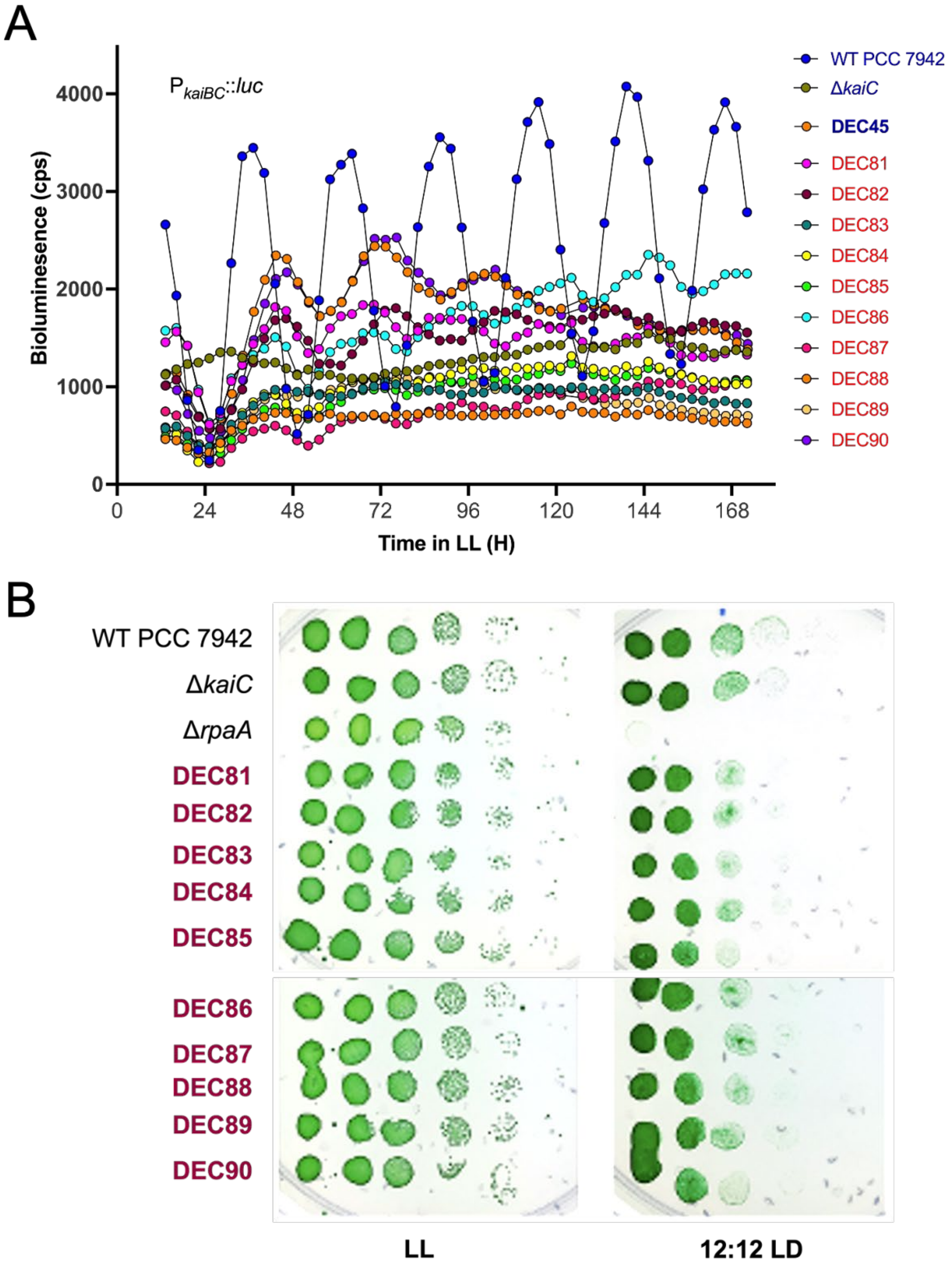


Figure 3-S8: Schematic of homologous recombination events in *crm* mutant construction. Homologous recombination events of the WT PCC 7942 chromosome with the UGS *crm*:Tn*5* disruption cosmid can produce either rhythmic mutants with the *crm*:Tn*5* insertion or arrhythmic mutants that also include *rpaA*^{G362A}.

Figure 3-S9: Suppressor screen mutants of RpaA-121Q. The arrhythmic PCC 7942 RpaA-R121Q CRISPR/Cpf1 edited strain DEC45 was selected for suppressors that restore LD survival (B), and strains that emerged from this selection were screened for restoration of the rhythmic phenotype. Dilution series of strains were grown in constant light (LL) or in 12 hour light/dark cycles (12:12LD) for 48 hours to assess LD fitness. Bioluminescence from strains carrying a P_{kaiBC}-*luc* reporter at NS2 was recorded as an assay for circadian rhythms of gene expression. LL, constant light after entrainment in a 12-h light:12-h dark cycle.


A
P_kaiBC::luc
Bioluminesence (cps)
0
1000
2000
3000
4000
0
24
48
72
96
120
144
168
Time in LL (H)
WT PCC 7942
ΔkaiC
DEC45
DEC81
DEC82
DEC83
DEC84
DEC85
DEC86
DEC87
DEC88
DEC89
DEC90
B
WT PCC 7942
ΔkaiC
ΔrpaA
DEC81
DEC82
DEC83
DEC84
DEC85
DEC86
DEC87
DEC88
DEC89
DEC90
LL
12:12 LD

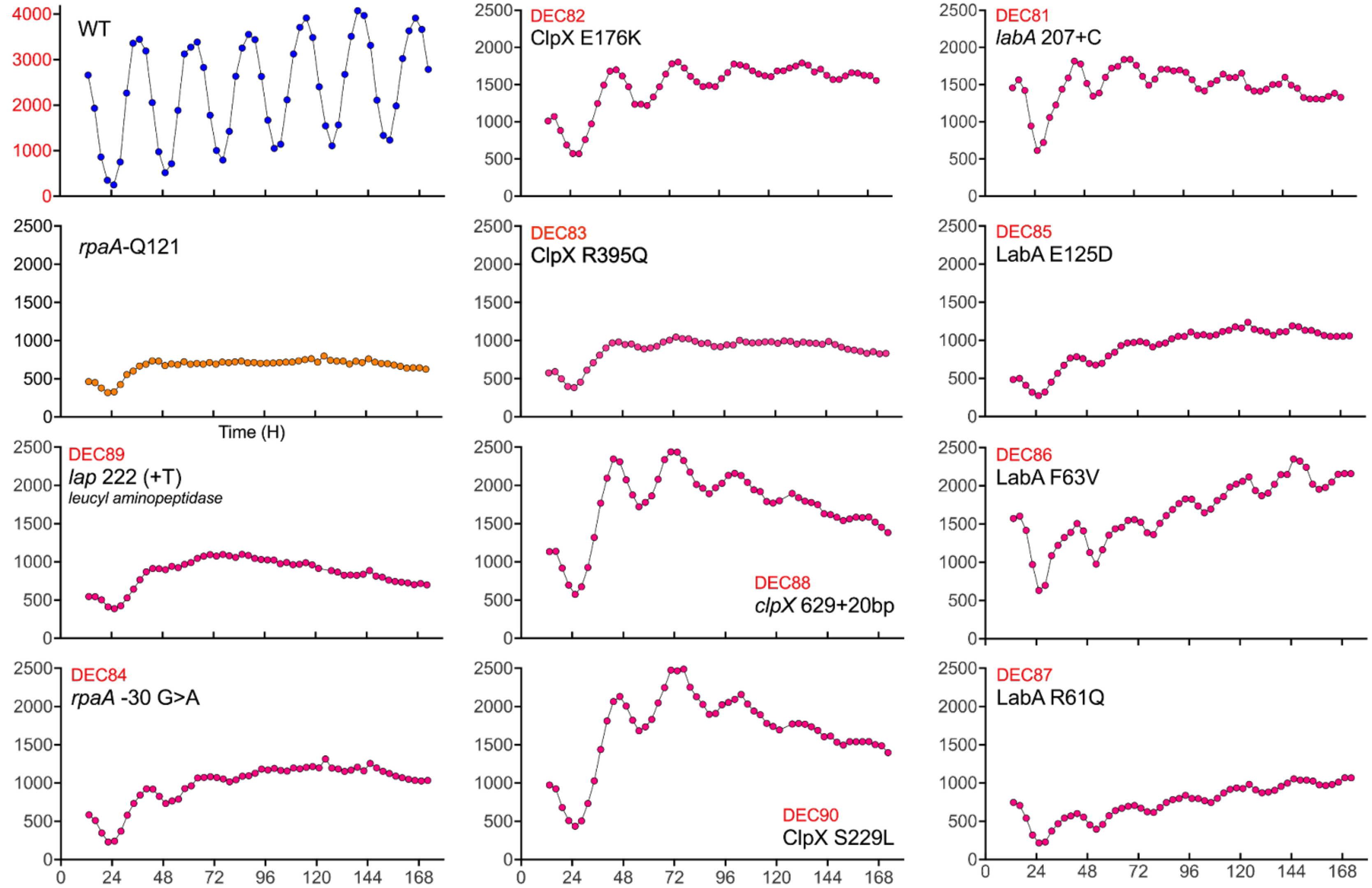
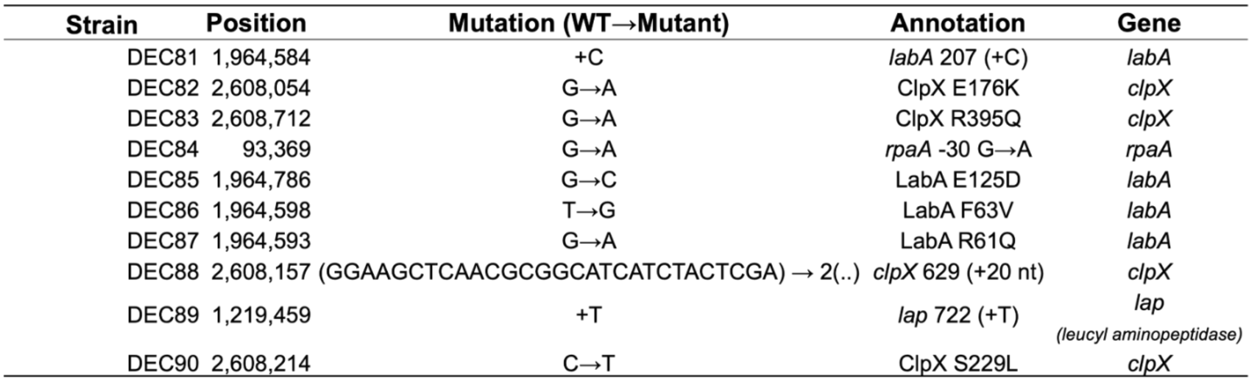


Figure 3-S10: Measurement of rhythmic gene expression in suppressor mutants. Nine of ten suppressor mutants of RpaA-Q211 show restored rhythmic gene expression. Bioluminescence from strains carrying a P_{kaiBC}-*luc* reporter at NS2 was recorded as an assay for circadian rhythms of gene expression. LL, constant light after entrainment in a 12-h light:12-h dark cycle.

Table 3-S6: Description of second-site mutations in suppressor mutants of RpaA-Q121.



Strain	Position	Mutation (WT→Mutant)	Annotation	Gene
DEC81	1,964,584	+C	*labA* 207 (+C)	*labA*
DEC82	2,608,054	G→A	ClpX E176K	*clpX*
DEC83	2,608,712	G→A	ClpX R395Q	*clpX*
DEC84	93,369	G→A	*rpaA* -30 G→A	*rpaA*
DEC85	1,964,786	G→C	LabA E125D	*labA*
DEC86	1,964,598	T→G	LabA F63V	*labA*
DEC87	1,964,593	G→A	LabA R61Q	*labA*
DEC88	2,608,157	(GGAAGCTCAACGCGGCATCATCTACTCGA) → 2(..)	*clpX* 629 (+20 nt)	*clpX*
DEC89	1,219,459	+T	*lap* 722 (+T)	*lap* *(leucyl aminopeptidase)*
DEC90	2,608,214	C→T	ClpX S229L	*clpX*

Table 3-S7: Strains, plasmids, and primer sequences used Chapter 3.

Strain	Genotype	Source
AMC06	WT *S. elongatus* PCC 7942	Lab collection
AMC18	WT *S. elongatus* PCC 6301	Lab collection
AMC2388	WT *S. elongatus* UTEX 3055	Lab collection
AMC2370	AMC06 / NS1-riboB-Synpcc7942_0765	This work
AMC2372	AMC06 / NS1-riboF-Synpcc7942_0765	This work
AMC2366	AMC06 / NS1-riboB-Synpcc7942_0766	This work
AMC2368	AMC06 / NS1-riboF-Synpcc7942_0766	This work
AMC2142	PCC 7942 with a 50kb deletion of the prophage	(45)
AMC2301 (D1K3)	AMC06 / Δprophage::Km	This work
AMC2343	AMC06 / ΔDs7::Km	This work
Synpcc7942_0756:Tn*5*	Tn*5* 21-F10 in AMC06	This work
AMC2399	AMC06 / ΔSynpcc7942_0757::Km	This work
AMC2412	AMC06 / ΔSynpcc7942_0759::Km	This work
AMC2410	AMC06 /ΔSynpcc7942_0760::Km	This work
AMC2344	AMC06 / ΔSynpcc7942_0761::Km	This work
AMC2346	AMC06 / ΔSynpcc7942_0762::Km	This work
D1K3 + Ds7	D1K3 / NS1-P*conII*::Ds7	This work
D1K3 +0759-0760	D1K3 / NS1-P*conII*::0759-0760	This work
UTEX3055_pg2266:Tn5	pKMW7 Tn*5* in UTEX3055_pg2266	This work
Δ*nPT*	AMC2388 / ΔUTEX3055_pg2263-2266::Km	This work
Δ*nPT* + pg_2266	ΔnPT / NS1-P*trc*::UTEX3055_pg2266	This work
Δ*nPT* + pg_2265-6	ΔnPT / NS1-P*trc*::UTEX3055_pg2265-2266	This work
Δ*nPT* + pg_2264-6	ΔnPT / NS1-P*trc*::UTEX3055_pg2264-2266	This work
Δ*nPT* + pg_2263-6	ΔnPT / NS1-P*trc*::UTEX3055_pg2263-2266	This work
AMC541	AMC06 / NS2-PkaiBC::luc	Lab collection
AMC704	AMC541 / Δ*kaiC*	Lab collection
ΔrpaA	*rpaA*::Gm / NSII-P$_{kaiBC}$::luc	(65)
AMC2609	*rpaA*(G362->A) / NS1I-PkaiBC::luc	(65)
AMC2610	*rpaA*(G11->T) / NS1I-PkaiBC::luc	This work
AMC2611	*rpaA*(G11->T; G362->A) / NS1I-PkaiBC::luc	This work
AM5643	CRISPR edited *rpaA*(G362->A; R121Q)	This work
AM5644	CRISPR edited *rpaA*(G11->T; R4L)	This work
AM5652	CRISPR edited *rpaA*(G362->A & G11->T; R4L & R121Q)	This work
AM4523	Deletion construct containing *rpaA*::Gm flanked by *S. elongatus* gDNA	This work

Table 3-S7 (cont.): Strains, plasmids, and primer sequences used Chapter 3.

Plasmid	Description	Source
pAM5206	riboB-Synpcc7942_0765 expressed from NS1	This work
pAM5207	riboF-Synpcc7942_0765 expressed from NS1	This work
pAM5198	riboB-Synpcc7942_0766 expressed from NS1	This work
pAM5199	riboF-Synpcc7942_0766 expressed from NS1	This work
pAM4998	Deletion vector to replace PCC 7942 prophage with Km	This work
pAM5013	Tagging vector to add sucrose sensitivity to NS3 within PCC 7942 prophage	This work
pAM5113	Deletion vector to replace Ds7 with Km	This work
pAM5194	Deletion vector to replace Synpcc7942_0757 with Km	This work
pAM5195	Deletion vector to replace Synpcc7942_0759 with Km	This work
pAM5196	Deletion vector to replace Synpcc7942_0760 with Km	This work
pAM5124	Deletion vector to replace Synpcc7942_0761 with Km	This work
pAM5114	Deletion vector to replace Synpcc7942_0762 with Km	This work
pAM5259	PconII-Ds7 expressed from NS1	This work
pAM5260	PconII-Synpcc7942_0759-0760 expressed from NS1	This work
pMA04	DH5a; deltanLPS plasmid	This work
pMA06	P*trc*::UTEX3055_pg2266 expressed from NS1	This work
pMA07	P*trc*::UTEX3055_pg2265-2266 expressed from NS1	This work
pMA08	P*trc*::UTEX3055_pg2264-2266 expressed from NS1	This work
pMA09	P*trc*::UTEX3055_pg2263-2266 expressed from NS1	This work
pDE32	pSL2680 with gRNA targetting *rpaA* and HDR encoding *rpaA*(G362->A; R121Q)	This work
pDE33	pSL2680 with gRNA targetting *rpaA* and HDR encoding *rpaA*(G11->T; R4L)	This work
pDE36	pSL2680 with gRNA targeting rpaA and HDR encoding rpaA (G11->T)	This work
pDE44	pSL2680 with gRNA targetting *rpaA* and HDR encoding *rpaA*(G362->A & G11->T; R4L & R121Q)	This work

Table 3-S7 (cont.): Strains, plasmids, and primer sequences used Chapter 3.

Primer	Sequence	Source
Phage F1	CGCAAGTTAGGCGCTGATATCAAGC	This work
Phage R1	GCGATTTTGGCCCCTACGAC	This work
Phage F2	GTACAACACTCACCTGGCAC	This work
Phage R2	GCACAGAAATCGGCAGGCGAC	This work
ARB1	GGCCACGCGTCGACTAGTACNNNNNNNNNNATGGC	(18)
ARB3	GGCCACGCGTCGACTAGTACNNNNNNNNNNGATG	(18)
ARB2	GGCCACGCGTCGACTAGTAC	(18)
pRL27_minus1	AGGAACACTTAACGGCTGAC	(18)
pRL27_IE_rev1	ACTGAGAAGCCCTTAGAGCC	(18)
rpaA_gRNA_F	AGATGCTGAACAGCTAAAGCCTGA	This work
rpaA_gRNA_R	AGACTCAGGCTTTAGCTGTTCAGC	This work
rpaA_HDR-UP_F	CATTTTTTTGTCTAGCTTTAATGCGGTAGTTGGTACCGAGTCCTGAGCTGCTACTGCC	This work
rpaA_HDR-UP_R	GGTCAGGCTTTAGCTGTGCAGCTCGTTCCCGACCTGAT	This work
rpaA_HDR-DWN_F	ATCAGGTCGGGAACGAGCTGCACAGCTAAAGCCTGACC	This work
rpaA_HDR-DWN_R	GCCCGGATTACAGATCCTCTAGAGTCGACGGTACCGCCATGTCAAACTCAATCAAGCG	This work
rpaA_cPCR_F	GAACAGGCTGAAACGATGGC	This work
rpaA_cPCR_R	GGATTGAGTAATTTGATGGTTGACTGC	This work

References

1. Falkowski PG. 1994. The role of phytoplankton photosynthesis in global biogeochemical cycles. Photosynth Res 39:235–258.

2. Callieri C. 2008. Picophytoplankton in Freshwater Ecosystems: The Importance of Small-Sized Phototrophs. frer 1:1–28.

3. Mager DM. 2010. Carbohydrates in cyanobacterial soil crusts as a source of carbon in the southwest Kalahari, Botswana. Soil Biol Biochem 42:313–318.

4. Hays SG, Ducat DC. 2015. Engineering cyanobacteria as photosynthetic feedstock factories. Photosynth Res 123:285–295.

5. Kondo T, Strayer CA, Kulkarni RD, Taylor W, Ishiura M, Golden SS, Johnson CH. 1993. Circadian rhythms in prokaryotes: luciferase as a reporter of circadian gene expression in cyanobacteria. Proc Natl Acad Sci U S A 90:5672–5676.

6. Zouni A, Witt HT, Kern J, Fromme P, Krauss N, Saenger W, Orth P. 2001. Crystal structure of photosystem II from *Synechococcus elongatus* at 3.8 A resolution. Nature 409:739–743.

7. Jordan P, Fromme P, Witt HT, Klukas O, Saenger W, Krauss N. 2001. Three-dimensional structure of cyanobacterial photosystem I at 2.5 A resolution. Nature 411:909–917.

8. Golden SS, Brusslan J, Haselkorn R. 1987. [12] Genetic engineering of the cyanobacterial chromosome, p. 215–231. *In* Methods in Enzymology. Academic Press.

9. Golden SS. 2018. The international journeys and aliases of *Synechococcus elongatus*. N Z J Bot 1–6.

10. Schatz D, Nagar E, Sendersky E, Parnasa R, Zilberman S, Carmeli S, Mastai Y, Shimoni E, Klein E, Yeger O, Reich Z, Schwarz R. 2013. Self-suppression of biofilm formation in the cyanobacterium *Synechococcus elongatus*. Environ Microbiol 15:1786–1794.

11. Yang Y, Lam V, Adomako M, Simkovsky R, Jakob A, Rockwell NC, Cohen SE, Taton A, Wang J, Lagarias JC, Wilde A, Nobles DR, Brand JJ, Golden SS. 2018. Phototaxis in a wild isolate of the cyanobacterium *Synechococcus elongatus*. Proc Natl Acad Sci U S A 115:E12378–E12387.

12. Stanier RY, Kunisawa R, Mandel M, Cohen-Bazire G. 1971. Purification and properties of unicellular blue-green algae (order *Chroococcales*). Bacteriol Rev 35:171–205.

13. Grigorieva GA, Shestakov SV. 1976. Application of the genetic transformation method for taxonomic analysis of unicellular blue-green algae, p. 220–221. *In* Proc 2nd Int Symp Photosynthetic Prokaryotes.

14. Shestakov SV, Khyen NT. 1970. Evidence for genetic transformation in blue-green alga *Anacystis nidulans*. Mol Gen Genet 107:372–375.

15. Yu J, Liberton M, Cliften PF, Head RD, Jacobs JM, Smith RD, Koppenaal DW, Brand JJ, Pakrasi HB. 2015. *Synechococcus elongatus* UTEX 2973, a fast growing cyanobacterial chassis for biosynthesis using light and CO2. Sci Rep 5:8132.

16. Vijayan V, Jain IH, O'Shea EK. 2011. A high resolution map of a cyanobacterial transcriptome. Genome Biol 12:1–18.

17. Billis K, Billini M, Tripp HJ, Kyrpides NC, Mavromatis K. 2014. Comparative transcriptomics between *Synechococcus* PCC 7942 and *Synechocystis* PCC 6803 provide insights into mechanisms of stress acclimation. PLoS One 9:e109738.

18. Rubin BE, Wetmore KM, Price MN, Diamond S, Shultzaberger RK, Lowe LC, Curtin G, Arkin AP, Deutschbauer A, Golden SS. 2015. The essential gene set of a photosynthetic organism. Proc Natl Acad Sci U S A 112:E6634–43.

19. Sugita C, Ogata K, Shikata M, Jikuya H, Takano J, Furumichi M, Kanehisa M, Omata T, Sugiura M, Sugita M. 2007. Complete nucleotide sequence of the freshwater unicellular cyanobacterium Synechococcus elongatus PCC 6301 chromosome: gene content and organization. Photosynth Res 93:55–67.

20. Ungerer J, Wendt KE, Hendry JI, Maranas CD, Pakrasi HB. 2018. Comparative genomics reveals the molecular determinants of rapid growth of the cyanobacterium *Synechococcus elongatus* UTEX 2973. Proc Natl Acad Sci U S A 115:E11761–E11770.

21. Li S, Sun T, Xu C, Chen L, Zhang W. 2018. Development and optimization of genetic toolboxes for a fast-growing cyanobacterium *Synechococcus elongatus* UTEX 2973. Metab Eng 48:163–174.

22. Daniell H, Sarojini G, McFadden BA. 1986. Transformation of the cyanobacterium *Anacystis nidulans* 6301 with the *Escherichia coli* plasmid pBR322. Proc Natl Acad Sci U S A 83:2546–2550.

23. Daniell H, McFadden BA. 1986. Characterization of DNA uptake by the cyanobacterium *Anacystis nidulans*. Mol Gen Genet 204:243–248.

24. Takeshima Y, Sugiura M, Hagiwara H. 1994. A novel expression vector for the cyanobacterium, *Synechococcus* PCC 6301. DNA Res 1:181–189.

25. Golden SS, Nalty MS, Cho DS. 1989. Genetic relationship of two highly studied *Synechococcus* strains designated *Anacystis nidulans*. J Bacteriol 171:24–29.

26. Taton A, Erikson C, Yang Y, Rubin BE, Rifkin SA, Golden JW, Golden SS. 2020. The circadian clock and darkness control natural competence in cyanobacteria. Nat Commun 11:1688.

27. Jaiswal D, Sengupta A, Sengupta S, Madhu S, Pakrasi HB, Wangikar PP. 2020. A novel cyanobacterium *Synechococcus elongatus* PCC 11802 has distinct genomic and metabolomic characteristics compared to its neighbor PCC 11801. Sci Rep 10:191.

28. Jaiswal D, Sengupta A, Sohoni S, Sengupta S, Phadnavis AG, Pakrasi HB, Wangikar PP. 2018. Genome features and biochemical characteristics of a robust, fast growing and naturally transformable cyanobacterium *Synechococcus elongatus* PCC 11801 isolated from India. Sci Rep 8:16632.

29. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. 2018. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat Commun 9:5114.

30. Akiyama H, Kanai S, Hirano M, Miyasaka H. 1998. A novel plasmid recombination mechanism of the marine cyanobacterium *Synechococcus* sp. PCC7002. DNA Res 5:327–334.

31. Lau RH, Doolittle WF. 1979. Covalently closed circular DNAs in closely related unicellular cyanobacteria. J Bacteriol 137:648–652.

32. Chauvat F, Astier C, Vedel F, Joset-Espardellier F. 1983. Transformation in the cyanobacterium *Synechococcus* R2: improvement of efficiency; role of the pUH24 plasmid. Mol Gen Genet 191:39–45.

33. Van der Plas J, Oosterhoff-Teertstra R, Borrias M, Weisbeek P. 1992. Identification of replication and stability functions in the complete nucleotide sequence of plasmid pUH24 from the cyanobacterium *Synechococcus* sp.PCC 7942. Mol Microbiol 6:653–664.

34. Chen Y, Taton A, Go M, London RE, Pieper LM, Golden SS, Golden JW. 2016. Self-replicating shuttle vectors based on pANS, a small endogenous plasmid of the unicellular cyanobacterium *Synechococcus elongatus* PCC 7942. Microbiology 162:2029–2041.

35. Chen Y, Holtman CK, Magnuson RD, Youderian PA, Golden SS. 2008. The complete sequence and functional analysis of pANL, the large plasmid of the unicellular freshwater cyanobacterium *Synechococcus elongatus* PCC 7942. Plasmid 59:176–192.

36. Makarova KS, Koonin EV. 2015. Annotation and Classification of CRISPR-Cas Systems, p. 47–75. *In* Lundgren, M, Charpentier, E, Fineran, PC (eds.), CRISPR: Methods and Protocols. Springer New York, New York, NY.

37. Makarova KS, Wolf YI, Alkhnbashi OS, Costa F, Shah SA, Saunders SJ, Barrangou R, Brouns SJJ, Charpentier E, Haft DH, Horvath P, Moineau S, Mojica FJM, Terns RM, Terns MP, White MF, Yakunin AF, Garrett RA, van der Oost J, Backofen R, Koonin EV. 2015. An updated evolutionary classification of CRISPR-Cas systems. Nat Rev Microbiol 13:722–736.

38. Makarova KS, Wolf YI, Iranzo J, Shmakov SA, Alkhnbashi OS, Brouns SJJ, Charpentier E, Cheng D, Haft DH, Horvath P, Moineau S, Mojica FJM, Scott D, Shah SA, Siksnys V, Terns MP, Venclovas Č, White MF, Yakunin AF, Yan W, Zhang F, Garrett RA, Backofen R, van der Oost J, Barrangou R, Koonin EV. 2020. Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants. Nat Rev Microbiol 18:67–83.

39. Shmakov SA, Sitnik V, Makarova KS, Wolf YI, Severinov KV, Koonin EV. 2017. The CRISPR spacer space is dominated by sequences from species-specific mobilomes. MBio 8.

40. Hein S, Scholz I, Voß B, Hess WR. 2013. Adaptation and modification of three CRISPR loci in two closely related cyanobacteria. RNA Biol 10:852–864.

41. Huang S, Wang K, Jiao N, Chen F. 2012. Genome sequences of siphoviruses infecting marine *Synechococcus* unveil a diverse cyanophage group and extensive phage–host genetic exchanges. Environ Microbiol 14:540–558.

42. Niederholtmeyer H, Wolfstädter BT, Savage DF, Silver PA, Way JC. 2010. Engineering cyanobacteria to synthesize and export hydrophilic products. Appl Environ Microbiol 76:3462–3466.

43. Fouts DE. 2006. Phage_Finder: automated identification and classification of prophage regions in complete bacterial genome sequences. Nucleic Acids Res 34:5839–5851.

44. Chénard C, Chan AM, Vincent WF, Suttle CA. 2015. Polar freshwater cyanophage S-EIV1 represents a new widespread evolutionary lineage of phages. ISME J 9:2046–2058.

45. Watanabe S, Ohbayashi R, Shiwa Y, Noda A, Kanesaki Y, Chibazakura T, Yoshikawa H. 2012. Light-dependent and asynchronous replication of cyanobacterial multi-copy chromosomes. Mol Microbiol 83:856–865.

46. Guerreiro ACL, Benevento M, Lehmann R, van Breukelen B, Post H, Giansanti P, Maarten Altelaar AF, Axmann IM, Heck AJR. 2014. Daily rhythms in the cyanobacterium *Synechococcus elongatus* probed by high-resolution mass spectrometry-based proteomics reveals a small defined set of cyclic proteins. Mol Cell Proteomics 13:2042–2055.

47. Nilsson AS, Haggård-Ljungquist E. 2007. Evolution of P2-like phages and their impact on bacterial evolution. Res Microbiol 158:311–317.

48. Fillol-Salom A, Martínez-Rubio R, Abdulrahman RF, Chen J, Davies R, Penadés JR. 2018. Phage-inducible chromosomal islands are ubiquitous within the bacterial universe. ISME J 12:2114–2128.

49. Bobay L-M, Touchon M, Rocha EPC. 2014. Pervasive domestication of defective prophages by bacteria. Proc Natl Acad Sci U S A 111:12127–12132.

50. Howard-Varona C, Hargreaves KR, Abedon ST, Sullivan MB. 2017. Lysogeny in nature: mechanisms, impact and ecology of temperate phages. ISME J 11:1511–1520.

51. Collier JL, Grossman AR. 1994. A small polypeptide triggers complete degradation of light-harvesting phycobiliproteins in nutrient-deprived cyanobacteria. EMBO J 13:1039–1047.

52. Gao E-B, Gui J-F, Zhang Q-Y. 2012. A novel cyanophage with a cyanobacterial nonbleaching protein A gene in the genome. J Virol 86:236–245.

53. Flores-Uribe J, Philosof A, Sharon I, Fridman S, Larom S, Béjà O. 2019. A novel uncultured marine cyanophage lineage with lysogenic potential linked to a putative marine *Synechococcus* “relic” prophage. Environ Microbiol Rep 11:598–604.

54. Nadel O, Rozenberg A, Flores-Uribe J, Larom S, Schwarz R, Béjà O. 2019. An uncultured marine cyanophage encodes an active phycobilisome proteolysis adaptor protein NblA. Environ Microbiol Rep 11:848–854.

55. Song S, Wood TK. 2020. A primary physiological role of toxin/antitoxin systems is phage inhibition. Front Microbiol 11:1895.

56. Xia K, Bao H, Zhang F, Linhardt RJ, Liang X. 2019. Characterization and comparative analysis of toxin-antitoxin systems in *Acetobacter pasteurianus*. J Ind Microbiol Biotechnol 46:869–882.

57. Pandey DP, Gerdes K. 2005. Toxin-antitoxin loci are highly abundant in free-living but lost from host-associated prokaryotes. Nucleic Acids Res 33:966–976.

58. Kunin V, Ouzounis CA. 2003. The balance of driving forces during genome evolution in prokaryotes. Genome Res 13:1589–1594.

59. Xu C, Wang B, Yang L, Zhongming Hu L, Yi L, Wang Y, Chen S, Emili A, Wan C. 2021. Global landscape of native protein complexes in *Synechocystis* sp. PCC 6803. Genomics Proteomics Bioinformatics https://doi.org/10.1016/j.gpb.2020.06.020.

60. Simkovsky R, Effner EE, Iglesias-Sánchez MJ, Golden SS. 2016. Mutations in novel lipopolysaccharide biogenesis genes confer resistance to amoebal grazing in *Synechococcus elongatus*. Appl Environ Microbiol 82:2738–2750.

61. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc 10:845–858.

62. Chen M-Y, Teng W-K, Zhao L, Hu C-X, Zhou Y-K, Han B-P, Song L-R, Shu W-S. 2021. Comparative genomics reveals insights into cyanobacterial evolution and habitat adaptation. ISME J 15:211–227.

63. Holtman CK, Chen Y, Sandoval P, Gonzales A, Nalty MS, Thomas TL, Youderian P, Golden SS. 2005. High-throughput functional analysis of the *Synechococcus elongatus* PCC 7942 genome. DNA Res 12:103–115.

64. Boyd JS, Bordowitz JR, Bree AC, Golden SS. 2013. An allele of the *crm* gene blocks cyanobacterial circadian rhythms. Proc Natl Acad Sci U S A 110:13950–13955.

65. Chavan AG, Swan JA, Heisler J, Sancar C, Ernst DC, Fang M, Palacios JG, Spangler RK, Bagshaw CR, Tripathi S, Crosby P, Golden SS, Partch CL, LiWang A. 2021. Reconstitution of an intact clock reveals mechanisms of circadian timekeeping. Science 374:eabd4453.

66. Diamond S, Rubin BE, Shultzaberger RK, Chen Y, Barber CD, Golden SS. 2017. Redox crisis underlies conditional light–dark lethality in cyanobacterial mutants that lack the circadian regulator, RpaA. Proc Natl Acad Sci U S A 114:E580–E589.

67. Taniguchi Y, Katayama M, Ito R, Takai N, Kondo T, Oyama T. 2007. *labA*: a novel gene required for negative feedback regulation of the cyanobacterial circadian clock protein KaiC. Genes Dev 21:60–70.

68. Cohen SE, McKnight BM, Golden SS. 2018. Roles for ClpXP in regulating the circadian clock in *Synechococcus elongatus*. Proc Natl Acad Sci U S A 115:E7805–E7813.

69. Cameron JC, Pakrasi HB. 2010. Essential role of glutathione in acclimation to environmental and redox perturbations in the cyanobacterium *Synechocystis sp. PCC 6803*. Plant Physiol 154:1672–1685.

70. Takatani N, Use K, Kato A, Ikeda K, Kojima K, Aichi M, Maeda S-I, Omata T. 2015. Essential role of acyl-ACP synthetase in acclimation of the cyanobacterium *Synechococcus elongatus* strain PCC 7942 to high-light conditions. Plant Cell Physiol 56:1608–1615.

71. Kaczmarzyk D, Fulda M. 2010. Fatty acid activation in cyanobacteria mediated by acyl-acyl carrier protein synthetase enables fatty acid recycling. Plant Physiol 152:1598–1610.

72. Ruffing AM, Jones HDT. 2012. Physiological effects of free fatty acid production in genetically engineered *Synechococcus elongatus* PCC 7942. Biotechnol Bioeng 109:2190–2199.

73. Darling ACE, Mau B, Blattner FR, Perna NT. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res 14:1394–1403.

74. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402.

75. Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, Mende DR, Letunic I, Rattei T, Jensen LJ, von Mering C, Bork P. 2019. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res 47:D309–D314.

76. Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. 2021. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. bioRxiv.

77. Webb AE, Walsh TA, O’Connell MJ. 2017. VESPA: Very large-scale Evolutionary and Selective Pressure Analyses. PeerJ Comput Sci 3:e118.

78. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30:772–780.

79. Nagar E, Zilberman S, Sendersky E, Simkovsky R, Shimoni E, Gershtein D, Herzberg M, Golden SS, Schwarz R. 2017. Type 4 pili are dispensable for biofilm development in the cyanobacterium *Synechococcus elongatus*. Environ Microbiol 19:2862–2872.

80. Yegorov Y, Sendersky E, Zilberman S, Nagar E, Waldman Ben-Asher H, Shimoni E, Simkovsky R, Golden SS, LiWang A, Schwarz R. 2021. A cyanobacterial component required for pilus biogenesis affects the exoproteome. MBio 12.

81. Parnasa R, Sendersky E, Simkovsky R, Waldman Ben-Asher H, Golden SS, Schwarz R. 2019. A microcin processing peptidase-like protein of the cyanobacterium *Synechococcus elongatus* is essential for secretion of biofilm-promoting proteins. Environ Microbiol Rep 11:456–463.

82. Parnasa R, Nagar E, Sendersky E, Reich Z, Simkovsky R, Golden S, Schwarz R. 2016. Small secreted proteins enable biofilm development in the cyanobacterium *Synechococcus elongatus*. Sci Rep 6:32209.

83. Zhang N, Chang Y-G, Tseng R, Ovchinnikov S, Schwarz R, LiWang A. 2020. Solution NMR structure of Se0862, a highly conserved cyanobacterial protein involved in biofilm formation. Protein Sci 29:2274–2280.

84. Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM. 2007. DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. Int J Syst Evol Microbiol 57:81–91.

85. Rodriguez-R LM, Konstantinidis KT. 2016. The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. e1900v1. PeerJ Preprints.

86. Wu M, Eisen JA. 2008. A simple, fast, and accurate method of phylogenomic inference. Genome Biol 9:R151.

87. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25:1972–1973.

88. Price MN, Dehal PS, Arkin AP. 2010. FastTree 2 - approximately maximum-likelihood trees for large alignments. PLoS One 5:e9490.

89. Rippka R, Deruelles J, Waterbury JB, Herdman M, Stanier RY. 1979. Generic assignments, strain histories and properties of pure cultures of cyanobacteria. Microbiology 111:1–61.

90. Shih PM, Wu D, Latifi A, Axen SD, Fewer DP, Talla E, Calteau A, Cai F, Tandeau de Marsac N, Rippka R, Herdman M, Sivonen K, Coursin T, Laurent T, Goodwin L, Nolan M, Davenport KW, Han CS, Rubin EM, Eisen JA, Woyke T, Gugger M, Kerfeld CA. 2013. Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing. Proc Natl Acad Sci U S A 110:1053–1058.

91. Clerico EM, Ditty JL, Golden SS. 2007. Specialized techniques for site-directed mutagenesis in cyanobacteria, p. 155–171. *In* Rosato, E (ed.), Circadian Rhythms: Methods and Protocols. Humana Press, Totowa, NJ.

92. Taton A, Unglaub F, Wright NE, Zeng WY, Paz-Yepes J, Brahamsha B, Palenik B, Peterson TC, Haerizadeh F, Golden SS, Golden JW. 2014. Broad-host-range vector system for synthetic biology and biotechnology in cyanobacteria. Nucleic Acids Res 42:e136.

93. Chen Y, Holtman CK, Taton A, Golden SS. 2012. Functional analysis of the *Synechococcus elongatus* PCC 7942 genome, p. 119–137. *In* Burnap, R, Vermaas, W (eds.), Functional Genomics and Evolution of Photosynthetic Systems. Springer Netherlands, Dordrecht.

94. Wetmore KM, Price MN, Waters RJ, Lamson JS, He J, Hoover CA, Blow MJ, Bristow J, Butland G, Arkin AP, Deutschbauer A. 2015. Rapid quantification of mutant fitness in diverse bacteria by sequencing randomly bar-coded transposons. MBio 6:e00306–15.

95. Ungerer J, Pakrasi HB. 2016. Cpf1 is a versatile tool for CRISPR genome editing across diverse species of cyanobacteria. Sci Rep 6:39681.

96. Mackey SR, Ditty JL, Clerico EM, Golden SS. 2007. Detection of rhythmic bioluminescence from luciferase reporters in cyanobacteria. Methods Mol Biol 362:115–129.

97. Hutchison AL, Maienschein-Cline M, Chiang AH, Tabei SMA, Gudjonson H, Bahroos N, Allada R, Dinner AR. 2015. Improved statistical methods enable greater sensitivity in rhythm detection for genome-wide data. PLoS Comput Biol 11:e1004094.

98. Zielinski T, Moore AM, Troup E, Halliday KJ, Millar AJ. 2014. Strengths and limitations of period estimation methods for circadian data. PLoS One 9:e96462.

CHAPTER 4: Creating an RB-TnSeq library in *Synechococcus elongatus* UTEX 3055

4.1 Abstract

The essential gene set of *Synechococcus elongatus* has been determined from the analysis of a randomly barcoded transposon mutant sequencing (RB-TnSeq) library in the model strain PCC 7942. The RB-TnSeq library of PCC 7942 has been used to estimate the fitness contributions of genes in specific environmental conditions. *S. elongatus* UTEX 3055 is a recent isolate that shares 98.46% average nucleotide identity with PCC 7942 but has unique phenotypes of phototaxis and robust biofilm formation in laboratory conditions. The same barcoded transposon plasmid library that was used to create the PCC 7942 RB-TnSeq library was used in a modified library building protocol to build an RB-TnSeq library in UTEX 3055. The UTEX 3055 library was constructed in multiple “mini-libraries” that were then pooled into a final library. This mini-library approach was necessary because the conjugation and transposition process is 50 times less efficient in UTEX 3055 than in PCC 7942. The library of UTEX 3055 has insertions an average of every 205 bp across the genome, with central insertions in 71% of protein-coding genes. The insertion density of the library is not enough to make definitive essentiality calls, but a preliminary essentiality analysis is possible in combination with the essentiality data of the PCC 7942. The UTEX 3055 RB-TnSeq library will be a powerful tool to explore phenotypes that were previously unavailable to RB-TnSeq analysis in PCC 7942. Additionally, the use of both RB-TnSeq libraries in tandem experiments can be used to explore the fitness of the entire pangenome of *S. elongatus*.

4.2 Introduction

The determination of gene sets that are essential for the survival of microorganisms in diverse environments is useful in many different applications in microbiology. The comparison of essential gene sets of diverse bacterial species can illuminate basic biological functions that are necessary for cellular life (1, 2) and assist synthetic biologists in the search for the blueprint of a minimal cell (3, 4). Essential gene sets can be determined through a bioinformatic approach, examining conserved genes in comparative genomic analyses (4), or through high-throughput sequencing of dense transposon mutant libraries (TnSeq) (5). TnSeq libraries can also be used in parallel sequencing strategies to compare the lethality of gene knockouts in different environmental conditions, revealing not only the essential gene set of an organism but also the fitness of genes in specific conditions. This parallel TnSeq approach has been used to study the specific biology of a species, such as finding genes that contribute to the virulence of pathogenic microbes (6, 7), or refining genome-scale metabolic models (8, 9).

A refinement of TnSeq through the addition of randomized DNA barcodes to each transposon mutant (RB-TnSeq) makes these types of parallel sequencing experiments easier, reproducible, and cost-effective (10). With an RB-TnSeq approach, a transposon library needs only to be constructed and characterized once to determine the insertion sites of the barcoded transposons, and subsequent fitness assays can be assayed through PCR amplification and sequencing of the barcodes. These libraries have been used to connect genotype to phenotype by comparing the functional gene fitness of mutant phenotypes in massive RB-TnSeq experiments, and providing experimental data to refine genome annotations for proteins of unknown functions (11).

A dense RB-TnSeq library of the model cyanobacterium *Synechococcus elongatus* PCC 7942 has been used to great effect to determine its essential gene set (12), construct a genome-scale metabolic model (9), elucidate the gene network required for survival in light-dark cycles (13, 14), and reveal genes required for natural transformation (15). A recent isolate of *S. elongatus*, UTEX 3055, shares 98.5% nucleotide identity with PCC 7942, and although clearly the same species as PCC 7942, has a genome with distinct differences and has phenotypes not observed in WT PCC 7942, such as phototaxis and biofilm formation in laboratory conditions (16). The identification and characterization of a biofilming mutant of PCC 7942 show that this lab strain is capable of forming biofilms, and supports a model of constitutive repression of the biofilm genetic program in PCC 7942. Additionally, although PCC 7942 is not phototactic, genetic transplantation of phototaxis pathway components from PCC 7942 to UTEX 3055 showed that those genes are functional.

These experimental data and the close genetic relationship of these strains suggest that a comparative genome analysis approach combined with RB-TnSeq fitness screens could shed light on the genes responsible for these phenotypes. The comparative genome analysis of *S. elongatus* UTEX 3055 and PCC 7942 has been described in Chapter 2 of this work. An RB-TnSeq library of UTEX 3055 was constructed using similar methods as those used for the PCC 7942 library. In this chapter I describe the characterized library of UTEX 3055 and its similarities and differences to the library of PCC 7942.

4.3 Results and Discussion

Library creation in S. elongatus UTEX 3055 is less efficient than in S. elongatus PCC 7942. The same protocols used to create the dense RB-TnSeq library of PCC 7942 (12) were used initially without modification to attempt to make a similar library in UTEX 3055. In the PCC 7942 library construction the plating of conjugations on nitrocellulose filters overlaid on selective media and subsequent replica-plating of the filters on fresh plates after growth resulted in fewer *Escherichia coli* conjugation donor cells carried forward in the final library. This step can help remove contaminating DNA from (dead) donor cells when the library is sequenced. The protocols designed for PCC 7942 met with limited success with UTEX 3055, with initial library construction attempts showing limited antibiotic selection when cells were plated on nitrocellulose filters. The same phenotype of biofilm formation that makes UTEX 3055 an attractive candidate for library construction may be partially responsible for this lack of antibiotic selection, but there is also some evidence (not shown) that UTEX 3055 may have different levels of sensitivity to kanamycin (the antibiotic used in library construction) than PCC 7942. A more "traditional" conjugation protocol was used instead for UTEX 3055 library construction: a conjugation reaction was made with an equal mix of donor and UTEX 3055 cells, and then plated in spots on permissive media. After 12 hours, the conjugation reactions were resuspended and plated on selective media at a density experimentally determined to allow for selection of transconjugants.

In addition to the weak kanamycin resistance phenotype of UTEX 3055, the process of conjugation and transposition appears to be 50-100X less efficient in UTEX 3055 when compared to PCC 7942. The same experiment and protocol introducing the barcoded transposon mutagenesis plasmids into cells via conjugation produced ~1500 transconjugants per conjugation plate when PCC 7942 was the recipient cell, but only 50 transconjugants with UTEX 3055 (Fig. 4-1). A similar differential in transformation efficiency has also been observed (data not shown). Other *S. elongatus* strains (PCC 6301 and UTEX 2973) are known to lack the natural transformation phenotype due to a nonsense mutation in *pilN*, a component of the Type IV pilus apparatus essential for competence in *S. elongatus* (15, 17). *S. elongatus* UTEX 3055 has a full copy of *pilN*, and the differences in transformation efficiency and success introducing the library transposon mutagenesis plasmids are likely due to other genomic differences, such as different restriction-modification systems and the presence of a CRISPR-Cas system.

In order to make a full RB-TnSeq library in UTEX 3055, accommodations for the low conjugation efficiency were necessary. In contrast to one large conjugation reaction that would result in a complete library, multiple conjugation reactions were performed to create “mini-libraries” that were frozen and archived as they were made. The mini-libraries were simultaneously revived and grown to density and pooled together into the final library, proportionally by the number of transconjugants present in each library at the time of archival, and also taking into account the density of each revived culture. This final pooled library was sequenced and archived as the complete UTEX 3055 RB-TnSeq library.

The RB-TnSeq library of S. elongatus UTEX 3055 has useful insertion density. The sequenced library was mapped to the genome of UTEX 3055 to determine the barcoded transposon insertions of each mutant strain in the pooled library. Approximately 20,100 transposon mutants were pooled to create the final library. There are 14,034 mutants with insertion locations mapped with at least 2 supporting sequencing reads. The insertion locations were checked for even insertion density between both DNA strands; the library has relatively evenly spaced insertions across the genome, with an insertion every ~205 bp, and transposon insertions in the central portion of 71% of the protein-coding genes (Fig. 4-2).

The essential gene set of S. elongatus UTEX 3055 is similar to that of S. elongatus PCC 7942. The same analysis used to determine the essential gene set of PCC 7942 through the insertion density of transposon insertions in that library was performed on the RB-TnSeq library of UTEX 3055. Briefly, an insertion index was created to compare the insertions across different genes by dividing the insertions in a gene by the length of the gene. In the PCC 7942 analysis, there was a positive bias for insertion into guanine-cytosine (GC) rich regions, and the insertion density was first normalized for GC% bias. The PCC 7942 insertion index was also calculated using only those insertions that mapped to the middle 80% of genes, to avoid including insertions that might be tolerated in the extreme 10% of otherwise essential genes. Finally, the insertion index in PCC 7942 did not include genes smaller than 70 bp, or genes that had high sequence similarity to other genes in the genome and could not have accurately mapped barcode insertions. The essentiality analysis of UTEX 3055 was first performed with the same parameters as described above for the PCC 7942 library, and the essentiality determinations for the UTEX 3055 gene set were compared to PCC 7942 essentiality. The essentiality analysis was performed again using insertions across 100% of genes without the GC% bias correction, but retaining the filter that excluded genes shorter than 70 bp or with high sequence similarity to other genome regions. The results of the essentiality analysis of the UTEX 3055 library without GC% bias correction and using all mapped insertions showed higher agreement with the essentiality analysis of PCC 7942, with 78% of essential and non-essential genes in PCC 7942 having the same determination in the UTEX 3055 library.

The essentiality analysis uses the fit of a γ-distribution to determine the essential and non-essential peaks of a bimodal distribution of the insertion index. In the case of TnSeq libraries that are not dense enough to produce this bimodal distribution, the determinations of the essentiality analysis cannot be used confidently, and in the absence of other data the only confident determination between groups in the analysis is that of the presence or absence of inserts. The UTEX 3055 library is not dense enough to produce this bimodal distribution and confidently fit a γ-distribution (Fig. 4-3). However, the essential gene set determined by the very dense PCC 7942 RB-TnSeq library can be compared to the UTEX 3055 essentiality analysis as a measure of similarities and differences in the essentiality of homologs of the two strains.

When the essential gene set of PCC 7942 is compared to that of UTEX 3055, there are 109 genes that are essential in PCC 7942 that are non-essential in UTEX 3055. One difference in the way the essentiality analysis was performed for these two libraries that could impact this outcome is the use of all gene insertions for the essentiality analysis in UTEX 3055, compared to the analysis in PCC 7942 only using insertions in the middle 80% of genes. When only insertions found in the middle 80% of genes in this "non-essential/essential" gene set of UTEX 3055 are examined further, only 11 genes have five or more central insertions (Table 4-1). One of these genes, *hik34* (UTEX3055_pg1727/Synpcc7942_1517), is the sensor kinase of a two component system that regulates the stress response gene *dnaK2*, and can be deleted in PCC 7942 (18). Taking the essentiality analysis data of both PCC 7942 and UTEX 3055 as well as this experimental data into account, it is likely that *hik34* is a beneficial gene, but is not essential to *S. elongatus*.

The non-essential/essential gene set also includes two genes that regulate translation through tRNA modification (UTEX3055_pg2024 and UTEX3055_pg2630), two genes that may regulate the translation or post-translational modifications of photosystem proteins (UTEX3055_pg0527 and UTEX3055_pg1647), and a trigger factor with functions as a chaperone, especially of exported proteins (UTEX3055_pg2735). For most of the non-essential/essential categorized genes in UTEX 3055 with few transposon insertions in the central portion of the gene, the non-essential categorization is likely due to the limitations of the essentiality analysis method with a non-saturated RB-TnSeq library. However, examination of enrichment of the genes in this non-essential/essential set with the greatest number of central insertions suggests that this subset of genes may not be essential in UTEX 3055. Approximately half of the genes in this subset code for proteins with functions in translational or post-translational regulatory functions, suggesting that UTEX 3055 may have alternative post-translational regulatory mechanisms.

The unique gene set of S. elongatus UTEX 3055 contains genes that may be essential. A major motivation for the creation of an RB-TnSeq library in UTEX 3055 is to use the library to screen for genes involved in biofilm or phototaxis phenotypes, based on the hypothesis that genes unique to UTEX 3055 are responsible for these phenotypes. There are 305 unique genes in UTEX 3055, of which 81 genes are possibly essential genes as determined by the preliminary essentiality analysis or their lack of transposon insertion. This unique and putatively essential gene set has 21 genes found in a novel prophage region and 10 anti-toxin components of toxin-antitoxin systems (TAS). The role of anti-toxin components in the fitness of bacteria that harbor TAS is well known (19), and this essentiality data can be used to refine the annotations of suspected TAS components that are currently annotated as hypothetical proteins. The majority of the putative essential prophage genes are annotated as hypothetical proteins and further experiments will be necessary to determine their role in the biology of UTEX 3055 and whether they are truly essential.

Although further experimental verification will be necessary to determine the true essentiality of genes in the unique and putative essential genes gene set of UTEX 3055, especially for those genes that have no transposon insertion, this gene set may point to key biological pathways that are unique to UTEX 3055 and important for its fitness (Table 4-2). One region of interest consists of three genes (UTEX3055_pg0056-0058) that exist in UTEX 3055 in lieu of an 11-gene region in PCC 7942. In PCC 7942, this region appears to contain a specialized carbohydrate synthesis pathway. UTEX 3055 contains none of these 11 genes, but the three genes it contains instead all appear to have functions involved in cell wall or secreted glycan synthesis. These are functions that are of great interest when searching for genetic components that would affect biofilm and phototaxis phenotypes. Other regions of interest in this unique and potentially essential region include gene clusters for cell wall/cell membrane components, and a heavy-metal resistance gene cluster.

4.4 Conclusions

The RB-TnSeq library of PCC 7942 has proven to be an invaluable tool for interrogating the genetics of phenotypes in a cost-effective and labor-effective manner. The power of an RB-TnSeq library is not only in the resulting analysis of an essential gene set for an organism, but in the ability to interrogate the fitness contributions of every non-essential gene in the genome in diverse environmental conditions. With RB-TnSeq libraries, even slight fitness effects can be determined, allowing for the study of subtle phenotype differences in a high-throughput manner.

The RB-TnSeq library of UTEX 3055 is a beneficiary of the great depth of information that has been synthesized through the use of the PCC 7942 library. The insertion density of the UTEX 3055 library is not dense enough to use the output of the essentiality analysis without additional data. The essentiality data of the PCC 7942 library, however, complements the essentiality analysis in UTEX 3055 and adds confidence to the essentiality determinations. Additional examination of disagreements between the essentiality determinations of each library indicate differences in the essential gene sets of these two strains. In some cases these differences may indicate that the essentiality call for the PCC 7942 library may need to be adjusted, supported by evidence in the literature that one essential gene in PCC 7942 that is not essential in UTEX 3055 has been successfully deleted in PCC 7942. In other cases, the differences between the essentiality determinations of the two libraries may indicate alternate genes or pathways in UTEX 3055 that can perform the essential functions.

The library of UTEX 3055 is dense enough to be used to screen for genes involved in phenotypes that are difficult or impossible to assay in PCC 7942 such as biofilm formation and phototaxis. In PCC 7942, biofilm formation is studied in mutant strains, adding an additional layer of complexity to planning and interpreting experiments using the RB-TnSeq library, and PCC 7942 is not phototactic, making it impossible to screen for genes related to this phenotype using RB-TnSeq. The two libraries can also be used in a complimentary manner to expand the genetic network of phenotypes that we have previously studied only in PCC 7942. We have recently used the RB-TnSeq library of PCC 7942 to find the genes essential for natural transformation in *S. elongatus*. UTEX 3055 has a lower transformation efficiency than PCC 7942, but doesn't appear to have major differences in the known essential competence genes. A competence fitness screen using the UTEX 3055 RB-TnSeq library, especially compared to the results of the same screen in PCC 7942, could lead to targeted genetic adjustments in UTEX 3055 that would improve competence in the strain and improve its genetic tractability as a model organism.

4.5 Materials and Methods

Mini-Library Creation. Small Tn*5* mutant libraries in *S. elongatus* UTEX 3055 were constructed in a similar manner as previously described for PCC 7942 (12). Briefly, UTEX 3055 was grown in BG-11 medium (20) as liquid cultures with continuous shaking (125 rpm) under continuous illumination of 100-150 µmol photons m^{-2} s^{-1} from fluorescent cool white bulbs until it reached OD_{750}=0.9. A diaminopimelic acid (DAP) auxotrophic *E. coli* donor strain carrying a library of barcoded Tn5 elements (pKMW7) (10) was grown in LB broth with 60 µg/mL DAP and 50 µg/mL kanamycin to an OD_{600}=1.0. Both *E. coli* and *S. elongatus* cells were washed twice and resuspended in BG-11 supplemented with 5% LB at a 1:1 donor cell:recipient cell ratio and spotted on BG-11 w/ 5% LB agar plates with

60 µg/mL DAP. The conjugation reaction was performed for 12 h under 40 µmol photons·m^{-2} ·s^{-1} of illumination and then resuspended in BG-11 and plated onto BG-11 Km agar plates for selection of exconjugants. After 10 d of growth under 100–140 µmol photons·m^{-2} ·s^{-1}, all transconjugant colonies were scraped and flushed into BG-11 medium and frozen at -80°C in 1-mL aliquots with 80 µl of DMSO.

Pooled Library Creation. Frozen mini-libraries were thawed quickly for 2 min at 37°C and then added to 100 mL of liquid BG-11 with 50 µg/mL kanamycin. Flasks were left at low light ~40 µmol photons·m^{-2} ·s^{-1} with no shaking for 24 hours, then transferred to continuous shaking (125 rpm) under continuous illumination of 100-150 µmol photons m^{-2} s^{-1} for 4 days. Libraries were pooled together based on the proportion of transconjugants present in the mini-library and normalized by the density of the culture as measured by OD_{750}. The pooled library was centrifuged and concentrated 20X and frozen at -80°C in 1-mL aliquots with 80 µl of DMSO. A sample was set aside and frozen for DNA sequencing.

Library sequencing. An Illumina-compatible sequencing library as described previously (10) with some modifications (21) was used to determine transposon insertion sites and link them to the random DNA barcodes. Briefly, genomic DNA was extracted by phenol-chloroform extraction (22), and library prep was conducted by the standard ≥100 ng protocol from the NEBNext Ultra II FS DNA Library Prep Kit for Illumina (NEB) with 500 ng of DNA and a 7 min fragmentation incubation. For adaptor ligation, a custom splinklerette adaptor was used to decrease non-specific amplification (Table 4-3) (23, 24). For size selection, 0.15X (by volume) NEBNext Sample Purification Beads (NEB) were used for the first and second bead selection steps. After digestion with BsrBI (NEB), the

DNA was purified using 1X AMPure XP beads (Beckman Coulter) in preparation for PCR enrichment where the transposon junction was amplified by nested PCR. The NEBNext Ultra II FS DNA Library Prep Kit for Illumina (NEB) PCR protocol was used, with custom primers specific to the transposon and the adaptor and a run of 30 cycles for the first PCR step (Table 2). The first PCR was purified with a 0.7X size selection using NEBNextSample Purification and eluted in 15 µl for the second PCR. The second PCR was purified with 0.7x size selection and the final library was eluted in 30 µl. Samples were quality-controlled and multiplexed using 1X HS dsDNA Qubit (Thermo Fisher) for total sample quantification, and Bioanalyzer DNA 12000 chip (Agilent) for sizing. Samples were sequenced by Novogene on the Novaseq platform.

Essentiality Analysis. Gene essentiality was determined by a previously described method (12); briefly, a normalized insertion index of the mapped transposon insertions of the library was created and statistically analyzed for genes with underrepresentation of insertions. The pangenome annotation of UTEX 3055 was used for mapping the transposon insertion sites. Genes from shorter than 70 bp, or previously identified as nearly identical to other parts of the genome, were removed from the analysis. For the remaining genes, the insertions per gene were divided by the length of the gene to calculate insertion density. Finally, a preliminary essentiality measure was determined using an approach described previously (7). Briefly, γ-distributions were fit to the essential and nonessential peaks in the insertion index, and log2 likelihood ratios were calculated from these distributions. Genes with log2 likelihood ratios below −2 were called as essential genes, and those with log2 likelihood ratios above 2 were called as nonessential genes. Scripts were adapted from the Bio::Tradis pipeline (github.com/sanger-pathogens/Bio-Tradis) (23). Because the insertion index does not fit a γ-distribution, the essential/non-essential determination from this analysis cannot be used confidently without additional data. The essentiality determination of the UTEX 3055 library was compared to the essentiality of the PCC 7942 library for shared homologs. PaperBLAST (25) was used to find additional functional information for genes.

4.6 Acknowledgements

Chapter 4 is coauthored with Emily Pierce and Susan S. Golden. The dissertation author is the primary investigator and author of this chapter.

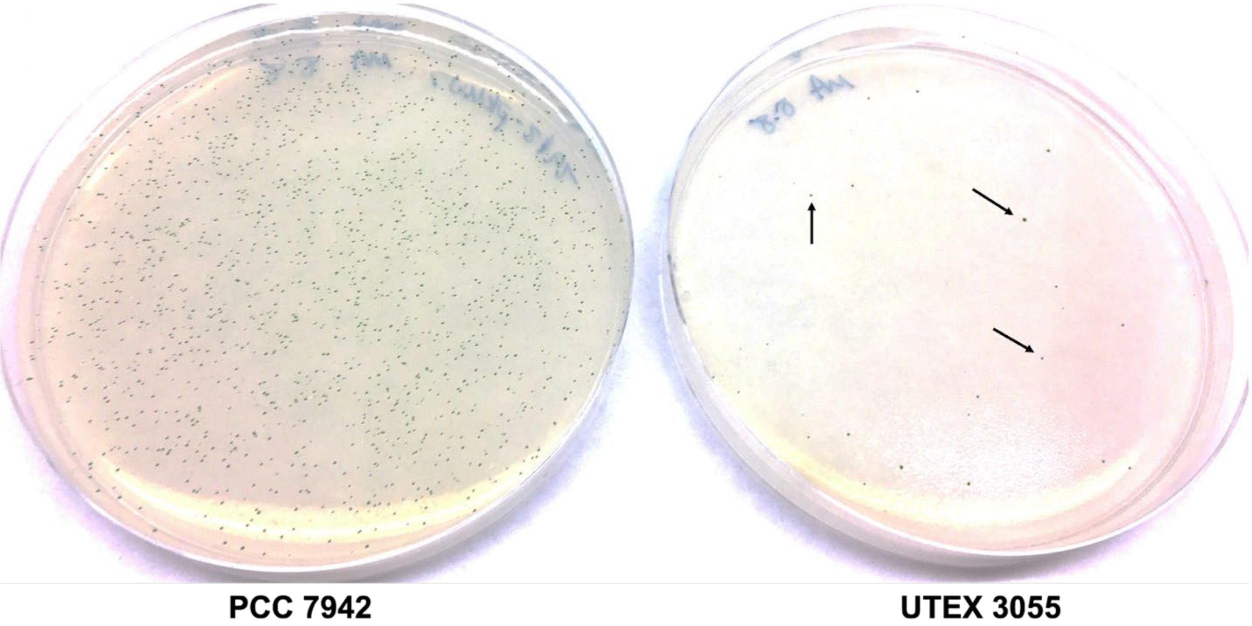


Figure 4-1: **Conjugation of *S. elongatus* PCC 7942 and UTEX 3055**. Conjugation with *E. coli* donor strain containing the barcoded transposon mutant plasmid library. The same growth conditions for both cyanobacterial strains, same donor strain, same donor:recipient ratio for conjugation reactions results in dramatically fewer numbers of conjugants in UTEX 3055.

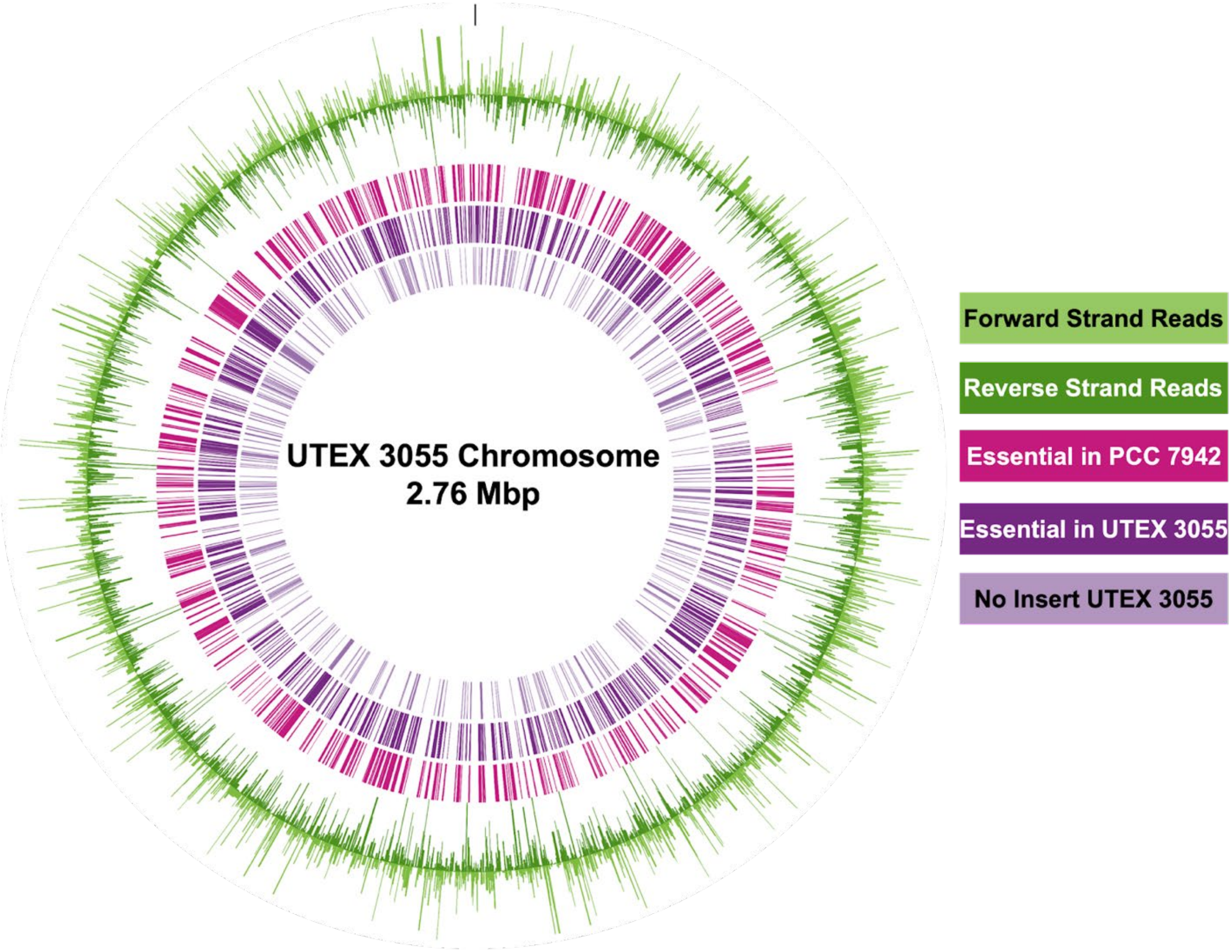


Figure 4-2: Map of UTEX 3055 RB-TnSeq insertions. The RB-TnSeq library of UTEX 3055 has even transposon insertions across the chromosome in both DNA strands (outer graph; light green and dark green graph). The genes determined to be essential in UTEX 3055 (middle ring; dark purple) agree in most cases with the genes that are essential in PCC 7942 (outer ring; pink). Those genes with no transposon insertions in UTEX 3055 are shown in the inner ring (light purple).

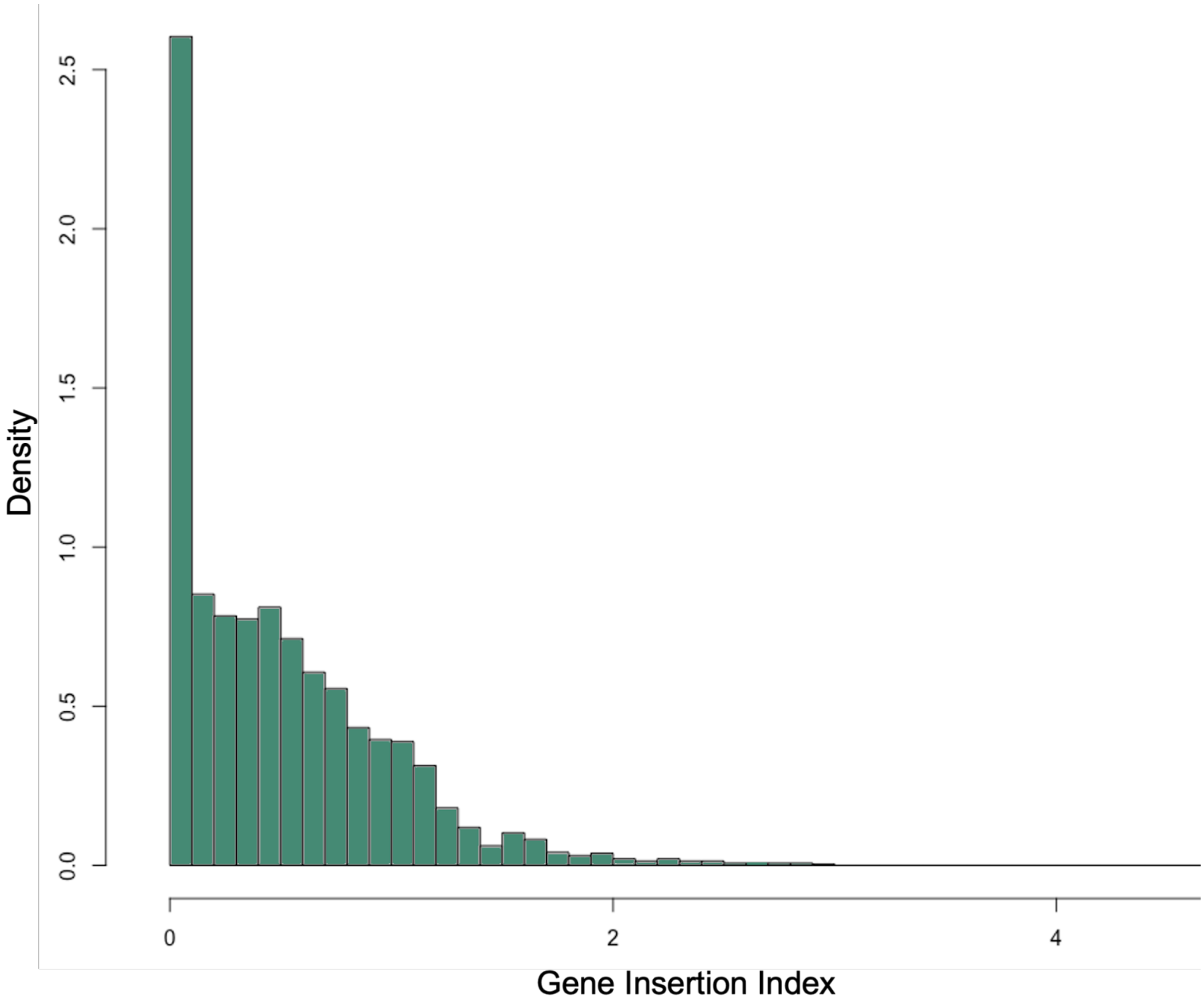


Figure 4-3: Distribution of RB-TnSeq library insertions. The distribution of gene insertions in the UTEX 3055 RB-TnSeq library does not follow a bimodal distribution that allows for a distinction between essential and non-essential genes. Instead, the clear distinction can be seen between genes with no insertions and genes with one or more insertions.

Table 4-1: Non-essential UTEX 3055 genes that are essential in PCC 7942. Some genes that are essential in PCC 7942 have insertions in the central portion of the UTEX 3055 homologs, suggesting that they are not essential in UTEX 3055. Eleven genes with more than 5 insertions in the central portion of the gene are described in the table below, with the number of insertions in that gene in the UTEX 3055 RB-TnSeq library and the percentage of identical amino acids shared with the gene's homolog in PCC 7942.

Locus	**Gene Description**	**COG Category; Description**	**Tn*5* insertions**	**% identical amino acids**
UTEX3055_pg0022	sulfite reductase; *sir*	**P**; Inorganic ion transport and metabolism	7	99
UTEX3055_pg0224	Nucleoside-diphosphate-sugar epimerase	**M**; Cell wall/membrane/envelope biogenesis	8	99
UTEX3055_pg0527	C-terminal processing peptidase-2; *ctpB*	**O**; Posttranslational modification, protein turnover, chaperones	7	100
UTEX3055_pg1219	hydroxymethylpyrimidine synthase; *thiC*	**H**; Coenzyme transport and metabolism	19	100
UTEX3055_pg1220	riboswitch binding thiamine pyrophosphate	**Y**; Not in COGs	6	--
UTEX3055_pg1647	Protein of unknown function DUF1092	**Y**; Not in COGs	6	100
UTEX3055_pg1727	histidine kinase	**T**; Signal transduction mechanisms	5	100
UTEX3055_pg2024	L-threonylcarbamoyladenylate synthase	**J**; Translation, ribosomal structure and biogenesis	6	97
UTEX3055_pg2492	hypothetical protein	**M**; Cell wall/membrane/envelope biogenesis	5	99
UTEX3055_pg2630	tRNA uridine 5-carboxymethylaminomethyl modification enzyme; *gidA*	**J**; Translation, ribosomal structure and biogenesis	5	100
UTEX3055_pg2735	trigger factor; *tig*	**O**; Posttranslational modification, protein turnover, chaperones	8	100

Table 4-2: UTEX 3055 unique and putative essential genes. Unique genes in UTEX 3055 that have putative essentiality, either by preliminary essentiality analysis or by the lack of transposon insert. Gene clusters with functions of interest, as determined by bioinformatic analysis, are highlighted.

Locus	Description	Essentiality	Putative Gene Cluster Function
UTEX3055_pg0037	hypothetical protein	essential	
UTEX3055_pg0056	hypothetical protein	no insert	Glycan biosynthesis
UTEX3055_pg0057	class I SAM-dependent methyltransferase	essential	
UTEX3055_pg0058	glycosyltransferase family 2 protein	no insert	
UTEX3055_pg0100	hypothetical protein	no insert	
UTEX3055_pg0102	hypothetical protein	essential	
UTEX3055_pg0182	O-acetyl-ADP-ribose deacetylase regulator of RNase III, contains Macro domain	essential	
UTEX3055_pg0801	hypothetical protein	essential	
UTEX3055_pg0802	hypothetical protein	no insert	
UTEX3055_pg0929	hypothetical protein	no insert	
UTEX3055_pg0963	hypothetical protein	no insert	
UTEX3055_pg1020	hypothetical protein	essential	
UTEX3055_pg1147	Protein of unknown function DUF1524	essential	
UTEX3055_pg1241	hypothetical protein	essential	
UTEX3055_pg1294	Phage integrase family protein	essential	Orphan response regulator
UTEX3055_pg1295	Response regulator receiver domain-containing protein	no insert	
UTEX3055_pg1296	hypothetical protein	no insert	
UTEX3055_pg1681	hypothetical protein	no insert	
UTEX3055_pg1682	hypothetical protein	no insert	
UTEX3055_pg1804	T5orf172 domain-containing protein	essential	
UTEX3055_pg1857	hypothetical protein	no insert	
UTEX3055_pg1859	hypothetical protein	no insert	
UTEX3055_pg2116	tellurium resistance protein TerD	no insert	
UTEX3055_pg2264	hypothetical protein	no insert	
UTEX3055_pg2309	hypothetical protein	essential	
UTEX3055_pg2310	hypothetical protein	no insert	
UTEX3055_pg2315	hypothetical protein	no insert	
UTEX3055_pg2317	Ubiquinone biosynthesis protein Coq4	no insert	
UTEX3055_pg2319	transcriptional regulator, TetR family	essential	
UTEX3055_pg2473	Uncharacterized membrane protein YeaQ/YmgE, transglycosylase-associated protein family	essential	Cell wall/membrane biogenesis
UTEX3055_pg2474	Peptidoglycan/xylan/chitin deacetylase, PgdA/CDA1 family	no insert	
UTEX3055_pg2475	hypothetical protein	essential	
UTEX3055_pg2476	hypothetical protein	essential	
UTEX3055_pg2479	hypothetical protein	essential	

Table 4-2 (cont.): UTEX 3055 unique and putative essential genes.

Locus	Description	Essentiality	Putative Gene Cluster Function
UTEX3055_pg2687	peptide/nickel transport system permease protein	no insert	
UTEX3055_pg2688	Uncharacterized conserved protein, contains HEPNdomain	no insert	
UTEX3055_pgB028	hypothetical protein	essential	
UTEX3055_pgB030	DUF3854 domain-containing protein	essential	
UTEX3055_pgB042	ABC transporter permease subunit	no insert	
UTEX3055_pgB044	methyltransferase domain-containing protein	no insert	
UTEX3055_pgB045	monooxygenase	essential	
UTEX3055_pgB048	hypothetical protein	essential	
UTEX3055_pgB059	hypothetical protein	no insert	
UTEX3055_pgB064	recombinase family protein	no insert	
UTEX3055_pgB066	hypothetical protein	essential	
UTEX3055_pgB076	metal ABC transporter permease	no insert	Heavy metal resistance
UTEX3055_pgB077	metal ABC transporter ATP-binding protein	essential	
UTEX3055_pgB079	cadmium-translocating P-type ATPase	essential	
UTEX3055_pgB083	DUF2513 domain-containing protein	essential	
UTEX3055_pgD011	translesion error-prone DNA polymerase V autoproteolytic subunit	no insert	

Table 4-3: Primer sequences for transposon-junction sequencing. The index region of the second stage PCR primers is highlighted in red.

Description	Sequence
splinkerette adaptor "top" strand for ligation with fragmented DNA	/5Phos/ GATCGGAAGAGCTTTTTTTTTTCAAAAAAAA
splinkerette adaptor "bottom" strand (with 3' A overhang) for ligation with fragmented DNA	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
Nested PCR round 1 primer to amplify from the insertion	GTAATACGACTCACTATAGGGGATA*G
Nested PCR round 1 primer to amplify from the adaptor (binding site created upon extension from insertion)	ATCGGTCTCGGCATTCCCA*G
Nested PCR Round 2 Primer (p5/i5) to amplify from the insertion	AATGATACGGCGACCACCGAGATCTACAC GTCGTC ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNGATGTCCACGAGGTCTC*T
Nested PCR Round 2 Primer (p7/i7) to amplify from the splinkerette adaptor	CAAGCAGAAGACGGCATACGAGAT ACATC GGTGACTGGAGTTCAGACGTGT*G

References

1. Juhas M, Eberl L, Glass JI. 2011. Essence of life: essential genes of minimal genomes. Trends Cell Biol 21:562–568.

2. Luo H, Lin Y, Liu T, Lai F-L, Zhang C-T, Gao F, Zhang R. 2021. DEG 15, an update of the Database of Essential Genes that includes built-in analysis tools. Nucleic Acids Res 49:D677–D686.

3. Glass JI, Assad-Garcia N, Alperovich N, Yooseph S, Lewis MR, Maruf M, Hutchison CA 3rd, Smith HO, Venter JC. 2006. Essential genes of a minimal bacterium. Proc Natl Acad Sci U S A 103:425–430.

4. Delaye L, González-Domenech CM, Garcillán-Barcia MP, Peretó J, de la Cruz F, Moya A. 2011. Blueprint for a minimal photoautotrophic cell: conserved and variable genes in Synechococcus elongatus PCC 7942. BMC Genomics 12:25.

5. van Opijnen T, Lazinski DW, Camilli A. 2015. Genome-Wide Fitness and Genetic Interactions Determined by Tn-seq, a High-Throughput Massively Parallel Sequencing Method for Microorganisms. Curr Protoc Microbiol 36:1E.3.1–24.

6. Dembek M, Barquist L, Boinett CJ, Cain AK, Mayho M, Lawley TD, Fairweather NF, Fagan RP. 2015. High-throughput analysis of gene essentiality and sporulation in Clostridium difficile. MBio 6:e02383.

7. Langridge GC, Phan M-D, Turner DJ, Perkins TT, Parts L, Haase J, Charles I, Maskell DJ, Peters SE, Dougan G, Wain J, Parkhill J, Turner AK. 2009. Simultaneous assay of every *Salmonella typhi* gene using one million transposon mutants. Genome Res 19:2308–2316.

8. Burger BT, Imam S, Scarborough MJ, Noguera DR, Donohue TJ. 2017. Combining Genome-Scale Experimental and Computational Methods To Identify Essential Genes in Rhodobacter sphaeroides. mSystems 2:e00015–17.

9. Broddrick JT, Rubin BE, Welkie DG, Du N, Mih N, Diamond S, Lee JJ, Golden SS, Palsson BO. 2016. Unique attributes of cyanobacterial metabolism revealed by improved genome-scale metabolic modeling and essential gene analysis. Proc Natl Acad Sci U S A 113:E8344–E8353.

10. Wetmore KM, Price MN, Waters RJ, Lamson JS, He J, Hoover CA, Blow MJ, Bristow J, Butland G, Arkin AP, Deutschbauer A. 2015. Rapid quantification of mutant fitness in diverse bacteria by sequencing randomly bar-coded transposons. MBio 6:e00306–15.

11. Price MN, Wetmore KM, Waters RJ, Callaghan M, Ray J, Liu H, Kuehl JV, Melnyk RA, Lamson JS, Suh Y, Carlson HK, Esquivel Z, Sadeeshkumar H, Chakraborty R, Zane GM, Rubin BE, Wall JD, Visel A, Bristow J, Blow MJ, Arkin AP, Deutschbauer AM. 2018.

Mutant phenotypes for thousands of bacterial genes of unknown function. Nature 557:503–509.

12. Rubin BE, Wetmore KM, Price MN, Diamond S, Shultzaberger RK, Lowe LC, Curtin G, Arkin AP, Deutschbauer A, Golden SS. 2015. The essential gene set of a photosynthetic organism. Proc Natl Acad Sci U S A 112:E6634–43.

13. Rubin BE, Huynh TN, Welkie DG, Diamond S, Simkovsky R, Pierce EC, Taton A, Lowe LC, Lee JJ, Rifkin SA, Woodward JJ, Golden SS. 2018. High-throughput interaction screens illuminate the role of c-di-AMP in cyanobacterial nighttime survival. PLoS Genet 14:e1007301.

14. Welkie DG, Rubin BE, Chang Y-G, Diamond S, Rifkin SA, LiWang A, Golden SS. 2018. Genome-wide fitness assessment during diurnal growth reveals an expanded role of the cyanobacterial circadian clock protein KaiA. Proc Natl Acad Sci U S A 115:E7174–E7183.

15. Taton A, Erikson C, Yang Y, Rubin BE, Rifkin SA, Golden JW, Golden SS. 2020. The circadian clock and darkness control natural competence in cyanobacteria. Nat Commun 11:1688.

16. Yang Y, Lam V, Adomako M, Simkovsky R, Jakob A, Rockwell NC, Cohen SE, Taton A, Wang J, Lagarias JC, Wilde A, Nobles DR, Brand JJ, Golden SS. 2018. Phototaxis in a wild isolate of the cyanobacterium *Synechococcus elongatus*. Proc Natl Acad Sci U S A 115:E12378–E12387.

17. Li S, Sun T, Xu C, Chen L, Zhang W. 2018. Development and optimization of genetic toolboxes for a fast-growing cyanobacterium *Synechococcus elongatus* UTEX 2973. Metab Eng 48:163–174.

18. Sato M, Nimura-Matsune K, Watanabe S, Chibazakura T, Yoshikawa H. 2007. Expression analysis of multiple dnaK genes in the cyanobacterium Synechococcus elongatus PCC 7942. J Bacteriol 189:3751–3758.

19. Leplae R, Geeraerts D, Hallez R, Guglielmini J, Drèze P, Van Melderen L. 2011. Diversity of bacterial type II toxin-antitoxin systems: a comprehensive search and functional analysis of novel families. Nucleic Acids Res 39:5513–5525.

20. Rippka R, Deruelles J, Waterbury JB, Herdman M, Stanier RY. 1979. Generic assignments, strain histories and properties of pure cultures of cyanobacteria. Microbiology 111:1–61.

21. Rubin BE, Diamond S, Cress BF, Crits-Christoph A, He C, Xu M, Zhou Z, Smock DC, Tang K, Owens TK, Krishnappa N, Sachdeva R, Deutschbauer AM, Banfield JF, Doudna JA. 2020. Targeted Genome Editing of Bacteria Within Microbial Communities. bioRxiv.

22. Clerico EM, Ditty JL, Golden SS. 2007. Specialized techniques for site-directed mutagenesis in cyanobacteria, p. 155–171. *In* Rosato, E (ed.), Circadian Rhythms: Methods and Protocols. Humana Press, Totowa, NJ.

23. Barquist L, Mayho M, Cummins C, Cain AK, Boinett CJ, Page AJ, Langridge GC, Quail MA, Keane JA, Parkhill J. 2016. The TraDIS toolkit: sequencing and analysis for dense transposon mutant libraries. Bioinformatics 32:1109–1111.

24. Devon RS, Porteous DJ, Brookes AJ. 1995. Splinkerettes--improved vectorettes for greater efficiency in PCR walking. Nucleic Acids Res 23:1644–1645.

25. Price MN, Arkin AP. 2017. PaperBLAST: Text Mining Papers for Information about Homologs. mSystems 2.

CHAPTER 5: Using Interaction RB-TnSeq to understand the genetic model of biofilm formation in *S. elongatus*

5.1 Abstract

Cyanobacterial biofilms are not only important as critical components of ecological habitats, but also have applications in wastewater treatment systems, bioremediation efforts, and prevention of biofouling. Despite the importance and promise of these microbial communities, most biofilm research is performed in heterotrophic bacterial species, and much less is known about the regulation and mechanisms of biofilm formation in cyanobacteria. The model cyanobacterium *Synechococcus elongatus* PCC 7942 is planktonic in lab conditions, but studies of biofilming mutants of the strain support a model of constitutive repression of biofilms in the lab and suggest that additional components of the biofilm genetic network can be found using this strain. An adaptation of the application of randomly barcoded transposon sequencing (RB-TnSeq) that introduces a known secondary mutation to the library to find gene interactions was used to expand the genetic network of biofilm formation in PCC 7942. An interaction RB-TnSeq (IRBSeq) screen adding a known insertion knockout mutation in *pilB* that causes biofilm formation in PCC 7942, found 197 synthetic gene interactions. Fitness estimate analysis of these interactions in the biofilm found the majority of Type IV pilus (T4P) components are necessary for biofilm formation and some metabolic changes that may promote survival in the biofilm. Genes whose products interact with second messenger signaling molecules were also scored as hits in the fitness analysis, and further exploration of these genes may expand our current model of the regulation of biofilm formation in PCC 7942.

5.2 Introduction

In liquid culture, *Synechococcus elongatus* PCC 7942 grows as a planktonic suspension with WT cells remaining buoyant in the absence of shaking or agitation of culture flasks. The mechanism of cellular suspension is not known in this strain, but may be related to pili, as the cells of non-piliated mutant strains show fast sedimentation rates compared to WT cells (1). A transposon insertion mutant of PilB in PCC 7942 forms biofilms, and experiments using the conditioned medium with this mutant support a model of constitutive repression of biofilm formation in PCC 7942 (2, 3). In this model, an unknown small molecule(s) secreted by WT PCC 7942 through the Type II secretion system represses the expression of small peptides that enable biofilm formation (Fig. 5-1). PilB is a component of the Type IV pilus (T4P) machinery, which is homologous with Type II secretion systems (T2SS), both of which push proteins through channels in the outer membrane (4). As would be expected for the disruption of both T4P and T2SS, the *pilB* mutant is non-piliated and has an altered protein secretion profile (1). Other biofilming mutants of PCC 7942 that have been found share a common theme of disruption of the T4P/T2SS (Table 5-1) and appear to act in the same pathway of constitutive repression of biofilm formation (Fig. 5-1).

A randomly-barcoded transposon insertion sequencing (RB-TnSeq) library in PCC 7942 has been used previously to determine the essential gene set of the organism (5) and for screening experiments to determine genes that contribute to various phenotypes, such as natural competence (6) and survival in light-dark cycles (7). In these screening experiments, the RB-TnSeq library is grown in a control and an experimental condition, sequenced, and the frequencies of previously mapped barcoded insertions are determined from the sequencing data. The barcode frequencies are a measurement of the abundance of mutant strains from the library in that environmental condition, and the comparison of these frequencies between the control and experimental conditions determine mutant strains that are overrepresented or underrepresented in the experimental condition, providing an estimate of the fitness contribution of each gene in that condition.

Because PCC 7942 constitutively represses biofilm formation, an adaptation of the RB-TnSeq method was used to screen for genes that contribute to the biofilm phenotype by first introducing a mutation that imposes biofilm proficiency to the entire library. Interaction RB-TnSeq, or IRB-Seq, measures the effects of a genetic perturbation in the library in addition to the effect of the environmental condition. IRB-Seq is accomplished by using the natural competence of PCC 7942 to introduce a known mutation to all mutants of the library, creating a synthetic interaction library in which each mutant strain contains a known mutation in addition to the library transposon insertion. The first stage of IRB-Seq, the interaction screen, measures the fitness effect of synthetic interactions, while the second stage of IRB-Seq, the sensitized interaction screen, incorporates the effect of an experimental condition to determine a fitness estimate for synthetic interactions in the experimental condition.

For this study, the *pilB* mutation was added to the library in an IRB-Seq experiment to promote biofilm formation, with subsequent screening for additional genes in the network that affects biofilm formation in PCC 7942. The IRB-Seq analysis found that most components of the Type 4 pilus and genes that are essential for transformation in PCC 7942 are also necessary for biofilm formation. Genes associated with the metabolic state of the cell, such as those associated with biosynthetic pathways, nutrient uptake mechanisms, or energy production, had altered fitness in the biofilm. The IRB-Seq screen also found two genes that regulate the cellular second messengers cyclic-AMP (cAMP) and cyclic-di-GMP (c-di-GMP) which previously had not been associated with biofilm regulation in PCC 7942.

5.3 Results and Discussion

A pilB IRB-Seq library in S. elongatus PCC 7942 screens for interactions in the biofilm pathway. A screen using an RB-TnSeq library can examine the fitness effects genes have on the reproduction of the cell in a specific environmental condition by comparing the relative frequency of each transposon insertion mutant in the library between a control condition and the experimental condition. With IRB-Seq, in addition to the measurement of the environmental perturbation of the library, a known genetic perturbation is added, and the fitness analysis measures the combined effect of these perturbations to find synthetic interactions. *S. elongatus* PCC 7942 constitutively represses biofilm formation by secreting a suppressor molecule(s) into the surrounding medium. This suppression phenotype complicates the use of the RB-TnSeq library for biofilm fitness screens in PCC 7942, but an IRB-Seq approach can take advantage of the known biofilm phenotype and genetic model of the *pilB* mutation.

For the first stage of IRB-Seq, the RB-TnSeq library was revived and initial T0 reference samples were archived. The T0 sample is the reference for the first stage comparison, the interaction screen (Fig. 5-2A). The library was split, and one portion was transformed with a *pilB* mutation plasmid and the second with a control plasmid. The library mutants contain a kanamycin-resistance cassette in the transposon insertion, so *pilB* and control plasmids were chosen that carry a gentamicin-resistance cassette, and library transformants were selected with both kanamycin and gentamicin to maintain the diversity of the library. The control plasmid transformation identifies mutants that are incapable of transformation or are affected by the transformation protocol. The transformants from the separate transformation reactions were collected and pooled, and Tɑ reference samples were archived. The Tɑ samples are the reference for the second comparison, the experimental interaction screen (Fig. 5-2B). The individual transformation libraries were used to inoculate flasks at a high density to promote biofilm formation, as the *pilB* biofilm phenotype develops during stationary phase growth. After 4 days in low light with no agitation, the flasks were separated into different fractions: cells in planktonic growth, cells that settled but were not incorporated into the biofilm, and the biofilm (Fig. 5-3). Analysis of the experimental interaction screen was done in pairwise fashion, comparing the planktonic fraction to either the biofilm or settled fraction to determine the fitness effect of gene interactions with *pilB*.

The synthetic interactions of pilB are alleviating interactions. The interaction analysis of *pilB* was filtered for genes with a false discovery rate (FDR, linear mixed-effects model) <0.01 and a fitness estimate > |1|. There were no genes with negative fitness estimates that would indicate aggravating or synthetic lethal gene interactions, but 65 genes with positive fitness estimates indicating alleviating interactions were present (Table 5-2). It isn’t surprising that no synthetic lethal interactions with *pilB* were observed, as mutations of the T4P apparatus are not uncommon in *S. elongatus*, including a mutation of PilN that renders several strains non-transformable (8). One third of the alleviating interactions are either core components of the T4P apparatus or members of the known set of genes required for competence in PCC 7942 (6). These competence genes appear in the interaction screen because they are depleted in both samples as a result of failing gentamicin selection. When the two sets are compared, any variation in the low frequency of barcodes associated with these genes in the *pilB* transformation relative to the control will cause a significant fitness estimate in the analysis. Additionally, synthetic alleviating interactions can indicate genes that are in the same pathway or are part of the same functional unit, as once the pathway has been rendered non-functional by one mutation, additional perturbations of the pathway will not have a fitness cost (9). Specifically, once the T4P has been rendered non-functional by the *pilB* mutation, additional mutations in components of the T4P machinery do not have an additional cost. Nine genes with transcriptional regulation or signal transduction functions also have alleviating interactions with *pilB*. Two of these genes encode CP12 proteins, which are theorized to be components of redox-mediated metabolic switches in cyanobacteria (10). A histidine kinase gene in this group of alleviating interactions has been previously observed to have co-fitness with T4P component genes in PCC 7942 in multiple fitness screens (11). These alleviating interactions with signal transduction or transcriptional regulation functions are targets that can expand the present model of the regulation of biofilm formation in *S. elongatus* to include the integration of a signaling pathway beyond the current model.

Fitness in the biofilm. The experimental interaction screen analysis revealed 131 genes with a significant (FDR > 0.01) fitness effect greater than |0.5| in the settled or biofilm fractions of the *pilB* interaction library (Fig. 5-4). The biofilm interaction screen analysis assesses the fitness effect of synthetic interactions on a mutant’s presence in the biofilm or in the settled fraction of the culture. In the case of the *pilB* IRB-Seq library, however, not all fitness estimates in the biofilm screen must be interpreted as gene interactions. This is partly due to the fact that the *pilB* mutation does not have a large fitness cost to cells, and also because the *pilB* mutant has a dramatically different secretion profile that promotes biofilm formation (12). The exoproteome of the *pilB* mutant has been shown to promote biofilm formation even in WT PCC 7942 cells that are capable of producing the repressor of biofilm formation. In this background and culture milieu, the interaction fitness effect of genes in biofilm and settled cell fractions do not have to be considered strictly as interaction mutants, and can also be considered on their own merit for their roles in the regulation of biofilm formation.

A gene set enrichment analysis based on categories for T4P machinery, competence, the biofilm pathway of PCC 7942, and COGs was performed on the sets of genes that had significant positive or negative interaction fitness effects in the biofilm interaction screen. The categories of T4P, competence, biofilm genes and motility or secretion COGs were enriched in the negative interaction gene sets (Fig. 5-5). The set of competence genes have significant fitness estimates in the biofilm experiment comparison due to their appearance at low but significant frequency in the interaction screen. Two proteins that are known to work in the same pathway of biofilm regulation as the *pilB* mutant and interact physically with PilB, EbsA and hfq, also have negative fitness in the interaction screen, as would be expected for interactions in the same pathway or function (12).

Many of the same mutants have negative fitness estimates in both biofilm and settled fractions, but the settled fraction contains some mutants that score with negative fitness estimates that do not have negative fitness in the biofilm fraction. There are no genes with a positive fitness estimate in the settled fraction that do not also have a positive fitness effect in the biofilm fraction. This pattern may indicate that the genes with negative fitness effects only in the settled fraction are necessary in specific stages of biofilm formation, as strains that become easily incorporated into the biofilm once they are no longer part of the planktonic fraction. An operon that synthesizes O-antigen in PCC 7942 has negative fitness in the settled fraction (13, 14). The previous work in PCC 7942 on the genes in this operon showed that mutations in these genes lead cells to autoflocculate and settle in flasks. It is likely that mutations in these genes allow cells to be easily incorporated into the biofilm, and their quick settling phenotype would have made them early components of the biofilm.

The gene set with positive fitness effects in the biofilm fraction include genes that regulate the second messengers cyclic-AMP (cAMP) and cyclic-di-GMP (c-di-GMP). The regulation of c-di-GMP levels in the cell is a canonical pathway for regulating the lifestyle switch between a sessile, biofilm lifestyle and a motile or planktonic lifestyle. Synpcc7942_1716 contains a phosphodiesterase domain to break down c-di-GMP, and no apparent regulatory region. I hypothesize that the knockout of this phosphodiesterase may allow c-di-GMP to rise in PCC 7942 to a level that promotes biofilm formation. Similarly, another gene with a positive fitness effect in the biofilm fraction is a cAMP receptor protein (*crp*). Previous literature did not find this protein in PCC 7942 and hypothesized that it had been lost as the result of adaptation to new environments, as Crps in other cyanobacteria are transcriptional regulators that impact diverse responses such as energy and motility (15–17).

5.4 Conclusions

An IRB-Seq approach was used to find the fitness contributions of genes to biofilm formation in PCC 7942. By adding the *pilB* mutation to the RB-TnSeq library of PCC 7942, the repression of biofilm formation was abolished and the fitness contributions of genes to the biofilm phenotype of PCC 7942 could be determined. Many genes that have already been shown to have functions in T4P or competence were found to have negative fitness in the biofilm, which supports previous knowledge that T4P provide an advantage in biofilm formation in other bacteria, even as they are dispensable for biofilm formation in PCC 7942 (1). Importantly, the screen identified genes that can be investigated further to expand the current model of biofilm formation in *S. elongatus* beyond the pathway controlled by constitutive repression, such as transcriptional regulators and genes involved in the metabolism of the second messengers cAMP and c-di-GMP.

5.5 Materials and Methods

IRB-Seq Screen and Sequencing. An IRB-Seq library was constructed and analyzed as previously described (18). Briefly, an aliquot of the RB-TnSeq library of PCC 7942 was thawed and grown as previously described (5). Four samples were taken for the T0 time point and pelleted and frozen at -80°C, then the library was split and transformed in triplicate reactions with standard transformation protocols (19) using a knockout mutation plasmid to create the *pilB* insertion knockout mutation (2) with gentamicin resistance (AM5369) and separately a Neutral Site I control mutation plasmid with gentamicin resistance (AM5610). After 4 days of growth in 140 µmol photons m^{-2} s^{-1} under kanamycin (for library selection) and gentamicin (for interaction mutation selection) transformants for each transformation reaction were harvested and pooled. One sample from each tube was pelleted and frozen at -80°C to serve as the post-selection sample for the interaction screen analysis and the T0 (Tα) sample for the sensitized interaction screen analysis. The harvested transformation plates were used as inocula for the biofilm interaction screen. Flasks containing 50 ml BG-11 were inoculated at a density of OD_{750} = 0.5 and placed in ~40 μmol photons m^{-2} s^{-1} without shaking for 4 days until visible biofilms had formed. Planktonic cells were harvested from each flask by drawing off all but 2 mL of culture. Settled cells were harvested by the gentle addition of 3 mL of fresh BG-11 and swirling the flasks for 5 seconds before drawing off the media and cells. The biofilm fraction was harvested by adding 5 mL of BG-11 medium and pipetting and scraping all biofilm cells off of the walls and bottom of the flasks. All fractions were pelleted and frozen at -80°C. The genomic DNA of all frozen archived samples was extracted using a phenol-chloroform protocol (19), and the barcoded mutant frequencies in each sample were quantified using the previously described BarSeq protocol (20).

IRB-Seq Analysis. The interaction screen and biofilm interaction screen analysis was performed as described previously (18). Briefly, to identify genetic interactions, the frequency of library constituents in the *pilB* mutation experiment was compared to the frequency of library constituents in the control mutation experiments to determine the interaction effects of the *pilB* mutation. The control mutation background was considered to be the control condition, and the *pilB* mutation background the experimental. For the biofilm screen, pairwise analyses were performed, with the second partner of the pair considered to be the sensitizing condition: Planktonic vs. Biofilm, Planktonic vs. Suspended, Suspended vs. Biofilm. A FDR <0.01 was used as a cutoff for candidates; no fitness estimate cutoff was used.

Gene Set Enrichment Analysis. Categorical meta-data for genes from multiple sources were curated for all genes in the PCC 7942 pangenome, such as: essentiality and conservation data (5), functional categories (COGs) determined in the pangenome analysis (see Chapter 2), pili and competence genes (6), and known biofilm genes (1–3, 12, 21). The IRB-Seq gene sets with significant positive or negative enrichment in biofilm or settled fractions were identified, and significant enrichments in the identified meta-data categories were determined using custom R scripts. Briefly, enrichment values were determined using two-sided Fisher's exact tests, with FDR-adjusted p-values ≤ 0.05 being designated as significant. Fold enrichment (F) was calculated as the number of genes in the IRB-Seq interest group that are also in the meta-data category (N_{gc}) divided by the number of genes expected in the group and category (E_{gc}). This expected number was calculated by multiplying the number of total genes in the IRB-Seq interest group (N_g) by the frequency of all genes in the genome that are found in the meta-data category (f_c), which was determined as the number of genes in the category (N_c) divided by the number of genes in the genome (N).

$$F = N_{gc}/E_{gc};\ E_{gc} = N_g * f_c;\ f_c = N_c/N$$

5.6 Acknowledgements

Chapter 5 is coauthored with Elliot Weiss, Ryan Simkovsky, and Susan S. Golden. The dissertation author is the primary investigator and author of this chapter.

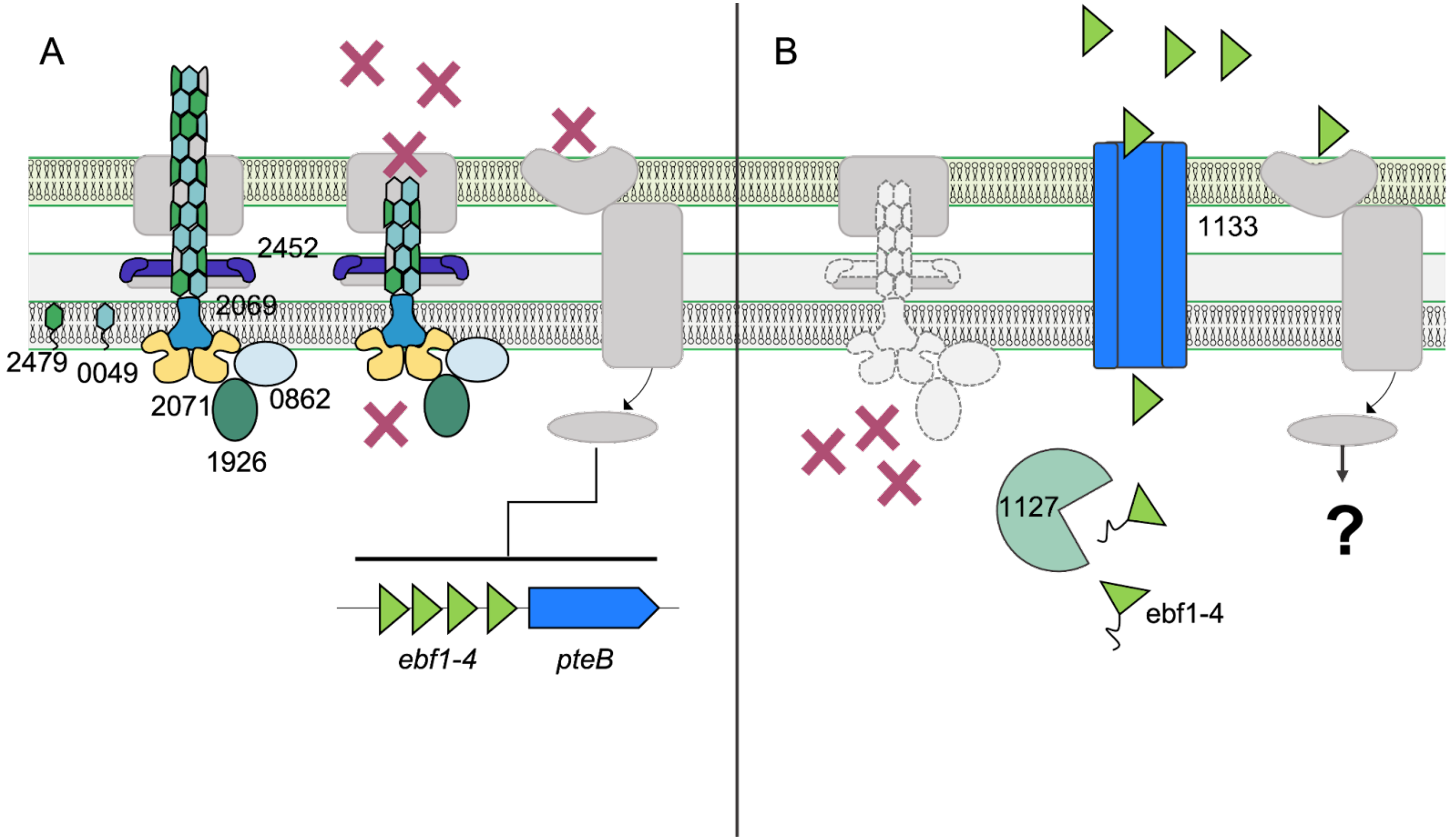


Figure 5-1: Model of constitutive repression of biofilm formation. (A) Biofilm formation is constitutively repressed in PCC 7942 by an unknown repressor that is secreted through the T2SS, suppressing expression (through an unknown mechanism) of Ebf small proteins and PteB, a peptidase that takes part in the secretion of these proteins. (B) When components of the T4P/T2SS are knocked out, ebf proteins, PteB and the cysteine peptidase EbsA (1127) are expressed, and the Ebf proteins are secreted and promote biofilm formation through an unknown regulatory pathway. Components of the T4P/T2SS that are not known to be relevant to the biofilm phenotype and the putative components of unknown regulatory pathways are represented in grey. Knockout components of the T4P/T2SS in (B) are represented in light grey.

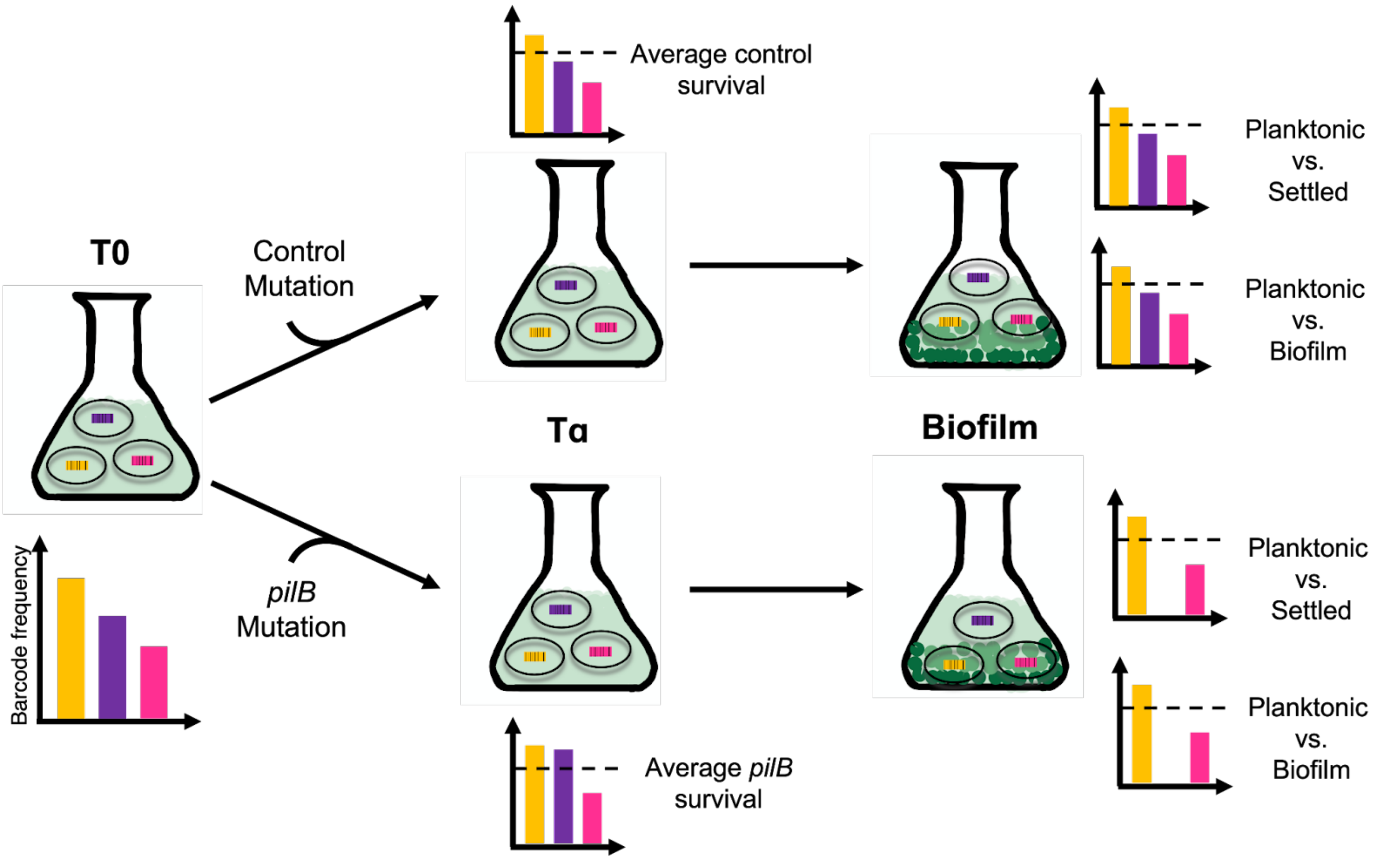


Figure 5-2: Schematic of the IRB-Seq experiment. The RB-TnSeq library contains a pool of loss-of-function mutants whose frequency is determined at T0. The library is split into two fractions, one receiving a control mutation and the other receiving the *pilB* mutation through transformation. After transformation, the frequencies of mutants in each fraction (Tɑ) are compared to determine genetic interactions between the library and the mutations. The two fractions are allowed to form biofilms, and the separate planktonic, settled, and biofilm cell fractions are sequenced separately. The biofilm and settled fractions in the *pilB* mutation library are compared to their planktonic fraction and the controls to determine the fitness estimates of genes in the settled and planktonic fraction.

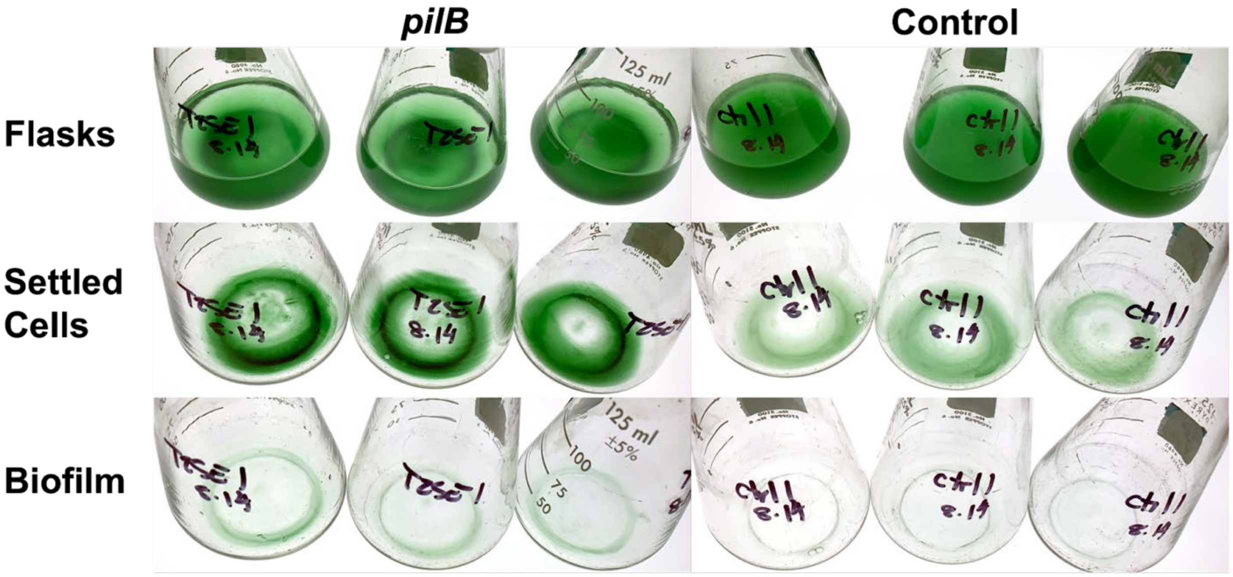


Figure 5-3: Representative biofilm IRB-Seq experiment flasks. Flasks of *pilB* and control-mutation cultures before the planktonic cells are removed, the settled fraction of the culture, and the biofilm fraction of the culture.

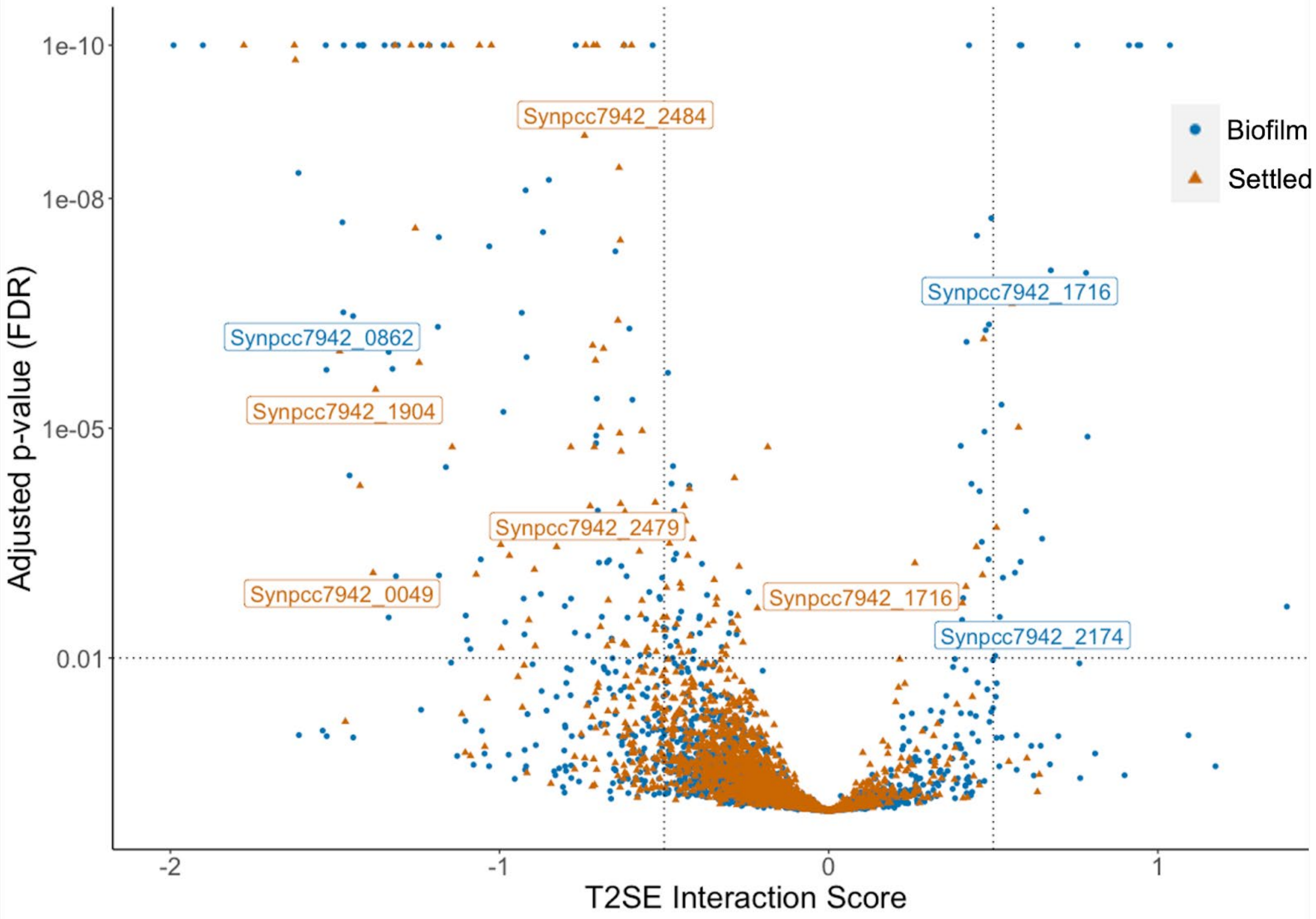


Figure 5-4: Volcano plot of IRB-Seq biofilm fitness estimates. The biofilm and settled fractions of the *pilB* interaction with the RB-TnSeq library. Genes above the horizontal dashed line have FDR < 0.01 (Linear mixed-effects model). Absolute values of fitness estimates of 1 are indicated by vertical dotted lines. All points with FDR < 10^{-10} are plotted as FDR = 10^{-10}.

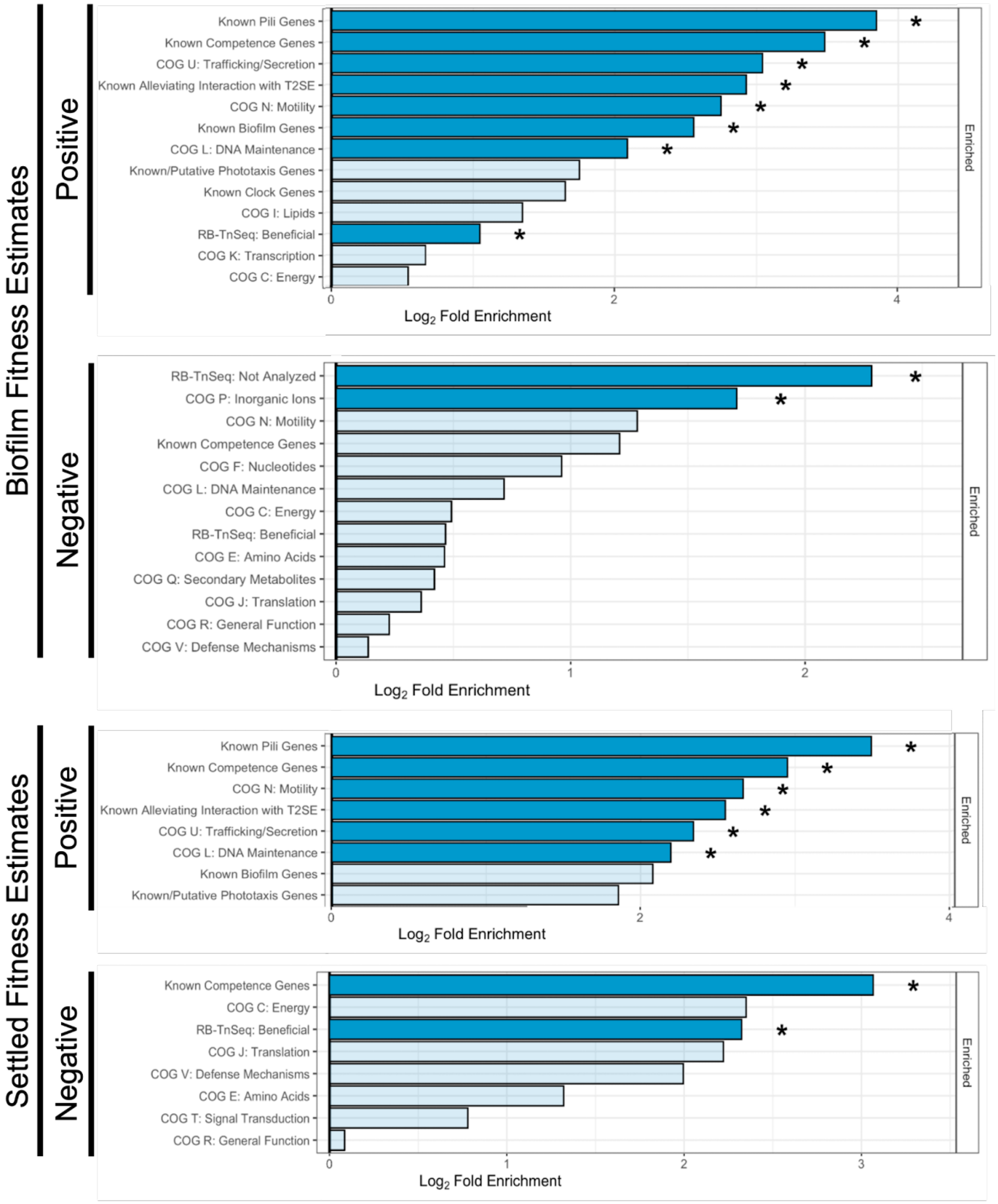


Figure 5-5: Gene set enrichment analysis for IRB-Seq biofilm experiment. Plots showing enriched categories in the gene sets with significant (FDR < 0.01) positive and negative fitness estimates in either the biofilm or settled fractions.

Table 5-1: List of genes and mutations with known effects on biofilm formation in PCC 7942.

PCC 7942 locus	Gene	Description	Knockout phenotype	Reference
Synpcc7942_0048	pilA	type IV pilus assembly protein PilA	biofilm	Nagar et al. 2017
Synpcc7942_0049	pilA	type IV pilus assembly protein PilA	biofilm	Nagar et al. 2017
Synpcc7942_0862	ebsA	hypothetical protein	biofilm	Nagar et al. 2017
Synpcc7942_1127	ebfE	Unknown type peptidase	suspended phenotype in double KO with *pilB*	Parnasa et al. 2019
Synpcc7942_1133	pteB	cysteine peptidase	suspended phenotype in double KO with *pilB*	Schatz et al. 2013; Parnasa et al. 2016
Synpcc7942_1134	Ebf1	peptide enabling biofilm formation	suspended phenotype in double KO with *pilB*	Schatz et al. 2013; Parnasa et al. 2016
Synpcc7942_ebf2	Ebf2	peptide enabling biofilm formation	suspended phenotype in double KO with *pilB*	Schatz et al. 2013; Parnasa et al. 2016
Synpcc7942_ebf3	Ebf3	peptide enabling biofilm formation	suspended phenotype in double KO with *pilB*	Schatz et al. 2013; Parnasa et al. 2016
Synpcc7942_ebf4	Ebf4	peptide enabling biofilm formation	suspended phenotype in double KO with *pilB*	Schatz et al. 2013; Parnasa et al. 2016
Synpcc7942_1926	hfq	host factor-**I** protein	biofilm	Yegorov et al. 2021
Synpcc7942_2069	pilC	type IV pilus assembly protein PilC	biofilm; altered secretion profile	Schatz et al. 2013; Nagar et al. 2017
Synpcc7942_2071	pilB	type IV pilus assembly protein PilB	biofilm; altered secretion profile	Schatz et al. 2013
Synpcc7942_2452	pilN	type IV pilus assembly protein PilN	biofilm	Nagar et al. 2017
Synpcc7942_2479	pilA2	type **II** secretion system protein	biofilm	Nagar et al. 2017

Table 5-2: IRB-Seq alleviating interactions. Alleviating reactions with *pilB*, those genes with known essential roles in competence in PCC 7942 are highlighted.

PCC7942 locus	Description	COG Description	Fitness estimate	FDR
Synpcc7942_0032	DUF86 domain-containing protein	Defense mechanisms	1.0315	6.20E-03
Synpcc7942_0049	type IV pilus assembly protein PilA	Cell motility	1.1482	1.86E-03
Synpcc7942_0100	hypothetical protein		1.2158	1.62E-10
Synpcc7942_0168	hypothetical protein		1.6277	1.12E-20
Synpcc7942_0252	hypothetical protein	Signal transduction mechanisms	1.0340	2.64E-03
Synpcc7942_0302	competence protein ComE	Lipid transport and metabolism	1.4354	8.88E-41
Synpcc7942_0361	CP12 domain-containing protein	Signal transduction mechanisms	1.3770	5.77E-04
Synpcc7942_0418	Protein of unknown function DUF1257		1.1000	1.41E-03
Synpcc7942_0529	6-phosphogluconolactonase	Carbohydrate transport and metabolism	1.1068	7.28E-10
Synpcc7942_0556	two component transcriptional regulator, winged helix family	Signal transduction mechanisms	1.0312	3.75E-07
Synpcc7942_0622	ABC transporter ATP-binding protein	General function prediction only	3.2165	5.17E-121
Synpcc7942_0675	CCB1 domain-containing protein		1.2785	1.58E-12
Synpcc7942_0703	DNA processing protein	Replication, recombination and repair	1.5207	4.06E-41
Synpcc7942_0764	HTH transcriptional regulator	Transcription	2.1102	4.90E-05
Synpcc7942_0862	hypothetical protein		1.5629	4.29E-25
Synpcc7942_0940	transcriptional regulator, XRE family		1.2176	8.57E-04
Synpcc7942_0956	GTPase, G3E family	General function prediction only	1.2089	3.94E-10
Synpcc7942_0994	hypothetical protein		1.1216	5.20E-12
Synpcc7942_1059	Uncharacterized membrane protein	Function unknown	1.0346	2.51E-08
Synpcc7942_1078	hypothetical protein		1.2988	1.88E-19
Synpcc7942_1108	bacterial nucleoid protein Hbs	Replication, recombination and repair	1.6952	6.06E-06
Synpcc7942_1174	photosystem II PsbJ protein		1.1870	3.74E-03
Synpcc7942_1184	6-pyruvoyltetrahydropterin/6-carboxytetrahydropterin synthase	Coenzyme transport and metabolism	1.0413	8.61E-09

Table 5-2 (cont.): IRB-Seq alleviating interactions.

PCC7942 locus	Description	COG Description	Fitness estimate	FDR
Synpcc7942_1250	photosystem I subunit 3		1.1025	5.23E-03
Synpcc7942_1288	Protein of unknown function DUF4240	Transcription	1.0646	9.01E-05
Synpcc7942_1301	ATP-dependent DNA helicase RecQ	Replication, recombination and repair	1.3151	2.35E-45
Synpcc7942_1460	Hypothetical protein Ycf34		1.0875	2.78E-03
Synpcc7942_1473	multisubunit sodium/proton antiporter, MrpD subunit	Energy production and conversion	1.1712	2.46E-06
Synpcc7942_1476	hypothetical protein	Signal transduction mechanisms	1.3310	7.48E-05
Synpcc7942_1567	hypothetical protein		1.1893	3.45E-03
Synpcc7942_1596	NADP-dependent 3-hydroxy acid dehydrogenase YdfG	Energy production and conversion	1.3613	2.14E-10
Synpcc7942_1638	hypothetical protein		1.4243	4.32E-05
Synpcc7942_1703	alpha-mannosidase	Carbohydrate transport and metabolism	1.0614	1.88E-51
Synpcc7942_1780	DNA mismatch repair protein MutL	Replication, recombination and repair	1.3526	4.06E-41
Synpcc7942_1784	sigma-70 family RNA polymerase sigma factor	Transcription	1.1529	1.16E-16
Synpcc7942_1794	Aspartate/methionine/tyrosine aminotransferase	Amino acid transport and metabolism	1.0340	1.91E-11
Synpcc7942_1849	RNA polymerase, sigma subunit, RpsC/SigC	Transcription	1.1760	5.97E-15
Synpcc7942_1919	transcriptional regulator, XRE family		1.0026	1.34E-13
Synpcc7942_1926	host factor-I protein	Signal transduction mechanisms	1.3752	1.59E-06
Synpcc7942_2006	hypothetical protein	Replication, recombination and repair	1.1006	2.10E-09
Synpcc7942_2023	ribosome maturation factor RimP	Translation, ribosomal structure and biogenesis	1.1402	7.62E-03
Synpcc7942_2069	type IV pilus assembly protein PilC	Cell motility	1.6059	9.81E-20
Synpcc7942_2070	twitching motility protein PilT	Cell motility	1.6446	2.95E-22
Synpcc7942_2071	type IV pilus assembly protein PilB	Cell motility	1.5626	1.84E-84

Table 5-2 (cont.): IRB-Seq alleviating interactions.

PCC7942 locus	Description	COG Description	Fitness estimate	FDR
Synpcc7942_2124	Acetoin utilization deacetylase AcuC	Secondary metabolites biosynthesis, transport and catabolism	1.1758	1.82E-06
Synpcc7942_2215	LSU ribosomal protein L15P	Translation, ribosomal structure and biogenesis	1.4668	6.66E-06
Synpcc7942_2242	histidine kinase	Signal transduction mechanisms	1.0836	2.67E-15
Synpcc7942_2255	AbrB-like transcriptional regulator	Transcription	1.1238	7.05E-05
Synpcc7942_2371	acyl-UDP-N-acetylglucosamine O-acyltransferase	Cell wall/membrane/envelope biogenesis	1.2914	5.88E-05
Synpcc7942_2404	ComF family protein	General function prediction only	1.3714	7.49E-19
Synpcc7942_2444	phosphate ABC transporter substrate-binding protein, PhoT family	Inorganic ion transport and metabolism	1.1374	8.74E-04
Synpcc7942_2450	type IV pilus assembly protein PilQ	Intracellular trafficking, secretion, and vesicular transport	1.4947	1.55E-78
Synpcc7942_2451	type IV pilus assembly protein PilO	Cell motility	1.3949	1.04E-31
Synpcc7942_2452	type IV pilus assembly protein PilN	Cell motility	1.4978	1.05E-42
Synpcc7942_2453	type IV pilus assembly protein PilM	Cell motility	1.4445	1.18E-54
Synpcc7942_2458	competence protein ComEC	Intracellular trafficking, secretion, and vesicular transport	1.3906	2.86E-87
Synpcc7942_2464	N-acylglucosamine-6-phosphate 2-epimerase	Carbohydrate transport and metabolism	1.0249	4.74E-12
Synpcc7942_2485	hypothetical protein required for natural transformation	Cell motility	1.3540	5.26E-13
Synpcc7942_2486	hypothetical protein required for natural transformation		1.6901	2.26E-62
Synpcc7942_2487	NDH-1 subunit O		1.3194	1.38E-03

Table 5-2 (cont.): IRB-Seq alleviating interactions.

PCC7942 locus	Description	COG Description	Fitness estimate	FDR
Synpcc7942_2590	prepilin-type N-terminal cleavage/methylation domain-containing protein	Cell motility	1.2988	5.23E-18
Synpcc7942_2591	hypothetical protein	Intracellular trafficking, secretion, and vesicular transport	1.3956	7.56E-24

References

1. Nagar E, Zilberman S, Sendersky E, Simkovsky R, Shimoni E, Gershtein D, Herzberg M, Golden SS, Schwarz R. 2017. Type 4 pili are dispensable for biofilm development in the cyanobacterium *Synechococcus elongatus*. Environ Microbiol 19:2862–2872.

2. Schatz D, Nagar E, Sendersky E, Parnasa R, Zilberman S, Carmeli S, Mastai Y, Shimoni E, Klein E, Yeger O, Reich Z, Schwarz R. 2013. Self-suppression of biofilm formation in the cyanobacterium *Synechococcus elongatus*. Environ Microbiol 15:1786–1794.

3. Parnasa R, Nagar E, Sendersky E, Reich Z, Simkovsky R, Golden S, Schwarz R. 2016. Small secreted proteins enable biofilm development in the cyanobacterium *Synechococcus elongatus*. Sci Rep 6:32209.

4. Berry J-L, Pelicic V. 2015. Exceptionally widespread nanomachines composed of type IV pilins: the prokaryotic Swiss Army knives. FEMS Microbiol Rev 39:134–154.

5. Rubin BE, Wetmore KM, Price MN, Diamond S, Shultzaberger RK, Lowe LC, Curtin G, Arkin AP, Deutschbauer A, Golden SS. 2015. The essential gene set of a photosynthetic organism. Proc Natl Acad Sci U S A 112:E6634–43.

6. Taton A, Erikson C, Yang Y, Rubin BE, Rifkin SA, Golden JW, Golden SS. 2020. The circadian clock and darkness control natural competence in cyanobacteria. Nat Commun 11:1688.

7. Welkie DG, Rubin BE, Chang Y-G, Diamond S, Rifkin SA, LiWang A, Golden SS. 2018. Genome-wide fitness assessment during diurnal growth reveals an expanded role of the cyanobacterial circadian clock protein KaiA. Proc Natl Acad Sci U S A 115:E7174–E7183.

8. Li S, Sun T, Xu C, Chen L, Zhang W. 2018. Development and optimization of genetic toolboxes for a fast-growing cyanobacterium *Synechococcus elongatus* UTEX 2973. Metab Eng 48:163–174.

9. Dixon SJ, Costanzo M, Baryshnikova A, Andrews B, Boone C. 2009. Systematic mapping of genetic interaction networks. Annu Rev Genet 43:601–625.

10. López-Calcagno PE, Howard TP, Raines CA. 2014. The CP12 protein family: a thioredoxin-mediated metabolic switch? Front Plant Sci 5:9.

11. Price MN, Wetmore KM, Waters RJ, Callaghan M, Ray J, Liu H, Kuehl JV, Melnyk RA, Lamson JS, Suh Y, Carlson HK, Esquivel Z, Sadeeshkumar H, Chakraborty R, Zane GM, Rubin BE, Wall JD, Visel A, Bristow J, Blow MJ, Arkin AP, Deutschbauer AM. 2018. Mutant phenotypes for thousands of bacterial genes of unknown function. Nature 557:503–509.

12. Yegorov Y, Sendersky E, Zilberman S, Nagar E, Waldman Ben-Asher H, Shimoni E, Simkovsky R, Golden SS, LiWang A, Schwarz R. 2021. A cyanobacterial component required for pilus biogenesis affects the exoproteome. MBio 12.

13. Simkovsky R, Effner EE, Iglesias-Sánchez MJ, Golden SS. 2016. Mutations in novel lipopolysaccharide biogenesis genes confer resistance to amoebal grazing in *Synechococcus elongatus*. Appl Environ Microbiol 82:2738–2750.

14. Simkovsky R, Daniels EF, Tang K, Huynh SC, Golden SS, Brahamsha B. 2012. Impairment of O-antigen production confers resistance to grazing in a model amoeba-cyanobacterium predator-prey system. Proc Natl Acad Sci U S A 109:16678–16683.

15. Agostoni M, Montgomery BL. 2014. Survival strategies in the aquatic and terrestrial world: the impact of second messengers on cyanobacterial processes. Life 4:745–769.

16. Xu M, Su Z. 2009. Computational prediction of cAMP receptor protein (CRP) binding sites in cyanobacterial genomes. BMC Genomics 10:23.

17. Song W-Y, Zang S-S, Li Z-K, Dai G-Z, Liu K, Chen M, Qiu B-S. 2018. Sycrp2 Is Essential for Twitching Motility in the Cyanobacterium Synechocystis sp. Strain PCC 6803. J Bacteriol 200.

18. Rubin BE, Huynh TN, Welkie DG, Diamond S, Simkovsky R, Pierce EC, Taton A, Lowe LC, Lee JJ, Rifkin SA, Woodward JJ, Golden SS. 2018. High-throughput interaction screens illuminate the role of c-di-AMP in cyanobacterial nighttime survival. PLoS Genet 14:e1007301.

19. Clerico EM, Ditty JL, Golden SS. 2007. Specialized techniques for site-directed mutagenesis in cyanobacteria, p. 155–171. *In* Rosato, E (ed.), Circadian Rhythms: Methods and Protocols. Humana Press, Totowa, NJ.

20. Wetmore KM, Price MN, Waters RJ, Lamson JS, He J, Hoover CA, Blow MJ, Bristow J, Butland G, Arkin AP, Deutschbauer A. 2015. Rapid quantification of mutant fitness in diverse bacteria by sequencing randomly bar-coded transposons. MBio 6:e00306–15.

21. Parnasa R, Sendersky E, Simkovsky R, Waldman Ben-Asher H, Golden SS, Schwarz R. 2019. A microcin processing peptidase-like protein of the cyanobacterium *Synechococcus elongatus* is essential for secretion of biofilm-promoting proteins. Environ Microbiol Rep 11:456–463.

CHAPTER 6: Conclusions

This work expanded the tools and information available to study complex phenotypes in *Synechococcus elongatus*, particularly phenotypes of phototaxis and biofilm formation. This was first accomplished through the development of *S. elongatus* UTEX 3055 as a model organism for phototaxis and biofilm research. The sequencing and annotation of the genome of this recent and novel isolate of *S. elongatus* led to the determination of an operon essential for phototaxis and made additional genome comparison analysis possible.

A comprehensive comparative genomics analysis in this work was the scaffold that supported the multiple discoveries in diverse topics relating to the biology of *S. elongatus*. Perhaps the finding with the potential for the largest impact in the cyanobacterial research community is the correction of the type strain PCC 6301 genome sequence, bringing it closer to the sequences of the legacy strains, specifically to UTEX 2793 that is presumably derived from it. The analysis also led to a pangenome analysis and annotation, making any future comparative analysis using *S. elongatus* strain data easier by standardizing and collating available annotations and metadata. The main proposed use of the comparative genome analysis, to use it to find specific loci or nucleotide changes that explain complex phenotypes, was a success. The analysis identified specific loci that explain a difference in pigmentation and phototaxis phenotypes between UTEX 3055 and other legacy *S. elongatus* strains as well as showing that the patterns of shared and unique SNPs and genes between UTEX 3055 and the legacy strains are compatible with a domestication hypothesis. The genomic analysis also helped explain a confusing allele present in PCC 7942 that is neither WT nor a sequencing error, but derives from a rare mutant clone used for the published reference sequence for PCC 7942.

The RB-TnSeq library of PCC 7942 has proven to be an invaluable tool for interrogating the genetics of phenotypes in a cost-effective and labor-effective manner, and as part of this work I developed a complementary library in UTEX 3055 that is a beneficiary of the great depth of information that has been synthesized through the use of the PCC 7942 library. The library of UTEX 3055 is dense enough to be used to screen for genes involved in phenotypes that are difficult or impossible to assay in PCC 7942 such as biofilm formation and phototaxis, and can be used in tandem with the PCC 7942 library to interrogate fitness effects across the entire *S. elongatus* pangenome. Experiments that mix the two libraries are also possible, because the unique barcode of each insertion is assigned to its respective genome. Such experiments will enable questions to be addressed related to the fitness contributions of unique portions of UTEX 3055, such as CRISPR and toxin-antitoxin systems, when presented with complex microbial assemblages such as water from Waller Creek.

In addition to genome comparison analysis and the creation of an RB-TnSeq library in UTEX 3055, an IRB-Seq approach in the PCC 7942 library was used to find additional genes in the biofilm formation pathway of PCC 7942. The analysis of this IRB-Seq experiment identified genes that can be investigated further to expand the current model of biofilm formation in *S. elongatus* beyond the pathway controlled by constitutive repression, such as transcriptional regulators and genes involved in the metabolism of the second messengers cAMP and c-di-GMP.

The combined approach of using comprehensive comparative genome analysis and deep mutant screens to interrogate complex phenotypes in *S. elongatus* not only bears fruit in revealing targets for exploring phototaxis and biofilm formation, but also provides tools for future researchers to use and build upon, such as an RB-TnSeq library in UTEX 3055, new sequence data for model strains, improved annotation and organism metadata, and a pangenome of *S. elongatus*.