

UC San Diego

UC San Diego Electronic Theses and Dissertations

Title

Massive Mutant Screens to Illuminate the Dark Side of the Cyanobacterium

Permalink

<https://escholarship.org/uc/item/7zk2p5mz>

Author

Rubin, Benjamin Emery

Publication Date

2017

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Massive Mutant Screens to Illuminate the Dark Side of the Cyanobacterium

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

in

Biology

by

Benjamin E. Rubin

Committee in Charge:

Susan S. Golden, Chair
Eric E. Allen
Mark Hildebrand
Stephen P. Mayfield
Brian Palenik

2017

Copyright

Benjamin E. Rubin, 2017

All rights reserved.

The Dissertation of Benjamin E. Rubin is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California, San Diego

2017

TABLE OF CONTENTS

SIGNATURE PAGE	iii
TABLE OF CONTENTS	iv
LIST OF FIGURES.....	v
LIST OF TABLES.....	vi
LIST OF SUPPLEMENTARY FIGURES.....	vii
LIST OF SUPPLEMENTARY TABLES	viii
ACKNOWLEDGMENTS.....	ix
VITA	xiii
ABSTRACT OF THE DISSERTATION	xv
CHAPTER 1: Introduction	1
1.1 Chapter Summary	1
1.2 Cyanobacterial Day-Night Physiology	3
1.3 RB-TnSeq	15
1.4 Acknowledgments.....	17
CHAPTER 2: Physiology Under Light-Dark Transitions.....	18
2.1 Chapter Summary	18
2.2 Physiology of the Dark to Light Transition	19
2.3 Physiology of the Light to Dark Transition	36
2.4 Acknowledgments.....	53
CHAPTER 3: Associating Genotype to Phenotype with RB-TnSeq.....	54
3.1 Chapter Summary.....	54
3.2 RB-TnSeq Development and Essential Genes in <i>S. elongatus</i>	56
3.3 RB-TnSeq Guided Metabolic Modeling	73
3.4 RB-TnSeq Screens: Biofilm Formation.....	92
3.5 RB-TnSeq Screens: Amoeba Grazing.....	95
3.6 RB-TnSeq Screens: Analysis Across Conditions	98
3.7 Acknowledgments.....	99
CHAPTER 4: Applying RB-TnSeq to Light-Dark Physiology.....	100
4.1 Chapter Summary.....	100
4.2 RB-TnSeq Under Light-Dark Cycles.....	102
4.3 The Role of c-di-AMP in Light-Dark Cycles	123
4.4 Acknowledgments.....	152
CHAPTER 5: Conclusion	153
5.1 Discussion	153
5.2 Future Directions	155

LIST OF FIGURES

Figure 1.3-1: RB-TnSeq approach	16
Figure 2.2-1: Overview of shared metabolic pathways	21
Figure 2.2-2: Average of normalized glycogen content	22
Figure 2.2-3: Summary of glycogen accumulation data over a 12-h period	22
Figure 2.2-4: Summary of glycogen degradation data	23
Figure 2.2-5: Summary of dimension reduction performed on metabolomics ..	24
Figure 2.2-6: Summary of metabolites that differ significantly	25
Figure 2.2-7: Heatmap of the correlation between the groupings	26
Figure 2.3-1: Absorbance and viability data	37
Figure 2.3-2: Summary of metabolic changes in WT and $\Delta rpaA$	38
Figure 2.3-3: Metabolic changes	39
Figure 2.3-4: Summary of enriched KEGG functional categories	40
Figure 2.3-5: Summary of data on MSM-	41
Figure 2.3-6: Plot of H ₂ DCFDA fluorescence over a 24-h LD cycle	42
Figure 3.2-1: The distribution of transposon mutations in the library	57
Figure 3.2-2: The determination of gene essentiality	58
Figure 3.2-3: Comparing the essential gene set	59
Figure 3.2-4: Gene essentiality in central metabolism	60
Figure 3.2-5: Essentiality of ncRNAs	61
Figure 3.2-6: Essentiality regulatory regions	62
Figure 3.3-1: Deriving light uptake rates for the GEM	75
Figure 3.3-2: Modeling the self-shading of cultures	76
Figure 3.3-3: Comparison of <i>in vivo</i> vs. <i>in silico</i> gene essentiality	77
Figure 3.3-4: Central carbon metabolism	78
Figure 3.3-5: The TCA pathway	80
Figure 3.4-1: RB-TnSeq for biofilm formation	94
Figure 3.5-1: HGG1 and LPG1 grazing	97
Figure 4.2-1: LDC sensitive genes	115
Figure 4.2-2: Growth and circadian rhythms of KaiA mutants	116
Figure 4.2-3: Phosphorylation time-course	117
Figure 4.3-1: Presence, synthase, and light dependence of c-di-AMP	142
Figure 4.3-2: Sensitivity of <i>dacA</i> mutant to LDCs	143
Figure 4.3-3: Identifying c-di-AMP binding proteins	144
Figure 4.3-4: IRB-Seq approach to genetic interaction screens	145
Figure 4.3-5: DacA genetic interactions using IRB-Seq	146

LIST OF TABLES

Table 3.3-1. Comparison of essentiality results <i>i</i> JB785 and <i>i</i> Syf715	77
Table 4.2-1. Strains	114
Table 4.3-1. C-di-AMP binding candidates.....	141

LIST OF SUPPLEMENTARY FIGURES

Figure 2.2-S1: Glycogen content in WT and $\Delta kaiC$	29
Figure 2.2-S2: Graphical representation of coefficients	29
Figure 2.2-S3: KEGG functional category analysis	30
Figure 2.2-S4: Enlarged area of heatmap from Fig. 7	31
Figure 2.2-S5: Plot of components 1 and 2 derived from PCA	31
Figure 2.2-S6: The effect of \log_2 normalization on raw metabolite abundance	32
Figure 2.3-S1: Summary of phenotypic effects of darkness.....	48
Figure 2.3-S2: Comparison of metabolite abundance	48
Figure 2.3-S3: Changes in expression of <i>glnN</i>	49
Figure 2.3-S4: Supporting data for EMS mutagenesis of the $\Delta rpaA$ mutant...	49
Figure 2.3-S5: Supporting data for potential mechanisms that suppress LD ..	50
Figure 2.3-S6: Controls for LD plating and ROS experiments	50
Figure 2.3-S7: Diagram of the growth and sampling scheme	51
Figure 2.3-S8: Supporting data for metabolomics statistical analysis	51
Figure 3.2-S1: Divergence of antibiotic and no antibiotic library outgrowths...	66
Figure 3.2-S2: Lack of sequence bias around insertion locations	67
Figure 3.2-S3: Normalizing for GC bias of insertion density	68
Figure 3.2-S4: Genotypic characterization of the recreated <i>tal</i> mutant	69
Figure 3.3-S1: Model development flowchart.....	86
Figure 3.3-S2: <i>In vivo</i> physiological data incorporated as constraints	87
Figure 3.3-S3: Diagram of cell-shading calculation	87
Figure 3.3-S4: Essential genes in <i>S. elongatus</i> not present in <i>iJB785</i>	88
Figure 3.3-S5: Nucleotide salvage pathway	88
Figure 3.3-S6: Photorespiration reactions	89
Figure 3.3-S7: Genotypic characterization and growth curve of the TCA	90

LIST OF SUPPLEMENTARY TABLES

Table 2.2-1. Metabolites with Significant Changes Between 0 h and 4 h	33
Table 2.2-2. Metabolites with Significant Difference in Abundance at 4 h	33
Table 2.2-3. Detailed Information on Each Target Cluster	34
Table 2.2-4. Cyanobacterial Strains Used in This Study	35
Table 2.3-1. Absorbance ratio (630/680 nm)	52
Table 3.2-1. The major variations of techniques used for conjugation	69
Table 3.2-2. Essential genes in central carbon metabolism	70
Table 3.2-3. Essential genes for the photosynthetic lifestyle	71
Table 3.3-1. Structural homology analysis of <i>S. elongatus</i> PGMs	91
Table 3.3-2. Primers for validation of the TCA pathway mutants	91

ACKNOWLEDGMENTS

I would like to first thank Professor Susan Golden. It goes without saying that I have learned from you a great deal about how to conduct science, but just as valuable is what I've learned about leadership. You emit a sense of calm that is contagious, lifting me in my lows and suggesting care in my highs. You manage our scientific egos so deftly that tensions in the lab are all but absent. Most importantly, you have been squarely in my court since day one, never overbearing, but always there when I'm in need. This has even resulted, at times, in you shifting opportunities and accolades aimed at you, towards me. For all this, and for developing me as a scientist, I am forever grateful.

I'm lucky enough to have had two labmates that have become great friends as well as collaborators. I am of course referring to David Welkie and Spencer Diamond. Spencer, you taught me most of what I know about Cyanobacteria and also that it was cool to talk to myself in the lab. You were always jovial on late nights of sampling, and if there was a way to salvage a collapsing experiment you knew it. David, you have become the close friend, collaborator, and the confidant that I lost when Spencer went on to his postdoc. You are perhaps the kindest person I have ever met, and that is certainly a refreshing characteristic to have next-door. You also have the best graphic design sense that I have run into, and I hope I have learned something from it. I thank both of you for being part of my PhD experience.

I have also had the opportunity to work with three exceptional undergrads. Laura, Jenny, and Gene it has been a privilege to be part of team BG11 with you.

Through the good times and the bad, you have stuck with it, and probably taught me more about mentoring than I taught you about science. Your contributions to my PhD are as essential as the genes we studied.

I would like to thank my committee members for each providing me with some crucial piece of wisdom at one point or another that has influenced my science and my future. I want to specifically thank Dr. Mark Hildebrand for sitting with me on multiple occasions and giving kind thoughtful advice on my project and my life, and Dr. Stephen Mayfield for pushing me towards several opportunities that advanced my career.

Finally I want to thank my family and friends. My parents and sister have provided me, throughout my PhD, with constant and unconditional support and love. I simply cannot imagine how I could have made it in life without this. The three of you are my biggest role models, each in different ways. Similarly I have leaned on many of my friends throughout the course of my PhD and they have never hesitated to lift me up when I was in need. Three people that have been particularly important in this respect are Jamie, my best friend and better half; Jarod, who has been an unimaginably perfect roommate; and Emily a friend who is as kind as she is thoughtful.

Chapter 1, in part, is made up of material from a manuscript in preparation for publication and an invention disclosure. The manuscript in preparation is: Rubin BE, Welkie DG, Diamond S, Hood RD, Savage DF, and Golden SS. The dark side of the cyanobacterium. (*In preparation*). The dissertation author is the primary author of this manuscript in preparation. The disclosure is: Rubin BE,

Wetmore KM, Price MN, Arkin AP, Deutschbauer A, and Golden SS (2015) Saturated Cyanobacterial Mutant Library Optimized for Screening.

Disclosure Case Number: 2016:013. The dissertation author was the primary author of this disclosure.

Chapter 2, in part, is made up of reprints of two published manuscripts. The first is: Diamond S, Jun D, Rubin BE, and Golden SS (2015) The circadian oscillator in *Synechococcus elongatus* controls metabolite partitioning during diurnal growth. *Proc Natl Acad Sci* 112(15). The dissertation author was the tertiary author of this paper. The second is: Diamond S, Rubin BE, Shultzaberger RK, Chen Y, Barber DB, Golden SS. (2017) Redox crisis underlies conditional light-dark lethality in cyanobacterial mutants that lack the circadian regulator, RpaA. *Proc Natl Acad Sci* 114(4). The dissertation author was the secondary author of this paper.

Chapter 3, in part, is made up of reprints of two published manuscripts and two manuscripts in preparation. The first published manuscript in section **3.2 RB-TnSeq Development and Essential Genes in *S. elongatus*** is: Rubin BE et al. (2015) The essential gene set of a photosynthetic organism. *Proc Natl Acad Sci*. 112(48). The dissertation author was the primary author of this paper. The second published manuscript in section **3.3 RB-TnSeq Guided Metabolic Modeling** is: Broddrick J*, Rubin BE*, Welkie DG*, Nui D, Nathan M, Diamond S, Jenny JL, 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* 113(51). The dissertation author was

the co-primary author of this paper. The first manuscript in preparation in section **3.4 RB-TnSeq Screens: Biofilm Formation** is: Simkovsky R, Rubin BE, Wang J, Nagar E, Parnasa R, Yevgeni J, Veltman B, Sendersky E, Schwarz R, Golden SS. RNA-Seq vs RB-TnSeq: Each reveal critical elements in biofilm repression and formation in the cyanobacterium *Synechococcus elongatus* PCC 7942. (In Preparation). The dissertation author will be the secondary author of this paper. The Second manuscript in preparation in section **3.5 RB-TnSeq Screens: Amoeba Grazing** is: Ota M, Simkovsky R, Rubin BE, Golden SS, Brahamsha B, Golden JW. A genome-wide fitness assay identifies determinants of cyanobacterial resistance against protozoan grazing. (*In preparation*). The dissertation author will be the tertiary author of this paper.

Chapter 4, in part, is made up of reprints of two manuscripts in preparation. The first manuscript in preparation in section **4.2 RB-TnSeq Under Light-Dark Cycles** is: Welkie DG, Rubin BE, Chang YG, Diamond S, LiWang A, Golden SS. The essential role of the circadian clock protein KaiA for survival of *Synechococcus elongatus* sp. PCC 7942 in alternating light-dark environments. (In Preparation). The dissertation author will be the secondary author of this paper. The second manuscript in preparation in section **4.3 The Role of c-di-AMP in Light-Dark Cycles** is: Rubin BE, Huynh TA, Welkie DG, Diamond S, Lowe LC, Jenny JL, Woodward JJ, Golden SS. High-throughput interaction screens illuminate the role of c-di-AMP in cyanobacterial nighttime survival. (In Preparation). The dissertation author will be the primary author of this paper.

VITA

EDUCATION

- University of California San Diego, San Diego, CA** 2017
PhD, Biology
- Middlebury College, Middlebury, VT** 2011
Honors Bachelor of Arts, Double Major in Molecular Biology and
Biochemistry/Economics

HONORS & AWARDS

- Invited Speaker at the ARC Centre of Excellence in Australia 2017
- Selected for Plenary Speech at The AGBT General Meeting 2017
- Invited Speaker at Chloroplast Biotechnology Gordon Conference 2017
- Excellence in Graduate Research Award & Speaker 2016
- Biology Founding Faculty Award for Graduate Excellence, UCSD 2016
- Goeddel Chancellor's Award, UCSD 2016
- UCSD Representative for Graduate Research Advocacy Day 2016
- Achievement Reward for College Scientists (ARCS), UCSD 2015
- Goeddel Graduate Scholar Award, UCSD 2015
- The Biology Program Award for Teaching Excellence, UCSD 2014
- 3rd place in Grad Slam, UCSD 2014

PUBLICATIONS

1. ***Rubin BE** et al. (2015) The essential gene set of a photosynthetic organism. *Proc Natl Acad Sci*. 112(48).
***Commentary**: Whitman WB (2015) Insights in the life of an oxygenic phototroph. *Proc Natl Acad Sci*. 112(48).
2. Broddrick J*, **Rubin BE***, Welkie DG*, Nui D, Nathan M, Diamond S, Jenny JL, 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* 113(51).
***Co-first Authors**
3. Diamond S, **Rubin BE**, Shultzaberger RK, Chen Y, Barber DB, Golden SS. (2017) Redox crisis underlies conditional light-dark lethality in cyanobacterial mutants that lack the circadian regulator, RpaA. *Proc Natl Acad Sci* 114(4)
4. Diamond S, Jun D, **Rubin BE**, and Golden SS (2015) The circadian oscillator in *Synechococcus elongatus* controls metabolite partitioning during diurnal growth. *Proc Natl Acad Sci* 112(15).

5. Price MN, Wetmore KM, Waters RJ, Callaghan M, Ray J, Kuehl JV, Melyk RA, Lamson JS, Suh Y, Esquivel Z, Sadeeshkumar H, Chakraborty R, **Rubin BE**, Bristow J, Blow MJ, Arkin AP, Deutschbauer AM. (2017) Deep annotation of protein function across diverse bacteria from mutant phenotypes. *Nature*. (In review)

PUBLICATIONS IN PREPARATION

1. **Rubin BE**, Huynh TA, Welkie DG, Diamond S, Lowe LC, Jenny JL, Woodward JJ, Golden SS. High-throughput interaction screens illuminate the role of c-di-AMP in cyanobacterial nighttime survival. (In Preparation).
2. **Rubin BE**, Welkie DG, Diamond S, Hood RD, Savage DF, and Golden SS. The dark side of the cyanobacterium. *Trends in Microbiology* (Review solicited for submission June 1st).
3. Welkie DG, **Rubin BE**, Chang YG, Diamond S, LiWang A, Golden SS. The essential role of the circadian clock protein KaiA for survival of *Synechococcus elongatus* sp. PCC 7942 in alternating light-dark environments. (In Preparation).
4. Simkovsky R, **Rubin BE**, Wang J, Nagar E, Parnasa R, Yevgeni J, Veltman B, Sendersky E, Schwarz R, Golden SS. RNA-Seq vs RB-TnSeq: Each reveal critical elements in biofilm repression and formation in the cyanobacterium *Synechococcus elongatus* PCC 7942. (In Preparation).
5. Ota M, Simkovsky R, **Rubin BE**, Golden SS, Brahamsha B, Golden JW. A genome-wide fitness assay identifies determinants of cyanobacterial resistance against protozoan grazing. (*In preparation*).

GRANTS

- Cell Molecular Genetics Training Program **2013-2016**

BROADER IMPACTS AND TEACHING

- Guest Lecturer at San Diego State University **2016**
- SciChats **2015**
- Frazier Elementary Mentoring Program **2012-2015**
- Biology Peer Mentor **2013-2016**
- Organized BGSE205 Microbial Biology Section **2014-2015**
- Head TA for Circadian Biology **2014-2014**
- Officer of Outreach for BioEASI **2012-2014**
- Biology PhD Recruitment Committee **2013-2014**
- Award for Teaching Excellence **2013-2013**
- Guest Lecturer at UCSD **2013**

ABSTRACT OF THE DISSERTATION

Massive Mutant Screens to Illuminate the Dark Side of the Cyanobacterium

by

Benjamin E. Rubin

Doctor of Philosophy in Biology

University of California, San Diego, 2017

Professor Susan S. Golden, Chair

Cyanobacteria are key primary producers in the environment, models for photosynthesis and the circadian clock in the lab, and emerging biological production platforms for industry. Despite their scope of importance some of the most fundamental components of their biology are poorly understood. We know little about the functions encoded by most of their genes, and research traditionally focused on constant light has left the organisms' physiology in day-night conditions obscure. In the studies comprising my thesis, we worked to illuminate both of these knowledge gaps in the model cyanobacterium

Synechococcus elongatus PCC 7942. We employed traditional molecular biology and -omic techniques, and developed and applied a high-throughput approach for whole genome mutant screens in *S. elongatus* (RB-TnSeq). **Chapter one** introduces what is known about cyanobacterial physiology in Light-Dark Cycles (LDCs). It also describes RB-TnSeq, which is used here to elucidate gene function both generally and specifically to LDCs. **Chapter two** reports the characterization of cellular activities upon light transitions that facilitate survival in LDCs. **Chapter three** presents the development of RB-TnSeq in *S. elongatus* and its use for the assignment of gene importance, the development of an improved metabolic model for *S. elongatus*, and the implementation of screens to further our understanding of these genes. **Chapter four** reports the application of RB-TnSeq to understand survival in LDCs. In it, we identified the set of genes specifically important for survival of LDCs and followed up on prioritized candidates. This work resulted in improved understanding for the roles of the circadian clock and nucleotide signaling in LDC survival. **Chapter five** concludes by synthesizing the core achievements of the dissertation and suggesting future directions. Together, these chapters explain the development of a powerful genomic approach, RB-TnSeq, and its use to illuminate the genetic unknowns in Cyanobacteria as well as the organisms' LDC physiology. These findings will be applicable to the basic understanding of this important phylum, its industrial use, and photosynthetic organisms more generally, for which Cyanobacteria are the most tractable models.

CHAPTER 1: Introduction

1.1 Summary

Cyanobacteria are the most genetically tractable models for photosynthesis, the premier prokaryotic organisms for circadian clock research, responsible for tremendous environmental impact as primary producers, and increasingly popular chassis for industrial product production. Nevertheless, there are two large dark spots in our understanding of the phylum. The first is darkness itself. The pathways necessary to survive day-night cycles in cyanobacteria are currently poorly understood because most research is conducted in constant light, leaving half the lifecycle of this keystone phylum shrouded in shadow. This shortcoming and the work that has been done so far to rectify it is the topic of a review on which I am primary author and which is included, in part, as section **1.2 Cyanobacterial Day-Night Physiology**. The second fundamental knowledge gap is gene annotation. Approximately 40% of genes in cyanobacterial species have no-functional prediction and for those that do, the annotation is largely inferred from distantly related non-photosynthetic organisms. A newly developed approach for whole genome mutant screens using barcoded transposon mutagenesis coupled with high-throughput sequencing (RB-TnSeq) provides a powerful tool for the illumination of both areas. The conceptual basis of this technique is introduced in section **1.3 RB-TnSeq**. In this dissertation, RB-TnSeq and traditional molecular genetic

approaches are used on the model cyanobacterium *Synechococcus elongatus* PCC 7942 for genome-wide functional annotation of genes, as well as targeted enquiries into the physiology of day-night cycles.

1.2 Cyanobacterial Day-Night Physiology

An Introduction to Day-Night Cycles in Cyanobacteria. The daily fluctuation of light is a nearly universal evolutionary challenge to life. For cyanobacteria, microorganisms that rely almost exclusively on light for energy, the response to these day-night cycles is particularly extreme ranging from redirection of central metabolism (Diamond et al., 2017, 2015) to sweeping changes in the cell's transcription (Hosokawa et al., 2011; Ito et al., 2009). Moreover, as progenitors via endosymbiosis to all oxygen-evolving photosynthetic organisms (de Vries and Archibald, 2017), their response to light-dark cycles (LDCs) has broad implications for understanding photosynthesis, and for characterizing a phylum that has tremendous ecological impact and biotechnological potential (Flombaum et al., 2013; Oliver et al., 2016). However, due to the practical advantages of experimentation in continuous light, most research, and all reviews on cyanobacteria to date, have focused on these unnaturally static conditions. As a result, half of the lifecycle of this keystone phylum is largely unexplored.

New research has begun to address this shortcoming by probing the physiology of cyanobacteria in LDCs. Here, we consolidate current knowledge on the cyanobacterial response to LDCs by discussing functions of importance for the day state, the night state, and the regulation that is required for the drastic transitions between the two. We will also discuss how our paradigm for the response to LDCs generalizes beyond cyanobacteria.

Surviving the Day. Each day, a cyanobacterium wakes to the formidable task of turning inorganic carbon into the organic molecules of life via photosynthetic carbon dioxide assimilation. Its metabolic challenges are numerous. The cell must simultaneously duplicate its molecules to prepare for division while also storing energy reserves for the night. These reactions take place in the background of photosynthesis, which requires significant cellular resources for efficient function and generates damaging reactive oxygen species (ROS) as a secondary byproduct (Latifi et al., 2009). As a consequence, cyanobacterial metabolism is carefully orchestrated in both space and time (Ito et al., 2009; Stanier and Cohen-Bazire, 1977; Vijayan et al., 2009).

Central Carbon Metabolism. The temporal organization of cyanobacterial metabolism can be generalized as anabolic during the day and catabolic at night. Daytime metabolism begins with shifting flux from the catabolic oxidative pentose phosphate pathway (OPPP) and initiating the anabolic Calvin-Benson-Bassham Cycle (CBBC) via the activity of the photosynthetic light reactions (Knoop et al., 2013; You et al., 2015; Young et al., 2011). One of the critical steps in this process is inactivation of CP12, a redox-sensitive protein that is a master regulator of the CBBC (Tanoi et al., 2005). During the night, oxidized CP12 structurally sequesters glyceraldehyde 3-phosphate dehydrogenase (GAPDH) and phosphoribulokinase (PRK) and inhibits the CBBC. At the onset of light, photosynthetic reducing equivalents are generated, CP12 is inhibited, and CBBC activity resumes. Metabolomic analysis supports upregulation of anabolic metabolism during this phase of the day including pathways related to amino

acid, nucleotide, and quinone biosynthesis (Diamond et al., 2015). Upregulation of amino acid and nucleotide production agrees with historical observations that protein translation and DNA replication largely occur only during the day (Mori et al., 1996; Singer and Doolittle, 1975).

Energy Storage and Electron Sinks. A principal activity during the daytime is accumulation of excess photosynthate, which is sequestered as the glucose polymer glycogen. Under LDC growth, glycogen daytime accumulation serves two primary purposes: it is synthesized as the primary storage polymer in preparation for night (Diamond et al., 2015; Gaudana et al., 2013; Hanai et al., 2014; Suzuki et al., 2007; Wyman and Thom, 2012), and it serves as a “regulatory valve” for excess reducing power produced under conditions of particularly high light intensity (Latifi et al., 2009; Li et al., 2014; Miao et al., 2003; Work et al., 2015). The importance of glycogen storage is highlighted by the fact that mutations targeting the its biosynthetic pathway genes *glgA*, *glgC*, and *glgP* significantly attenuate the ability of cells to grow in LDC (Gründel et al., 2012; Osanai et al., 2007).

Surviving the Night.

Transcription and Translation. It is perhaps not surprising that darkness triggers widespread changes in transcription and translation in the cyanobacteria *Synechococcus elongatus*. Rates of both processes generally decrease (Doolittle, 1979; Hosokawa et al., 2011; Takano et al., 2015), but specific genes are induced upon a transition to darkness. Many of these genes are of unknown

function, and some are annotated simply as *dig*, for dark-induced gene (Hosokawa et al., 2011). Better-annotated examples within this list include a protein with a high degree of sequence similarity to CP12, (Tamoi et al., 2005); *hpf (IrtA)*, a regulator of ribosomal status (Hood et al., 2016); and heat shock protein A (*hspA*), a predicted chaperone. Learning more about the identities and functions of these genes may provide us with clues about how *S. elongatus* survives periods of darkness.

Are transcript decreases in darkness mirrored by protein level? The answer to this question is less clear. Ansong et al. (Ansong et al., 2014) showed that only 4% of proteins change in abundance between light and dark, suggesting that overall protein levels remain relatively constant. However, several studies have shown that translation rates decrease in the dark (Hood et al., 2016; Singer and Doolittle, 1975), and that dark-synthesized proteins differ from those synthesized in the light (Suranyi et al., 1987). Determining the identities of these dark-synthesized proteins using a technique such as ribosomal profiling represents an important next step in understanding translational responses to darkness.

Though a cyanobacterium in darkness is typically viewed as being in a dormant state, the cell is not inactive – many processes still operate dynamically. Studies on transcription, translation, and metabolism have demonstrated a specific, adaptive response to darkness in *S. elongatus*. While overall rates of these processes may be lower or close to zero, as is the case with DNA replication (Ohbayashi et al., 2013; Watanabe et al., 2012), they are coordinated

in such a way that the cell can conserve energy, ensuring its survival until the light shines again.

Regulation.

Transcription factors. RpaA and RpaB are transcription factors in cyanobacteria that act as control hubs of LDC physiology. RpaA has been studied thoroughly as the output mechanism for the cyanobacterial circadian clock and controls oscillation of thousands of genes in a clock-dependent manner (Shultzaberger et al., 2015). While RpaA serves to connect the circadian clock and transcription, RpaB, an essential response regulator of the histidine kinase NblS, integrates environmental signals. RpaB has been studied extensively as a light-responsive regulator of gene expression (Seino et al., 2009; Wilde and Hihara, 2016). Mutations that cause changes in RpaB level and phosphorylation have fitness phenotypes in LDCs (Espinosa et al., 2015). RpaB's activity in LDCs appears to be independent of the circadian clock; however, its output overlaps with that of the clock through antagonism of RpaA activity. These studies suggest that whereas RpaA acts as the output of the clock, RpaB integrates signals of LDCs into the transcriptional regime of the cell in a clock-independent way, and that both signals are necessary for survival of LDCs.

Chromosome Topology. One of the best-described examples of spatial organization changes in cyanobacteria during LDCs is that of chromosome compaction. Several studies have shown that the extent of chromosome or plasmid compaction in *S. elongatus* changes depending on the time of day (Murata et al., 2016; Smith and Williams, 2006; Vijayan et al., 2009; Woelfle et al., 2007). This phenomenon persists in constant light, but not constant darkness, and is thought to be circadian-regulated.

Furthermore, when supercoiling is altered using an inhibitor of DNA gyrase, gene expression patterns change (Min et al., 2004; Vijayan et al., 2009).

These studies provide insight into how gene expression might be controlled during circadian or diel cycles, but they are mostly correlative, relating DNA compaction or supercoiling with transcriptional outputs. The mechanistic details of how chromatin state and nucleoid organization are controlled – by the circadian oscillator, by histone-like DNA-binding proteins, and/or by changes in DNA topology – remain largely unknown.

Signaling nucleotides. Evidence is accumulating that signaling nucleotides act as intracellular messengers of LDCs. Levels of cAMP, c-di-AMP, c-di-GMP, and ppGpp have all been found to be light dependent in cyanobacteria (Agostoni and Montgomery, 2014) (See Section 4.3). Synthesis of ppGpp in *S. elongatus* after a light-to-dark transition leads to transcriptional responses, which in at least one case affect the physiological status of the cell (Hood et al., 2016). Survival of a strain that cannot make ppGpp is impaired after exposure to darkness, though the mechanisms behind this phenotype are not yet known.

C-di-AMP, a newly discovered signaling nucleotide in cyanobacteria, is also important for *S. elongatus* survival of darkness. Inactivation of its cyclase, *dacA*, leads to increased oxidative stress and decreased survival of the night periods of LDCs (See Section 4.3). C-di-AMP and ppGpp levels are linked in Firmicutes (Corrigan et al., 2015; Whiteley et al., 2015), suggesting the possibility that their activity in LDC cycles is coordinated, but evidence does not yet exist in cyanobacteria. Although there are many unknowns in the roles of nucleotide signaling in Cyanobacteria, they appear to have a role in regulation of LDC physiology.

Beyond Cyanobacteria. The response to changing light conditions induced by diel cycles is also a dominating force for plants and eukaryotic algae and influences global geochemical cycles such as CO₂ fixation and climate change. In plants, constant adjustments to light quality, intensity, and duration are made through the use of photoreceptors. Varying LDCs cue plants to undergo different phases of growth, development, and metabolism (Seluzicki et al., 2017). In *Arabidopsis*, for example, darkness elicits the expression of over 80 proteins that code for functions involved in photosystem II inhibition, starch degradation, chloroplastic translation inhibition, and redox regulation, all common themes that are familiar to what we observe in cyanobacteria (Wang et al., 2016).

There are also relevant responses plants make to darkness and LDCs from an economic/agricultural point of view. It has been shown that post-harvest storage and exposure of green leafy vegetables, such as kale and cabbage, compared to storage in non-LDC conditions, resulted in significantly improved appearance and health value of crops due to increased tissue integrity, chlorophyll content, and levels of glucosinolates, a phytonutrient (Liu et al., 2015). These consequences of crop storage in LDCs were comparable to that of refrigeration and reiterate the importance of understanding physiological responses to LDCs. As the most tractable photosynthetic model organisms, cyanobacteria can inform the industrial and applied sciences. For further discussion of how darkness, carbon storage, and plant productivity are interrelated, we direct readers to a review by Graf and Smith (Graf and Smith, 2011).

Some of the mechanisms developed to deal with the stresses inherent to LDCs in photosynthetic organisms may be conserved far beyond them. Cyanobacteria are not alone in struggling with oxidative stress. Hovering flight of nectarivores is an immensely energetic endeavor and comes with high metabolic turnover and ROS. In hawkmoths,

like cyanobacteria, this oxidative stress is likely detoxified by activity of the OPPP during rest, to produce the antioxidants NADPH and reduced glutathione (Levin et al., 2017). Furthermore, this strategy for oxidative stress management after intense exercise may be of importance in animals far beyond hawkmoths (Del Rio and Dillon, 2017). Therefore, strategies evolved in cyanobacteria to support LDC survival are present at distant ends of the tree of life for unique environmental concerns.

References.

- Agostoni, M., Montgomery, B., 2014. Survival Strategies in the Aquatic and Terrestrial World: The Impact of Second Messengers on Cyanobacterial Processes. *Life* 4, 745–769.
- Ansong, C., Sadler, N.C., Hill, E.A., Lewis, M.P., Zink, E.M., Smith, R.D., Beliaev, A.S., Konopka, A.E., Wright, A.T., 2014. Characterization of protein redox dynamics induced during light-to-dark transitions and nutrient limitation in cyanobacteria. *Front. Microbiol.* 5, 325.
- Corrigan, R.M., Bowman, L., Willis, A.R., Kaeffer, V., Gründling, A., 2015. Cross-talk between two nucleotide-signaling pathways in *Staphylococcus aureus*. *J. Biol. Chem.* 290, 5826–5839.
- Del Rio, C.M., Dillon, M.E., 2017. Sweet relief for pollinators. *Science* 355, 686–687.
- de Vries, J., Archibald, J.M., 2017. Endosymbiosis: Did Plastids Evolve from a Freshwater Cyanobacterium? *Curr. Biol.* 27, R103–R105.
- Diamond, S., Jun, D., Rubin, B.E., Golden, S.S., 2015. The circadian oscillator in *Synechococcus elongatus* controls metabolite partitioning during diurnal growth. *Proc. Natl. Acad. Sci. U. S. A.* 112, E1916–25.
- Diamond, S., Rubin, B.E., Shultzaberger, R.K., Chen, Y., Barber, C.D., Golden, S.S., 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.
- Doolittle, W.F., 1979. The cyanobacterial genome, its expression, and the control of that expression. *Adv. Microb. Physiol.* 20, 1–102.

- Espinosa, J., Boyd, J.S., Cantos, R., Salinas, P., Golden, S.S., Contreras, A., 2015. Cross-talk and regulatory interactions between the essential response regulator RpaB and cyanobacterial circadian clock output. *Proc. Natl. Acad. Sci. U. S. A.* 112, 2198–2203.
- Flombaum, P., Gallegos, J.L., Gordillo, R.A., Rincon, J., Zabala, L.L., Jiao, N., Karl, D.M., Li, W.K., Lomas, M.W., Veneziano, D., Vera, C.S., Vrugt, J.A., Martiny, A.C., 2013. Present and future global distributions of the marine Cyanobacteria *Prochlorococcus* and *Synechococcus*. *Proc. Natl. Acad. Sci. U. S. A.* 110, 9824–9829.
- Gaudana, S.B., Alagesan, S., Chetty, M., Wangikar, P.P., 2013. Diurnal rhythm of a unicellular diazotrophic cyanobacterium under mixotrophic conditions and elevated carbon dioxide. *Photosynth. Res.* 118, 51–57.
- Graf, A., Smith, A.M., 2011. Starch and the clock: the dark side of plant productivity. *Trends Plant Sci.* 16, 169–175.
- Gründel, M., Scheunemann, R., Lockau, W., Zilliges, Y., 2012. Impaired glycogen synthesis causes metabolic overflow reactions and affects stress responses in the cyanobacterium *Synechocystis* sp. PCC 6803. *Microbiology* 158, 3032–3043.
- Hanai, M., Sato, Y., Miyagi, A., Kawai-Yamada, M., Tanaka, K., Kaneko, Y., Nishiyama, Y., Hihara, Y., 2014. The Effects of Dark Incubation on Cellular Metabolism of the Wild Type Cyanobacterium *Synechocystis* sp. PCC 6803 and a Mutant Lacking the Transcriptional Regulator *cyAbrB2*. *Life* 4, 770–787.
- Hood, R.D., Higgins, S.A., Flamholz, A., Nichols, R.J., Savage, D.F., 2016. The stringent response regulates adaptation to darkness in the cyanobacterium *Synechococcus elongatus*. *Proc. Natl. Acad. Sci. U. S. A.* 113, E4867–76.
- Hosokawa, N., Hatakeyama, T.S., Kojima, T., Kikuchi, Y., Ito, H., Iwasaki, H., 2011. Circadian transcriptional regulation by the posttranslational oscillator without de novo clock gene expression in *Synechococcus*. *Proc. Natl. Acad. Sci. U. S. A.* 108, 15396–15401.
- Ito, H., Mutsuda, M., Murayama, Y., Tomita, J., Hosokawa, N., Terauchi, K., Sugita, C., Sugita, M., Kondo, T., Iwasaki, H., 2009. Cyanobacterial daily life with Kai-based circadian and diurnal genome-wide transcriptional control in *Synechococcus elongatus*. *Proc. Natl. Acad. Sci. U. S. A.* 106, 14168–14173.
- Knoop, H., Gründel, M., Zilliges, Y., Lehmann, R., Hoffmann, S., Lockau, W., Steuer, R., 2013. Flux balance analysis of cyanobacterial metabolism: the metabolic network of *Synechocystis* sp. PCC 6803. *PLoS Comput. Biol.* 9, e1003081.

- Latifi, A., Ruiz, M., Zhang, C.-C., 2009. Oxidative stress in cyanobacteria. *FEMS Microbiol. Rev.* 33, 258–278.
- Levin, E., Lopez-Martinez, G., Fane, B., Davidowitz, G., 2017. Hawkmoths use nectar sugar to reduce oxidative damage from flight. *Science* 355, 733–734.
- Liu, J.D., Goodspeed, D., Sheng, Z., Li, B., Yang, Y., Kliebenstein, D.J., Braam, J., 2015. Keeping the rhythm: light/dark cycles during postharvest storage preserve the tissue integrity and nutritional content of leafy plants. *BMC Plant Biol.* 15, 92.
- Li, X., Shen, C.R., Liao, J.C., 2014. Isobutanol production as an alternative metabolic sink to rescue the growth deficiency of the glycogen mutant of *Synechococcus elongatus* PCC 7942. *Photosynth. Res.* 120, 301–310.
- Miao, X., Wu, Q., Wu, G., Zhao, N., 2003. Changes in photosynthesis and pigmentation in an *agp* deletion mutant of the cyanobacterium *Synechocystis* sp. *Biotechnol. Lett.* 25, 391–396.
- Min, H., Liu, Y., Johnson, C.H., Golden, S.S., 2004. Phase determination of circadian gene expression in *Synechococcus elongatus* PCC 7942. *J. Biol. Rhythms* 19, 103–112.
- Mori, T., Binder, B., Johnson, C.H., 1996. Circadian gating of cell division in cyanobacteria growing with average doubling times of less than 24 hours. *Proc. Natl. Acad. Sci. U. S. A.* 93, 10183–10188.
- Murata, K., Hagiwara, S., Kimori, Y., Kaneko, Y., 2016. Ultrastructure of compacted DNA in cyanobacteria by high-voltage cryo-electron tomography. *Sci. Rep.* 6, 34934.
- Ohbayashi, R., Watanabe, S., Kanesaki, Y., Narikawa, R., Chibazakura, T., Ikeuchi, M., Yoshikawa, H., 2013. DNA replication depends on photosynthetic electron transport in cyanobacteria. *FEMS Microbiol. Lett.*
- Oliver, N.J., Rabinovitch-Deere, C.A., Carroll, A.L., Nozzi, N.E., Case, A.E., Atsumi, S., 2016. Cyanobacterial metabolic engineering for biofuel and chemical production. *Curr. Opin. Chem. Biol.* 35, 43–50.
- Osanai, T., Azuma, M., Tanaka, K., 2007. Sugar catabolism regulated by light- and nitrogen-status in the cyanobacterium *Synechocystis* sp. PCC 6803. *Photochem. Photobiol. Sci.* 6, 508–514.
- Seino, Y., Takahashi, T., Hihara, Y., 2009. The Response Regulator RpaB Binds to the Upstream Element of Photosystem I Genes To Work for Positive Regulation under Low-Light Conditions in *Synechocystis* sp. Strain PCC 6803. *J. Bacteriol.* 191, 1581–1586.

- Seluzicki, A., Burko, Y., Chory, J., 2017. Dancing in the dark: darkness as a signal in plants. *Plant Cell Environ.*
- Shultzaberger, R.K., Boyd, J.S., Diamond, S., Greenspan, R.J., Golden, S.S., 2015. Giving Time Purpose: The *Synechococcus elongatus* Clock in a Broader Network Context. *Annu. Rev. Genet.* 49, 485–505.
- Singer, R.A., Doolittle, W.F., 1975. Control of gene expression in blue-green algae. *Nature* 253, 650–651.
- Smith, R.M., Williams, S.B., 2006. Circadian rhythms in gene transcription imparted by chromosome compaction in the cyanobacterium *Synechococcus elongatus*. *Proc. Natl. Acad. Sci. U. S. A.* 103, 8564–8569.
- Stanier, R.Y., Cohen-Bazire, G., 1977. Phototrophic prokaryotes: the cyanobacteria. *Annu. Rev. Microbiol.* 31, 225–274.
- Suranyi, G., Korcz, A., Palfi, Z., Borbely, G., 1987. Effects of light deprivation on RNA synthesis, accumulation of guanosine 3'(2')-diphosphate 5'-diphosphate, and protein synthesis in heat-shocked *Synechococcus* sp. strain PCC 6301, a cyanobacterium. *J. Bacteriol.* 169, 632–639.
- Suzuki, E., Umeda, K., Nihei, S., Moriya, K., Ohkawa, H., Fujiwara, S., Tsuzuki, M., Nakamura, Y., 2007. Role of the GlgX protein in glycogen metabolism of the cyanobacterium, *Synechococcus elongatus* PCC 7942. *Biochim. Biophys. Acta* 1770, 763–773.
- Takano, S., Tomita, J., Sonoike, K., Iwasaki, H., 2015. The initiation of nocturnal dormancy in *Synechococcus* as an active process. *BMC Biol.* 13, 36.
- Tamoi, M., Miyazaki, T., Fukamizo, T., Shigeoka, S., 2005. The Calvin cycle in cyanobacteria is regulated by CP12 via the NAD(H)/NADP(H) ratio under light/dark conditions. *Plant J.* 42, 504–513.
- Vijayan, V., Zuzow, R., O'Shea, E.K., 2009. Oscillations in supercoiling drive circadian gene expression in cyanobacteria. *Proceedings of the National Academy of Sciences* 106, 22564–22568.
- Wang, J., Yu, Q., Xiong, H., Wang, J., Chen, S., Yang, Z., Dai, S., 2016. Proteomic Insight into the Response of Arabidopsis Chloroplasts to Darkness. *PLoS One* 11, e0154235.
- 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.

- Whiteley, A.T., Pollock, A.J., Portnoy, D.A., 2015. The PAMP c-di-AMP Is Essential for *Listeria monocytogenes* Growth in Rich but Not Minimal Media due to a Toxic Increase in (p)ppGpp. *Cell Host Microbe* 17, 788–798.
- Wilde, A., Hihara, Y., 2016. Transcriptional and posttranscriptional regulation of cyanobacterial photosynthesis. *Biochim. Biophys. Acta* 1857, 296–308.
- Woelfle, M.A., Xu, Y., Qin, X., Johnson, C.H., 2007. Circadian rhythms of superhelical status of DNA in cyanobacteria. *Proc. Natl. Acad. Sci. U. S. A.* 104, 18819–18824.
- Work, V.H., Melnicki, M.R., Hill, E.A., Davies, F.K., Kucek, L.A., Beliaev, A.S., Posewitz, M.C., 2015. Lauric Acid Production in a Glycogen-Less Strain of *Synechococcus* sp. PCC 7002. *Front Bioeng Biotechnol* 3, 48.
- Wyman, M., Thom, C., 2012. Temporal orchestration of glycogen synthase (GlgA) gene expression and glycogen accumulation in the oceanic picoplanktonic cyanobacterium *Synechococcus* sp. strain WH8103. *Appl. Environ. Microbiol.* 78, 4744–4747.
- You, L., He, L., Tang, Y.J., 2015. Photoheterotrophic fluxome in *Synechocystis* sp. strain PCC 6803 and its implications for cyanobacterial bioenergetics. *J. Bacteriol.* 197, 943–950.
- Young, J.D., Shastri, A.A., Stephanopoulos, G., Morgan, J.A., 2011. Mapping photoautotrophic metabolism with isotopically nonstationary (¹³C) flux analysis. *Metab. Eng.* 13, 656–665.

1.3 RB-TnSeq

RB-Tn-seq is a method that exploits transposon mutagenesis coupled to high throughput sequencing to elucidate gene function. In RB-TnSeq a dense mutant library is created with transposons, each of which contains a unique barcode. The library is grown up and genomic DNA is sequenced to connect each barcoded transposon to the surrounding sequence, so the barcode serves as an identity tag for each mutant in the *S. elongatus* library (Fig. 1-1). In addition to linking barcode to surrounding sequence, library analysis can be used to provide genomic essentiality information by looking for genes and intergenic regions where transposons are underrepresented because the mutants are not viable. In this way initial characterization of the library allows unique barcodes be used as identifiers for each mutant, and provides genome wide essentiality data.

After initial linking of transposon barcodes to surrounding sequence the library can be used for pooled screening for fitness contribution of individual loci under specific growth conditions. In this approach the library is split into one control condition and into any number of experimental conditions. After some prescribed period of growth the culture grown in each condition is screened by high throughput sequencing of the barcodes. The count of each barcode is inversely correlated with the importance of the associated gene in the condition sampled (Fig. 1.3-1). For example, in a screen conducted in high temperatures, an underrepresented barcode indicates that the gene linked to the barcode is important for survival in high heat because few mutants for that gene survived the condition. By using pooled barcoded mutants, this approach plays into the

strengths of next generation sequencing, resulting in an unbiased, inexpensive, and quantitative whole-genome screen.

Another use of the library is arrayed screening for novel mutants. In this approach the mutant library would be plated on agar in Petri dishes to separate out single cloned mutants. The mutation responsible for any phenotype of interest could be determined by PCR using universal primers surrounding the barcode and then Sanger sequencing. The sequence of the barcode would identify the mutation location, which was determined during library analysis.

Figures.

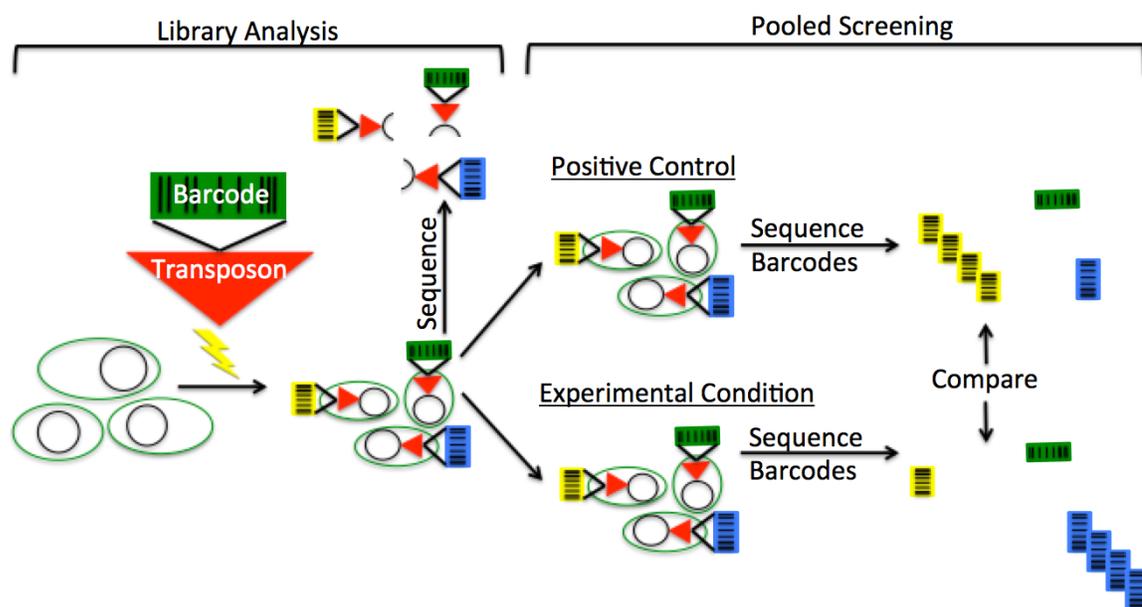


Figure 1.3-1. RB-TnSeq Approach. In library analysis the location of each barcoded transposon in the genome is found using high-throughput sequencing. In pooled screening the survival of every mutant in the library is tracked in a control and experimental conditions using high-throughput sequencing of the barcodes. The representation of the mutants, as determined by the number of barcodes specific to it, can be used to determine the survival of that mutant in the experimental condition.

1.4 Acknowledgments

Chapter 1, in part, is made up of material from a manuscript in preparation for publication and an invention disclosure. The manuscript in preparation is: Rubin BE, Welkie DG, Diamond S, Hood RD, Savage DF, and Golden SS. The dark side of the cyanobacterium. (*In preparation*). The dissertation author is the primary author of this manuscript in preparation. The disclosure is: Rubin BE, Wetmore KM, Price MN, Arkin AP, Deutschbauer A, and Golden SS (2015) Saturated Cyanobacterial Mutant Library Optimized for Screening. Disclosure Case Number: 2016:013. The dissertation author was the primary author of this disclosure.

CHAPTER 2: Physiology Under Light-Dark Transitions

2.1 Chapter Summary

Although it has long been known that the circadian clock provides a fitness advantage under LDCs, nothing was previously known about how the output of this simple yet robust time-keeping mechanism afforded such an advantage. Through observation of the metabolic changes of *S. elongatus* upon transitions into light and into darkness, many of the important roles of the circadian clock in LDC survival have come into focus. Upon transition into light, the active state of the clock, present at dawn, draws flux away from central metabolism and towards the production of secondary metabolites that likely have small but significant fitness roles. This transition is explained in the form of a *PNAS* paper on which I was third author, in section **2.2 Physiology of the Dark to Light Transition**. Upon transition into darkness, the inactive state of the circadian clock, present at dusk, allows for a remarkable metabolic stability to be maintained despite the perturbation of darkness. This active maintenance of metabolism is likely necessary to detoxify oxidative stress that can become lethal at night if not subdued. Mutants in which the circadian clock is constitutively active and that are unable to achieve the normal night-time state have catastrophic phenotypes in LDCs. These findings are elaborated in a *PNAS* paper on which I was second author, in section **2.2 Physiology of the Dark to Light Transition**. Together, the beneficial roles of each state of the circadian clock shed light on the necessity of its oscillation between a dawn and dusk state.

2.2 Physiology of the Dark to Light Transition

PNAS

The circadian oscillator in *Synechococcus elongatus* controls metabolite partitioning during diurnal growth

Spencer Diamond, Darae Jun¹, Benjamin E. Rubin, and Susan S. Golden²

Center for Circadian Biology, California Center for Algal Biotechnology, Division of Biological Sciences, University of California, San Diego, La Jolla, CA 92093

Contributed by Susan S. Golden, March 6, 2015 (sent for review February 6, 2015; reviewed by Shota Atsumi and Dan Ducat)

Synechococcus elongatus PCC 7942 is a genetically tractable model cyanobacterium that has been engineered to produce industrially relevant biomolecules and is the best-studied model for a prokaryotic circadian clock. However, the organism is commonly grown in continuous light in the laboratory, and data on metabolic processes under diurnal conditions are lacking. Moreover, the influence of the circadian clock on diurnal metabolism has been investigated only briefly. Here, we demonstrate that the circadian oscillator influences rhythms of metabolism during diurnal growth, even though light-dark cycles can drive metabolic rhythms independently. Moreover, the phenotype associated with loss of the core oscillator protein, KaiC, is distinct from that caused by absence of the circadian output transcriptional regulator, RpaA (regulator of phycobilisome-associated A). Although RpaA activity is important for carbon degradation at night, KaiC is dispensable for those processes. Untargeted metabolomics analysis and glycogen kinetics suggest that functional KaiC is important for metabolite partitioning in the morning. Additionally, output from the oscillator functions to inhibit RpaA activity in the morning, and *kaiC*-null strains expressing a mutant KaiC phosphomimetic, KaiC-pST, in which the oscillator is locked in the most active output state, phenocopies a $\Delta rpaA$ strain. Inhibition of RpaA by the oscillator in the morning suppresses metabolic processes that normally are active at night, and *kaiC*-null strains show indications of oxidative pentose phosphate pathway activation as well as increased abundance of primary metabolites. Inhibitory clock output may serve to allow secondary metabolite biosynthesis in the morning, and some metabolites resulting from these processes may feed back to reinforce clock timing.

metabolomics | metabolism | circadian clock | cyanobacteria | diurnal

Cyanobacteria comprise a promising engineering platform for the production of fuels and industrial chemicals. These organisms already have been engineered to produce ethanol, isobutyraldehyde, alkanes, and hydrogen (1–4). However, the efficient industrial-scale application of these photosynthetic organisms will require their growth and maintenance in the outdoors where they will be subjected to light–dark (LD) cycles (5). Phototrophic cyanobacteria present a completely different engineering challenge relative to heterotrophic bacteria such as *Escherichia coli*: their cellular activities respond strongly to the presence and absence of light because their metabolism is centered on photosynthesis (6, 7). Diverse cyanobacteria also possess a true circadian clock that synchronizes with external LD cycles and has been demonstrated to drive both gene expression and metabolic rhythms (8–10). It is important to understand how signals from the external environment and the internal circadian clock are integrated to modulate metabolic processes in environmentally relevant LD cycles to optimize the engineering of these organisms. In this work we attempt to separate the influences of environment and circadian control using the cyanobacterium *Synechococcus elongatus* PCC7942, because it is both a highly tractable genetic system and the foundational model for the prokaryotic circadian clock.

The circadian clock in *S. elongatus* is based on a central oscillator formed by the proteins KaiA, KaiB, and KaiC (11). The reversible phosphorylation of KaiC over a 24-h period sets the

timing of the clock mechanism. The clock synchronizes to the environment through KaiA and a histidine protein kinase, CikA. Both proteins bind quinone cofactors, likely plastoquinone present in the photosynthetic membrane, that reflect the cellular redox state (12, 13). KaiC activity also is modulated by the cellular ATP/ADP ratio (14), and both the cellular redox state and ATP/ADP ratio are dependent on the availability of external light. Thus, it has been demonstrated that changes in energy metabolism feed back in setting the timing of clock oscillations (15). The output of the clock is relayed to gene expression through the *Synechococcus* adaptive sensor (SasA)–regulator of phycobilisome-associated A (RpaA) two-component system (16) in which RpaA is a transcription factor that binds more than 170 gene targets. Many of the genes strongly activated by RpaA function in nighttime metabolic processes, including glycogen degradation, glycolysis, and the oxidative pentose phosphate pathway (OPPP) (17).

Under constant-light (LL) growth conditions circadian control in *S. elongatus* is quite pervasive, with up to 64% of transcripts displaying 24-h clock-dependent oscillations (10). Gene expression has roughly two distinct phases in LL: genes with an expression peak at subjective dusk (class 1) and genes with an expression peak at subjective dawn (class 2) (18). Recent work by Paddock et al. (19) suggests that a single output from the central oscillator is responsible for both out-of-phase rhythms and that the oscillator has maximum output activity in the morning when KaiC-pST becomes the most prevalent phosphorylation state. Furthermore, there is evidence that oscillator activity is inhibitory (20), and rhythms may manifest as different responses to

Significance

Cyanobacteria are increasingly being considered for use in large-scale outdoor production of fuels and industrial chemicals. Cyanobacteria can anticipate daily changes in light availability using an internal circadian clock and rapidly alter their metabolic processes in response to changes light availability. Understanding how signals from the internal circadian clock and external light availability are integrated to control metabolic shifts will be important for engineering cyanobacteria for production in natural outdoor environments. This study has assessed how “knowing” the correct time of day, via the circadian clock, affects metabolic changes when a cyanobacterium goes through a dark-to-light transition. Our data show that the circadian clock plays an important role in inhibiting activation of the oxidative pentose phosphate pathway in the morning.

Author contributions: S.D., B.E.R., and S.S.G. designed research; S.D., D.J., and B.E.R. performed research; S.D. and S.S.G. analyzed data; and S.D. and S.S.G. wrote the paper. Reviewers: S.A., University of California, Davis; and D.D., Michigan State University.

The authors declare no conflict of interest.

¹Present address: Department of Plant and Microbial Biology, University of California, Berkeley, CA 94720.

²To whom correspondence should be addressed. Email: sgolden@ucsd.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1504576112/-DCSupplemental.



the alleviation and return of this inhibition over a daily period. It also is likely that metabolism is strongly influenced by the clock in constant light, because a statistically high proportion of genes involved in energy metabolism are rhythmic in LL (21). However, no metabolic pathways are specifically enriched in class 1 or class 2 genes with the exception of ribosome biogenesis and photosynthesis, respectively (10).

A few studies have investigated the transcriptome, proteome, and physiological dynamics of particular species of cyanobacteria over a 24-h period under LD growth (6, 22, 23). In general, systems for oxygenic photosynthesis are activated during the day, and systems for respiratory metabolism are activated at night. Additionally, the day and night periods are used by cyanobacteria to segregate incompatible metabolic processes (22). For example, *S. elongatus* activates light-independent protochlorophyllide reduction, which is an oxygen-sensitive process, at night, a time when oxygen is not being produced by photosystem II (24). However, the degree to which the circadian clock and light availability independently affect metabolic events is poorly understood. In *S. elongatus*, only two studies have investigated the behavior of mutants that lack a functional clock under an LD cycle (21, 25). The available studies investigate these effects only over a light-to-dark transition, so currently there is an incomplete understanding of the circadian influence on cellular events over a full 24-h LD cycle. Finally, although there is a proteomics dataset for *S. elongatus* that covers a full 24-h LD period, that study tracked only WT cells and does not decouple clock and environmental influences (23).

When cyanobacteria are grown in a 24-h LD cycle, cells perform photosynthesis and store fixed carbon as the branched glucose polymer glycogen during the day. Glycogen subsequently is degraded at night for energy and reducing power via the OPPP (26, 27). Pattanayak et al. (15) recently showed that glycogen in *S. elongatus* oscillates in LL and that this oscillation depends on a functional clock. Rhythms of glycogen accumulation and degradation also have been observed during LD growth in *S. elongatus* (28); however, the influence of the clock under LD conditions is not clear. In fact, enzymes in glycogen metabolism are sensitive to the cellular redox state, and LD transitions alone may trigger changes in glycogen content (29). Glycogen is essential for survival in LD: Mutants defective for the *glgA* (glycogen synthase) or *glgC* (ADP-glucose pyrophosphorylase) genes, which are required for glycogen synthesis, are not viable under LD growth regimes (30). In turn, the deletion of the OPPP gene *zwf* (glucose-6-phosphate 1-dehydrogenase) or glycolysis gene *gap1* (glyceraldehyde 3-phosphate dehydrogenase), both of which participate in pathways that consume glycogen, results in mutants that are impaired in LD growth (31, 32). Null mutations in the circadian oscillator, including deletions of *kaiA*, *kaiB*, or *kaiC*, do not impair LD growth. However, disruptions in the SasA-RpaA clock output pathway dramatically stifle growth in LD (16, 33), and genes involved in catabolism of carbon, including *glgP* (glycogen phosphorylase), *gap1*, and *zwf*, are all known RpaA targets (17). Thus, although the clock output pathway likely activates important metabolic processes that occur at night, it is not clear if or how the circadian oscillator modulates these processes.

In this study we applied genetic, biochemical, and metabolomic methods to *S. elongatus* to dissect how the circadian oscillator and activation of the clock output pathway specifically control metabolism under an LD growth regime. We tracked glycogen content in WT *S. elongatus* and a Δ *kaiC* mutant over a 72-h time course under both LL and LD conditions. Subsequently, we characterized glycogen kinetics at LD transitions in WT, Δ *kaiC*, Δ *rpaA*, and a Δ *kaiC*::KaiC-pST phosphomimetic mutant (KaiC-ET) to address whether circadian oscillator output exerts a negative or positive control over glycogen levels. Finally, we performed untargeted metabolic profiling of WT cells and Δ *kaiC* mutants to investigate how oscillator activity affects global

metabolite abundance at the transition from darkness into light. We present a hypothesis for clock regulation of diurnal metabolism that combines our data with previous reports on *S. elongatus* transcript and protein rhythms (17, 21, 23) and that highlights the importance of circadian output for proper metabolite partitioning under LD growth regimes.

Results

The Circadian Clock Segregates Anabolic and Catabolic Carbon Metabolism in LL. To determine whether carbon metabolic pathways are under circadian control, we mined existing datasets using a bioinformatic approach that breaks larger pathways into anabolic and catabolic components. Using the Kyoto Encyclopedia of Genes and Genomes (KEGG), we determined which reactions of glycolysis, OPPP, and the Calvin cycle act exclusively within the OPPP (catabolic) or the Calvin cycle (anabolic). We subsequently annotated the enzymes that enable these reactions with their circadian class of transcript [class 1 (peaks at dusk) and class 2 (peaks at dawn)] using available microarray data collected from cells grown in LL (10). Our analysis showed that catabolic reactions are catalyzed exclusively by enzymes with class 1 gene-expression profiles, whereas anabolism is catalyzed almost exclusively by enzymes with class 2 gene-expression profiles (Fig. 1 *A* and *B*).

Like the OPPP and the Calvin cycle, glycogen metabolism shows strong temporal segregation in the expression of anabolic and catabolic pathway genes (gray box in Fig. 1*A*). To gauge circadian influence on cellular flux of carbon, we tracked glycogen content for 72 h in WT and in a clockless Δ *kaiC* mutant grown in a photobioreactor under constant and stringently controlled turbidity, temperature, and light conditions (*Materials and Methods*). A recent report from Pattanayak, et al. (15) demonstrated that WT cells show 24-h glycogen oscillations under LL conditions, whereas Δ *kaiC* mutants lack these oscillations. Our data confirmed a *kaiC*-dependent 24-h rhythm of glycogen oscillation in LL (period = 24.7 ± 0.13 h) (Fig. 2*A*). We propose that oscillations in glycogen content under LL conditions result from clock-controlled oscillations of gene expression that segregate pathways for storage and degradation of carbon temporally.

During LD Growth KaiC Has a Repressive Effect on Glycogen Synthesis and Is Not Required for Glycogen Degradation. The daily oscillations in glycogen abundance that occur when cells are grown in a 24-h LD cycle (28) could be controlled by the circadian oscillator or driven by the environmental cycle. We observed glycogen synthesis and degradation rhythms in both WT cells and a Δ *kaiC* mutant during growth in a 12:12 LD cycle over a 72-h period (Fig. 2*B*). Thus, the environment can drive cycles of glycogen accumulation independently of the clock. However, the kinetics of glycogen accumulation were different in the WT and Δ *kaiC* strains. Kinetic profiling revealed that glycogen accumulation occurs significantly faster in Δ *kaiC* mutants than in WT cells during the 12-h light period, particularly within the first 6 h of light exposure (Fig. 3). More rapid accumulation resulted in glycogen reaching its peak content 4–5 h earlier in the Δ *kaiC* mutants than in WT cells. The Δ *kaiC* mutant had different rates of glycogen accumulation in the first and last 6-h blocks of the light period, whereas accumulation in WT cells was maintained at a steady rate over the full 12-h period (Fig. 3*B*). Also, Δ *kaiC* mutants had higher overall glycogen levels than WT cells through the time course (Fig. S1). Thus, the observed rapid accumulation kinetics is not the result of normalization to a smaller starting pool but occurs despite elevated glycogen content in these cells.

In contrast, kinetic profiling of glycogen degradation when cultures were transferred to darkness showed little difference between WT and Δ *kaiC* strains (Fig. 4*A* and Fig. S2). In all tested cases glycogen degradation could be modeled as a first-order

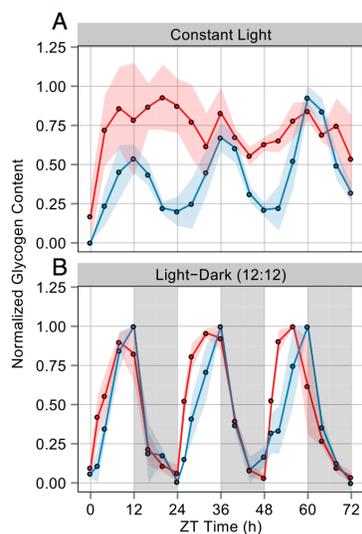


Fig. 2. Average of normalized glycogen content in WT (blue) and $\Delta kaiC$ (red) strains of *S. elongatus* over a 72-h period under both LL and LD growth conditions. The area of shaded color around the solid lines represents SEM. ZTO represents subjective dawn after circadian entrainment (*Materials and Methods*). (A) Glycogen sampling every 4 h from cells grown in LL for 72 h. The WT strain shows a 24-h rhythm of glycogen content, whereas $\Delta kaiC$ has arrhythmic fluctuations. Glycogen was normalized for each biological replicate to the maximum value in that replicate's 72-h period; the solid line is the average of these values. The experiment was performed in triplicate for each strain. (B) Glycogen sampling every 4 h from cells grown in alternating periods of 12 h light and 12 h darkness; darkness is indicated by the gray bars. Both WT and $\Delta kaiC$ strains display a 24-h rhythm of glycogen content. Glycogen was normalized for each biological replicate to the maximum value in that replicate's 24 h period; the solid line is the average of these values. The experiment was performed in duplicate for WT cells and in triplicate for $\Delta kaiC$.

is locked in the most active output state of the clock and phenocopies an RpaA-null strain. This finding agrees with previous reports that overexpression of KaiC has a repressive effect on

expression of class 1 genes, which normally are activated by RpaA (17, 20). Finally, this result demonstrates that RpaA has a positive effect on carbon catabolism; moreover, the ability to grow in a diel cycle strongly correlates with the extent to which glycogen is metabolized in the dark.

Metabolomic Profiling During Dark-to-Light Transition Reveals That the Clock Is Important for Proper Metabolite Partitioning in the Morning. Because disruption of *kaiC* does not cause major changes in glycogen degradation (Fig. 4A), the difference in glycogen accumulation observed between WT and $\Delta kaiC$ strains (Fig. 3A) suggests that a functioning circadian oscillator may be important for metabolite partitioning in the morning. To gain a clearer understanding of early-day metabolic changes in an LD cycle, we performed untargeted metabolic profiling using gas chromatography-time of flight-mass spectrometry (GC-TOF-MS) on both entrained WT and $\Delta kaiC$ strains directly before (0 h) and 4 h after a dark-to-light transition. The analysis successfully identified 130 known metabolites across a broad array of metabolic pathways and an additional 195 unknown metabolites that correspond to previously observed mass spectra to which no purified standard compound has been matched (*Dataset S1*) (34).

Factors contributing to metabolite variability. Because both sampling time and genotype potentially contribute to differences between samples, we first used partial least squares discriminant analysis (PLS-DA) to determine which factors contribute most of the variability in the dataset (35). Plotting PLS-DA components 1 and 2 showed that the sample replicates are well segregated from each other and that the variability from genotype differences is captured by component 1, whereas the variability from sampling time is captured by component 2 (Fig. 5A). Given the association of time and genotype with the respective components, it is apparent that genotype explains a much larger percentage of dataset variability than response to an environmental signal (41.2 and 14.3%, respectively). Also, samples collected at 0 h are not very different from each other, because there is a slight overlap of the 95% confidence interval (CI) ellipse between these groupings (Fig. 5A).

A loading plot was produced that gives a relative score showing how much an individual compound influences the variability of each component among samples (Fig. 5B). Unknown compounds contribute strongly to genotype-derived variability (component 1), whereas many compounds that contribute to sampling time-derived variability (component 2) are known primary metabolites such as glucose-6-phosphate and branched-chain amino acids. The connection of primary metabolites to

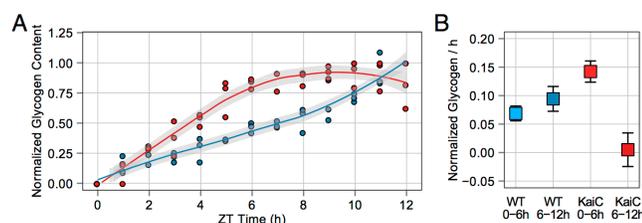


Fig. 3. Summary of glycogen accumulation data over a 12-h light period collected from WT and $\Delta kaiC$ cells growing in a 12:12 LD cycle. (A) Normalized glycogen content from WT (blue circles) and $\Delta kaiC$ (red circles) cells collected at 1-h intervals after cells were released into the light. Glycogen content for each replicate was normalized to the maximum value in the 12-h period. The data indicate that $\Delta kaiC$ accumulates glycogen more rapidly than WT early in the day. Best-fit curves were calculated for WT (blue line) and $\Delta kaiC$ (red line) cells using LOESS regression; the gray shaded area indicates the 95% CI for the regression line. Sampling for each strain was conducted in triplicate. (B) Slope calculated using linear regression of normalized glycogen content for the given time intervals. The glycogen accumulation rate for WT does not significantly differ over the time course, whereas $\Delta kaiC$ displays significantly different rates of glycogen accumulation in the first and last 6 h of the day period. The $\Delta kaiC$ strain also shows significantly more rapid accumulation than WT in the first 6 h. Error bars indicate the 95% CI of the slope estimate. Each slope was calculated from 18 data points.

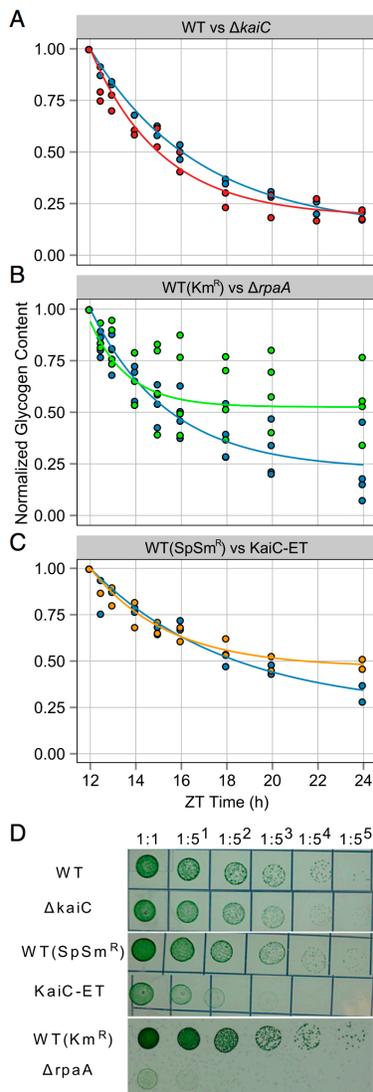


Fig. 4. Summary of glycogen degradation data and LD growth phenotypes for WT, $\Delta kaiC$, $\Delta rpaA$, and KaiC-ET strains. Samples for all glycogen degradation rate experiments were collected at 0, 0.5-, 1-, 2-, 3-, 4-, 6-, 8-, and 12-h time points after cells entered a dark period during a 12:12 LD diurnal cycle. Glycogen content for each replicate was normalized to the glycogen value at 12 h after lights on. The best fit for each set of data was modeled using first-order decay and is indicated by a solid line; coefficients are given in the text. (A) Normalized glycogen content from WT (blue circles) and $\Delta kaiC$ (red circles). First-order decay model for WT (blue line) and $\Delta kaiC$ (red line) indicates that glycogen degradation is similar in these strains. The experiment was performed in duplicate for both strains. (B) Normalized glycogen content from WT (blue circles) and $\Delta rpaA$ (green circles). The first-order decay model

time is indicative of the activation of primary metabolism after a dark-to-light transition. Some metabolites also contribute strongly to both components. These metabolites, such as sucrose and tryptophan, are interesting because, although they change after the dark-to-light transition, the nature of their variability is strongly affected by the presence or absence of KaiC. Overall, the status of the circadian oscillator contributes more to the variability than a dark-to-light transition. Strikingly, the compounds that contribute most strongly to genotype-related differences are unknowns. Finally, it is likely that metabolic differences accumulate over the time course, because the most divergent samples are the WT and $\Delta kaiC$ mutant strains at 4 h after lights on.

Metabolites significantly altered in dynamics or abundance. We identified 21 known and 29 unknown compounds that differed significantly in at least one pairwise comparison between sample types (Dataset S1). Based on PLS-DA, we focused on compounds that (i) changed significantly between the 0 h and 4 h time points (Fig. 6A) and (ii) had significantly different abundances in the WT and $\Delta kaiC$ strains at the 4-h time point (Fig. 6B). The metabolites that changed significantly over time in both the WT and $\Delta kaiC$ strains are primarily known metabolites (Fig. 6A and Table S1). Also, the direction of change over time was similar for many of these compounds in both strains. In contrast the majority of metabolites (11 of 12) that change over time only in the WT strain are unknown species. Some of these metabolites, such as BBID#106943 and BBID#101299, change strongly with time in WT cells but show effectively no change over time in the $kaiC$ mutant (Fig. 6A and Table S1). Only four compounds changed significantly over time uniquely in $\Delta kaiC$. One target, fructose-6-phosphate, is a known intermediate of the OPPP and shows a fourfold increase. Previous work on *S. elongatus* suggests that flux through this compound is indicative of OPPP activity (36). Additionally, the $\Delta kaiC$ mutant showed a 2.5-fold decrease of the unknown BBID#106921. This compound shows opposite metabolic movement between genotypes over the time course.

In $\Delta kaiC$ a number of primary metabolites were elevated relative to WT by 4 h in the light (Fig. 6B). Most notably, sucrose was elevated more than sixfold. Glucose-6-phosphate, fructose-6-phosphate, and inulotriose, which are connected to glycolysis, the OPPP, and glycogen biosynthesis, were also elevated significantly (Fig. 6B and C). Tryptophan, a product of the shikimate pathway, which is fed directly by the OPPP, was ~4.5-fold more abundant in $\Delta kaiC$ at 4 h. In contrast, a number of unknown compounds that were very abundant in the WT strain had extremely depressed levels in $\Delta kaiC$ mutants. Two of these compounds, BBID#106921 and BBID#1721, were more than 100-fold less abundant in $\Delta kaiC$, but, respectively, they were the third and sixth most abundant compounds detected in WT cells at 4 h (Fig. 6B and Table S2). In $\Delta kaiC$ these metabolites are only the 219th and 187th most abundant at 4 h, respectively.

In summary, the inactivation of *kaiC* appears to have a direct impact on how metabolites are partitioned in the cell after a dark-to-light transition. Both strains increase pool sizes of primary metabolites over the time course; however, $\Delta kaiC$ accumulates

for WT (blue line) and $\Delta rpaA$ (green line) indicates that glycogen degradation is severely attenuated in the $\Delta rpaA$ strain. The experiment performed in quadruplicate because of the known high variability in the $\Delta rpaA$ strain. (C) Normalized glycogen content from WT (blue circles) and KaiC-ET (orange circles). The first-order decay model for WT (blue line) and KaiC-ET (green line) indicates that glycogen degradation is attenuated in the KaiC-ET strain. The experiment was performed in duplicate. (D) Dilution series of strains grown on solid BG-11 medium for 5–7 d in a 12:12 LD cycle. (Top) WT and $\Delta kaiC$ have similar growth kinetics under these conditions. However, KaiC-ET (Middle) and $\Delta rpaA$ (Bottom) have severely attenuated growth when grown in a diel cycle. Images are representative of multiple experiments.

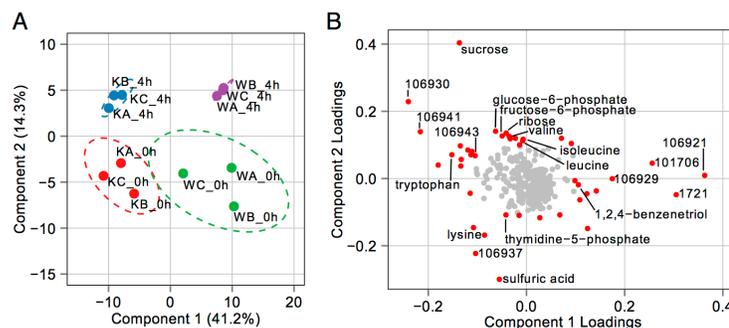


Fig. 5. Summary of dimension reduction performed on metabolomics data from WT and $\Delta kaiC$ cells grown in a 12:12 LD cycle at the 0-h and 4-h time points after entering light. (A) Plot of PLS-DA components 1 and 2 for all metabolomics samples. Components 1 and 2 account for 55.5% of the variance in the dataset, and, based on these components, all samples show good clustering with biological replicates. The plot indicates that component 1 describes genotype-derived variability, whereas component 2 describes sampling time-derived variability. Ellipses indicate the 95% CI of each grouping of samples on the plot. “W” indicates a WT sample while “K” indicates a $\Delta kaiC$ mutant. The letters A, B, and C represent the three biological replicates taken for each sample time point. (B) Loading plot derived from PLS-DA components 1 and 2 indicating the importance of each metabolite to the variability of a given component. Points in red are compounds for which one of the loadings was at least ± 0.1 . Points in gray are compounds for which no loading was greater than ± 0.1 . The plot shows that many unknown compounds drive variability in component 1 whereas known and unknown compounds drive variability in component 2.

much larger amounts of primary metabolites, specifically those involved in and directly connected to the OPPP, such as fructose-6-phosphate and sucrose. In contrast, WT cells mobilize carbon into a number of unknown compounds that are present only at low levels in $\Delta kaiC$ and make up a significant portion of the overall WT sample.

Correlations in Metabolite Abundance Can Help Classify Unknown Compounds. To identify shared pathways and suggest the biochemical context for the unknown metabolites that change remarkably in WT, we applied intermetabolite correlation analysis to look for groups of metabolites that share similar patterns of abundance (37). We compared the set of 50 metabolites with significant changes identified by ANOVA, which includes our unknown metabolites of interest (Dataset S1), with all of the known metabolites that were used in the ANOVA analysis (Materials and Methods). Pearson correlations were computed between the abundances of compounds in these two groups, which contained 50 and 111 compounds, respectively (Dataset S2). This analysis yielded 5,550 correlation coefficients from all possible pairwise comparisons. Subsequently, we used hierarchical clustering to group the correlation coefficients into clusters with similarity to each other. For the 50 metabolites with at least one significant change between samples, we could identify three distinct groups that we call “target clusters” (TC), for which the correlations to the 111 known metabolites formed a unique pattern. Similarly, when we looked at all 111 known metabolites we could identify six distinct groups, which we call “metabolite clusters” (MC), for which a group of known metabolites has a unique pattern of correlations across TCs. The correlations are presented as an ordered heat map with TCs on the x axis and MCs on the y axis (Fig. 7, Table S3, and Dataset S2).

We found that 11 of the 14 metabolites identified as more abundant in $\Delta kaiC$ at 4 h are clustered in TC2, whereas all seven metabolites significantly depressed in KaiC cells relative to WT cells at 4 h are found in TC3 (Table S3). Although TC1 and TC2 share similar correlation patterns across the six metabolite clusters, this pattern is very different from the pattern of TC3 across the same clusters (Fig. 7). TC1 and TC2 correlate positively with MC3 and MC5 and negatively with MC1 and MC2,

whereas TC3 has the opposite pattern, correlating negatively with MC3 and MC5 and positively with MC1 and MC2 (Fig. 7). The compounds that make up MC3 and MC5 are strongly enriched for roles in primary metabolic pathways, such as starch and sucrose metabolism ($P = 9.95e-7$), the pentose phosphate pathway ($P = 2.20e-6$), branched chain amino acid biosynthesis ($P = 3.26e-3$), and purine metabolism ($P = 9.99e-4$) (Fig. S3). Thus, TC1 and TC2 represent groupings of metabolites that increase together with primary metabolic activity, including sugar phosphates, nucleotides, and amino acids. This pattern is clearly evident in TC1, because this cluster contains many of the primary metabolites that increase in both strains after a dark-to-light transition (Table S3). In contrast, the compounds that make up MC1 and MC2 are enriched for roles in secondary metabolic pathways, such as fatty acid biosynthesis ($P = 1.96e-3$) and glycerolipid metabolism ($P = 5.97e-3$) (Fig. S3). MC1 and MC2 also contain a number of benzoate compounds that have been detected previously in cyanobacteria, including benzoic acid and 4-hydroxybenzoate (4HB) (Fig. S4 and Dataset S2) (38). Recent work has shown that plastoquinone biosynthesis in cyanobacteria uses 4HB as an intermediate (39, 40). Indeed, the benzoate compounds in MC1 and MC2 correlate negatively with the aromatic amino acids, which are consumed in plastoquinone biosynthesis (Fig. S4). Thus, it is likely MC1 and MC2 also are enriched in compounds with roles in biosynthesis of plastoquinone or other quinone-like molecules. Overall, our correlation analysis suggests that unknown compounds elevated in $\Delta kaiC$ are located primarily in TC2 and likely function in primary metabolic pathways or increase during their activation. In contrast, the unknown compounds elevated in the WT strain, which are exclusively found in TC3, likely function in secondary metabolic roles associated with lipid, glycerolipid, and possibly quinone biosynthesis.

Discussion

Before this work very limited data were available on the diurnal metabolism of *S. elongatus*, and no study had attempted to decouple the influences of the circadian clock and dark-to-light transitions on metabolism when cells are grown in a diurnal cycle. Our major conclusions from the collected data are that

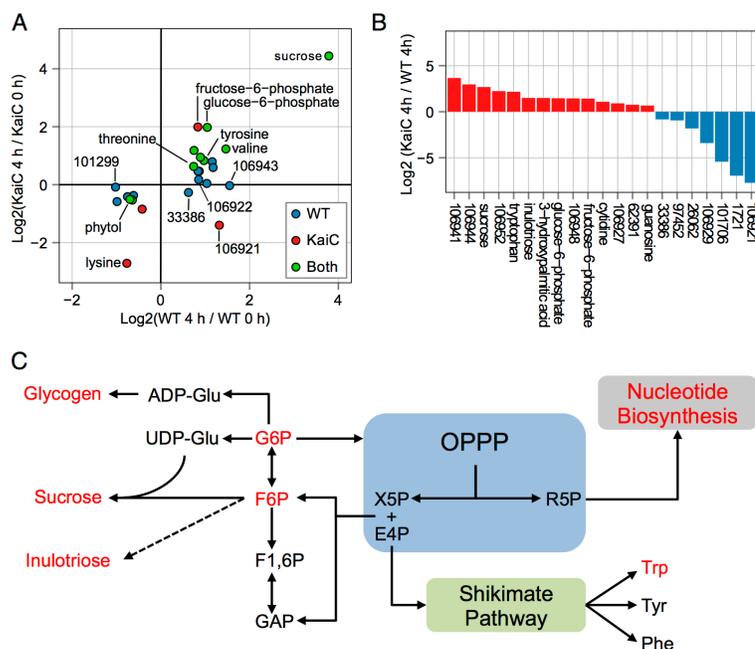


Fig. 6. Summary of metabolites that differ significantly in the WT and $\Delta kaiC$ strains. (A) Scatter plot of metabolites that show a significant change in abundance from 0 h to 4 h in WT, $\Delta kaiC$, or both strains. A significant change of a compound in a strain is indicated by the dot color. The \log_2 fold change from 0 h to 4 h after entering light is indicated on the x axis for WT and on y axis for $\Delta kaiC$ strains. (B) Plot of all metabolites that differ in abundance between WT and $\Delta kaiC$ at the 4-h sampling time point. Metabolite bars in red are significantly elevated and metabolite bars in blue are significantly reduced in $\Delta kaiC$ relative to WT. Although many primary metabolites are relatively elevated in $\Delta kaiC$ strains, all the metabolites in which $\Delta kaiC$ is reduced relative to WT are unknown compounds. Some of the unknowns are >100-fold less abundant in $\Delta kaiC$ strains. (C) Pathway diagram detailing the interconnections of the OPPP to glycolysis/glycogen metabolism, the Shikimate pathway, and nucleotide metabolism and indicating compounds that were significantly elevated in $\Delta kaiC$ relative to WT at the 4-h time point (red). Many of the elevated metabolites share the OPPP as a precursor hub; the monomers of many elevated sugar polymers were elevated also.

(i) the output from the core oscillator is dispensable for the degradation of carbon at night in a diel cycle; (ii) KaiC output inhibits RpaA, which serves to block activation of nighttime metabolic processes in the morning; and (iii) the importance of the circadian oscillator with respect to metabolism is primarily to modulate the balance between the Calvin cycle and the OPPP under diurnal growth conditions. The data are consistent with a model in which the clock serves to regulate RpaA activity negatively, and hence class 1 gene expression, in the morning. A decrease in inhibitory oscillator output over the day would allow RpaA to activate class 1 genes closer to dusk. This model agrees with data from Paddock et al. (19) suggesting that maximum output activity from the circadian oscillator occurs when KaiC is in the KaiC-pST (KaiC-ET) phospho-state, which is most abundant at dawn. The question remains as to what metabolic processes are driven by RpaA that are important for LD viability. Inactivation of a number of RpaA targets, such as *zwf* and *gnd* in the OPPP, also causes an LD sensitivity phenotype. However, it is unclear where carbon flows at night in *S. elongatus* and why these pathways are so critical for survival under these conditions. Our data suggest that normal KaiC output activity primarily affects metabolic processes that occur in the morning, because the largest differences between the WT and $\Delta kaiC$ strains in both glycogen kinetics and global metabolite partitioning are seen at this time.

The $\Delta kaiC$ mutant accumulates larger pools of glycogen precursors and primary carbon metabolites early in the day period (Figs. 3A and 6B). However, gene-expression data from LL conditions show that, relative to WT, the $\Delta kaiC$ mutant has significantly higher morning expression of transcripts involved in glycogen and carbon catabolism (21). Under diurnal growth conditions it is likely that multiple factors influence the flow of carbon in *S. elongatus*, including transcription, allosteric regulation of enzymes, and stoichiometric ratios of metabolites. Upon entering a morning period, when glycogen stores are low and photosynthesis is active, glycogen levels may not be strongly influenced by transcript levels from catabolism genes and instead reflect changes in other connected metabolic processes and allosteric regulation of glycogen biosynthetic enzymes. Indeed, GlgC is allosterically activated by the photosynthetic product 3-phosphoglycerate and a reducing cellular environment (29, 41). Alternatively, when cells enter a dark period, glycogen content is high, and GlgC is allosterically inactive. Under these conditions transcriptional activation by RpaA and the availability of degradative transcripts is a primary driving factor in glycogen catabolism. Activation of the OPPP in the morning by RpaA may, in fact, increase the availability of precursors for glycogen biosynthesis during a time when GlgC is strongly activated.

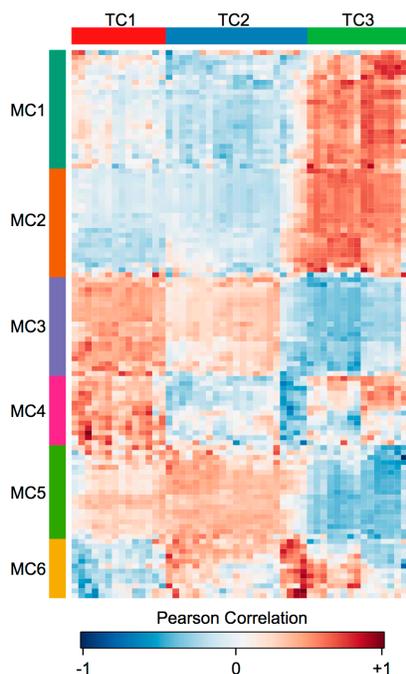


Fig. 7. Heatmap of the correlation between the groupings of metabolites identified by ANOVA to have some significant change (TCs) and a filtered set of all detected known compounds (MCs). More intense red color indicates the abundance patterns between two compounds in all collected samples are more positively correlated; more intense blue color indicates a negatively correlated abundance pattern. TC1 and TC2 have similar patterns of correlations across all known compounds, whereas TC3 displays a unique pattern of correlation. Almost all the unknown compounds that are highly abundant in WT and significantly reduced in $\Delta kaiC$ can be found in TC3. Thus, TC3 may give metabolic context to the possible placement of these unknown compounds in metabolism.

Other transcriptional changes in the $\Delta kaiC$ strain also may indirectly affect the regulatory protein CP12 (Calvin cycle protein 12), a master regulator of the Calvin cycle conserved between cyanobacteria and plants (36, 42). Reexamining the transcriptomics data from LL reveals that two of the most highly up-regulated genes in a $\Delta kaiC$ mutant are the pyridine nucleotide transhydrogenase subunits A and B (*pntA* and *pntB*), which also are known RpaA targets (17, 21). Products of these genes allow the interconversion of NADP(H) to NAD(H), and their overexpression may lower the normally high NADP(H)/NAD(H) ratio present during active photosynthesis. In *S. elongatus* low NADP(H)/NAD(H) levels activate CP12, causing a shift from the reductive Calvin cycle to the OPPP (36). In *S. elongatus* inactivation of CP12 resulted in decreased OPPP activity, in which a decrease in cellular fructose-6-phosphate could be detected directly (36). Additionally, in tobacco plants more active CP12 was associated with more starch, soluble sugars (including sucrose), and amino acids (43). The metabolic shifts observed in a $\Delta kaiC$ mutant in the morning mirror those seen when CP12 is active, including increased levels of fructose-6-phosphate, sucrose, nucleotides, and amino acids (Fig. 6 B and C). In contrast, the

repression of CP12 in tobacco resulted in accumulation of complex insoluble metabolites such as protein and cell wall components (43). In WT *S. elongatus* we observe increased abundance of unknowns that correlate strongly with compounds involved in fatty acid and glycerolipid biosynthesis; both these biosynthetic pathways would be important for cell wall and membrane biosynthesis in cyanobacteria. Thus, clock control may be important for regulating a shift between Calvin cycle activity and OPPP activity.

These data suggest a model in which KaiC output activity is important for inhibiting RpaA-driven OPPP activity in the morning. Inhibition of OPPP and other primary metabolic pathways frees up carbon so that it can be used in secondary biosynthetic processes. When inhibition of RpaA is relieved, it can activate its targets (including *pntA* and *pntB*) so that a lowering of the NADP(H)/NAD(H) ratio and activation of CP12 occurs. The strong correlation in WT samples of elevated unknown compounds with metabolites that participate in plastoquinone biosynthesis suggests that inhibition of primary metabolism by the clock in the morning may be important for this process (Fig. S4). Accumulation of plastoquinone in the morning not only would support photosynthesis through the day period but also would be important for its known role in resetting the circadian clock (44). The influence of the clock on accumulation of these compounds may represent a metabolic feedback loop in which the oscillator output is important for the biosynthesis of compounds that reinforce the correct oscillator timing in LD. In fact, both circadian control over starch metabolism and metabolic feedback to circadian timing have been observed previously in plants (45, 46). Thus, already there is some precedent for the existence of circadian timing reinforcement by metabolism in photosynthetic organisms.

Overall, this study highlights the importance of understanding the interaction of the circadian clock with light-to-dark transitions to gain insights into diurnal physiology and metabolism under day-night cycles. Some aspects of metabolism may be heavily dependent on the circadian clock, whereas others integrate both circadian influences and light availability. The expansion of mass spectral libraries and metabolic networks in photosynthetic organisms will be highly beneficial in determining the response to both internal circadian control and the external environment.

Materials and Methods

Cyanobacterial Strains, Media, and Culture Conditions. All strains were constructed in the *S. elongatus* PCC 7942 WT strain archived as AMC06 in our laboratory. Strains were constructed using standard procedures for cyanobacterial transformation (47) and are described in Table S4. All gene disruptions were validated by PCR of native loci. For all experiments precultures were prepared first by transferring 3 mL of stationary-phase culture into 100-mL flasks of fresh BG-11 medium (48) with appropriate antibiotics (5 $\mu\text{g}/\text{mL}$ kanamycin or 2 $\mu\text{g}/\text{mL}$ combination streptomycin/spectinomycin). Precultures were grown for 3–4 d at 30 °C, 150 rpm shaking (Thermo Fisher MaxQ 2000 Orbital Shaker), and 150 $\mu\text{E}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ constant light.

For all glycogen tracking and metabolomics experiments, the precultures were used to inoculate Phenometrics ePBR v1.1 photobioreactors (Phenometrics Inc.). Polycarbonate bioreactor vessels were inoculated to a volume of 400 mL, $\text{OD}_{750} = 0.1$ in medium that contained appropriate antibiotics. For all experiments temperature was maintained at 30 °C, 0.2 μm filtered air was sparged at a rate of 50 mL/min, and light intensity was 150 $\mu\text{E}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$, provided from the top of the culture, whenever lights were on. Controlled airflow was important for reproducibility of glycogen levels. After inoculation, all cultures were allowed to grow in LL until $\text{OD}_{750} = 0.3$. Cells then were maintained turbidostatically at this density for the duration of the experiments. For all strains, with the exception of the dark-sensitive strains $\Delta rpaA$ and KaiC-ET, circadian rhythms were entrained by growth in a 12:12 LD cycle for 3 d before release into experimental conditions and sampling. Dark-sensitive strains were maintained in LL before sampling periods.

For testing LD sensitivity, precultures were diluted initially to $\text{OD}_{750} = 0.2$, and subsequently were serially diluted five times 1:5 in fresh BG-11 medium. Drops of 4 μL from each dilution were plated on solid BG-11 medium with appropriate antibiotics and 1 mM $\text{Na}_2\text{S}_2\text{O}_3$. Plates were placed

at 30 °C/150 $\mu\text{E}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ constant light for 24 h and subsequently were transferred to 30 °C/150 $\mu\text{E}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ in a 12:12 light:dark cycle for 5–7 d.

KEGG Pathway Analysis. The KEGG pathways syf00030 (pentose phosphate pathway), syf00710 (carbon fixation in photosynthetic organisms), and syf00010 (glycolysis and gluconeogenesis) were cross-referenced for shared and unshared metabolic reactions. Peak circadian expression of genes that control metabolic pathway reactions was determined by data from Vijayan et al. (10). The number of dawn- or dusk-peaking genes unique to each pathway was compared with expected numbers of dawn- or dusk peaking genes in a random sample of genes, and *P* values were calculated using Fisher's exact test.

Glycogen Extraction and Analysis. For glycogen assay, 10 mL of culture (OD_{750} ~0.3) was collected and placed on ice. Cells were collected by centrifugation for 10 min at 4,000 \times g and 4 °C. The supernatant fraction was discarded, and pellets were frozen at –80 °C. Glycogen was extracted using methods modified from Ernst et al. (49). Specifically, a solution of 50 μL of sterile water and 200 μL of KOH [30% (wt/vol)] was used to resuspend cell pellets, which then were placed at 100 °C for 1.5 h. Glycogen was precipitated from extracts by adding 1 mL of 100% EtOH, and placing extracts on ice for 1 h. Precipitated glycogen was collected by centrifugation. Supernatant was discarded, and extracted glycogen was washed two times with 1 mL of 100% ethanol. Extracts were dried in a Speed-Vac (catalog no. 7810010; Labconco) for 15 min at 60 °C. Extracted glycogen was resuspended in 500 μL of 25 mM sodium acetate buffer (pH = 5) and stored overnight at 4 °C before assay. To quantify glycogen, 200 μL of each sample as well as purified glycogen standards (250, 200, 150, 100, 50, 25, 0 $\mu\text{g}/\text{mL}$) were mixed with 5 μL (5.5 U) of amyloglucosidase (catalog no. 10115; Sigma) and incubated at 37 °C for 1 h. Glucose in the resulting digest was determined by mixing 10 μL of digested glycogen with 190 μL of a solution containing 0.5 U glucose oxidase/0.1U peroxidase (catalog no. G3660; Sigma), 50 μM Amplex Red (catalog no. 10010469; Cayman Chemical), and 25 mM sodium acetate (pH 5). Reactions were incubated for 45 min at 23 °C, and absorbance at 540 nm was determined with a Tecan Infinite M200 plate reader. Unknown glycogen content was determined by comparison with purified standards, and background glucose content was determined by assay of samples untreated with amyloglucosidase.

Glycogen Kinetic Analysis. Glycogen accumulation was modeled using the Local regrESSion (LOESS) algorithm for local fitting with default parameters in the R plotting package ggplot2 (50). Accumulation rates for early and late time points were modeled using the linear modeling function in the base R statistical package (51). Glycogen degradation was modeled as a first-order decay process using the following mathematical expression:

$$G_T = (1 - G_{T12})^{(-\lambda T)} + G_{T12}.$$

Glycogen values (G_T) at the indicated time points (T) were provided to the model. The model was solved for the degradation rate constant (λ) and

terminal glycogen content (G_{T12}) using the nonlinear least squares function in the base R statistical package (51). Errors indicated for all modeled coefficients and graphs are presented as values encompassing the 95% CI of the data (52). All graphics were produced using the R plotting package ggplot2 (50).

Metabolomics and Data Analysis. Strains for metabolomics analysis were grown in photobioreactors as described above. At sampling time points 40 mL of culture was collected over ice in a 50-mL conical tube ($n = 3$ for all samples). Cells were collected immediately by centrifugation for 10 min at 4,000 \times g at –10 °C. Cell pellets were frozen rapidly in liquid nitrogen and placed at –80 °C before analysis. During sampling, the glycogen content of cells was tracked and confirmed to be similar to the accumulation behavior observed in Fig. 3A. Cell pellets were shipped on dry ice to the West Coast Metabolomics Center at the University of California, Davis for subsequent analysis. Metabolite extraction, derivatization, and analysis by GC-TOF-MS have been described in previous publications by Fiehn, et al. (34, 53). Metabolites were identified from MS spectra using the BinBase algorithm (34).

Raw abundance data for all known and unknown metabolites, consisting of unique ion peak heights, were analyzed with MetaboAnalyst (54). Principal component analysis (PCA) was applied to raw data as a quality-control measure to observe sample replicate groupings (Fig. 55). Raw data subsequently were filtered using interquartile range (IQR) to remove metabolites that showed very little variability over all samples. Filtered data were plotted using \log_2 normalization (Fig. 56). A mixture of univariate and multivariate statistics then was applied to investigate changes between genotypes and through dark-to-light transitions. PLS-DA was applied using default settings and was cross-validated using a maximum of two components (permutation $P < 0.01$). Differences in mean abundance between metabolites in different samples were assessed with ANOVA, and significance was determined using Tukey's honestly significant difference with a threshold of $P < 0.05$. To build the correlation matrix, metabolites identified as statistically significant by ANOVA were compared with all known metabolites present in the IQR-filtered set. Correlation between metabolites was calculated using Pearson's correlation statistic (r). Metabolite correlations were clustered with hierarchical clustering using Pearson correlation for the distance measure and average linkage for leaf ordering (Multiple Array Viewer v10.2). Cluster groupings were selected by eye, and KEGG pathway enrichment analysis was conducted on clusters using MBRole (55) with a false discovery rate (FDR) of 5% ($q < 0.5$).

ACKNOWLEDGMENTS. We thank Ryan Simkovsky and Mark Paddock for helpful discussions that improved the manuscript, and Anish Pal and Emily Effener for assistance in sample collection and strain maintenance. This work was supported by National Science Foundation Grant MCB1244108. S.D. was supported in part by National Institute of Health Cell and Molecular Genetics Training Grant T32GM007240.

- Deng MD, Coleman JR (1999) Ethanol synthesis by genetic engineering in cyanobacteria. *Appl Environ Microbiol* 65(2):523–528.
- Atsumi S, Higashide W, Liao JC (2009) Direct photosynthetic recycling of carbon dioxide to isobutyraldehyde. *Nat Biotechnol* 27(12):1177–1180.
- Schirmer A, Rude MA, Li X, Popova E, del Cardayre SB (2010) Microbial biosynthesis of alkanes. *Science* 329(5991):559–562.
- Kruse O, Hankamer B (2010) Microalgal hydrogen production. *Curr Opin Biotechnol* 21(3):238–243.
- Wijffels RH, Kruse O, Hellingwerf KJ (2013) Potential of industrial biotechnology with cyanobacteria and eukaryotic microalgae. *Curr Opin Biotechnol* 24(3):405–413.
- Guo J, et al. (2014) Proteome-wide light/dark modulation of thiol oxidation in cyanobacteria revealed by quantitative site-specific redox proteomics. *Mol Cell Proteomics* 13(12):3270–3285.
- Vermaas WF (2001) Photosynthesis and respiration in cyanobacteria. *Encyclopedia of Life Sciences* (Nature Publishing Group, London), pp 245–251.
- Dong G, Kim Y-I, Golden SS (2010) Simplicity and complexity in the cyanobacterial circadian clock mechanism. *Curr Opin Genet Dev* 20(6):619–625.
- Yang Q, Pando BF, Dong G, Golden SS, van Oudenaarden A (2010) Circadian gating of the cell cycle revealed in single cyanobacterial cells. *Science* 327(5972):1522–1526.
- Vijayan V, Zuzov R, O'Shea EK (2009) Oscillations in supercoiling drive circadian gene expression in cyanobacteria. *Proc Natl Acad Sci USA* 106(52):22564–22568.
- Mackey SR, Golden SS, Ditty JL (2011) The itty-bitty time machine: Genetics of the cyanobacterial circadian clock. *Adv Genet* 74:13–53.
- Nieva NB, Gao T, LiWang AC, Golden SS (2006) Quinone sensing by the circadian input kinase of the cyanobacterial circadian clock. *Proc Natl Acad Sci USA* 103(46):17468–17473.
- Wood TL, et al. (2010) The KaiA protein of the cyanobacterial circadian oscillator is modulated by a redox-active cofactor. *Proc Natl Acad Sci USA* 107(13):5804–5809.
- Rust MJ, Golden SS, O'Shea EK (2011) Light-driven changes in energy metabolism directly entrain the cyanobacterial circadian oscillator. *Science* 331(6014):220–223.
- Pattanayak GK, Phong C, Rust MJ (2014) Rhythms in energy storage control the ability of the cyanobacterial circadian clock to reset. *Curr Biol* 24(16):1934–1938.
- Takai N, et al. (2006) A KaiC-associating SasA-RpaA two-component regulatory system as a major circadian timing mediator in cyanobacteria. *Proc Natl Acad Sci USA* 103(32):12109–12114.
- Markson JS, Piechura JR, Puszynska AM, O'Shea EK (2013) Circadian control of global gene expression by the cyanobacterial master regulator RpaA. *Cell* 155(6):1396–1408.
- Johnson CH, Golden SS (1999) Circadian programs in cyanobacteria: Adaptiveness and mechanism. *Annu Rev Microbiol* 53:389–409.
- Paddock ML, Boyd JS, Adin DM, Golden SS (2013) Active output state of the *Synechococcus* Kai circadian oscillator. *Proc Natl Acad Sci USA* 110(40):E3849–E3857.
- Taniguchi Y, et al. (2007) labA: A novel gene required for negative feedback regulation of the cyanobacterial circadian clock protein KaiC. *Genes Dev* 21(1):60–70.
- Ito H, et al. (2009) Cyanobacterial daily life with Kai-based circadian and diurnal genome-wide transcriptional control in *Synechococcus elongatus*. *Proc Natl Acad Sci USA* 106(33):14168–14173.
- Stöckel J, et al. (2011) Diurnal rhythms result in significant changes in the cellular protein complement in the cyanobacterium *Cyanothece* 51142. *PLoS ONE* 6(2):e16680.
- Guerreiro ACL, et al. (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(8):2042–2055.

24. Reinbothe C, et al. (2010) Chlorophyll biosynthesis: Spotlight on protochlorophyllide reduction. *Trends Plant Sci* 15(11):614–624.
25. Hosokawa N, et al. (2011) Circadian transcriptional regulation by the posttranslational oscillator without de novo clock gene expression in *Synechococcus*. *Proc Natl Acad Sci USA* 108(37):15396–15401.
26. Yang C, Hua Q, Shimizu K (2002) Integration of the information from gene expression and metabolic fluxes for the analysis of the regulatory mechanisms in *Synechocystis*. *Appl Microbiol Biotechnol* 58(6):813–822.
27. Osanai T, et al. (2005) Positive regulation of sugar catabolic pathways in the cyanobacterium *Synechocystis* sp. PCC 6803 by the group 2 sigma factor sigE. *J Biol Chem* 280(35):30653–30659.
28. Suzuki E, et al. (2007) Role of the GlgX protein in glycogen metabolism of the cyanobacterium, *Synechococcus elongatus* PCC 7942. *Biochimica et Biophysica Acta* 1770(5):763–773.
29. Díaz-Troya S, López-Maury L, Sánchez-Riego AM, Roldán M, Florencio FJ (2014) Redox regulation of glycogen biosynthesis in the cyanobacterium *Synechocystis* sp. PCC 6803: Analysis of the AGP and glycogen synthases. *Mol Plant* 7(1):87–100.
30. Gründel M, Scheunemann R, Lockau W, Zilliges Y (2012) Impaired glycogen synthesis causes metabolic overflow reactions and affects stress responses in the cyanobacterium *Synechocystis* sp. PCC 6803. *Microbiology* 158(Pt 12):3032–3043.
31. Scanlan DJ, Sundaram S, Newman J, Mann NH, Carr NG (1995) Characterization of a *zwf* mutant of *Synechococcus* sp. strain PCC 7942. *J Bacteriol* 177(9):2550–2553.
32. Doolittle WF, Singer RA (1974) Mutational analysis of dark endogenous metabolism in the blue-green bacterium *Anacystis nidulans*. *J Bacteriol* 119(3):677–683.
33. Boyd JS, Bordowitz JR, Bree AC, Golden SS (2013) An allele of the *crn* gene blocks cyanobacterial circadian rhythms. *Proc Natl Acad Sci USA* 110(34):13950–13955.
34. Fiehn O, et al. (2010) Plasma metabolomic profiles reflective of glucose homeostasis in non-diabetic and type 2 diabetic obese African-American women. *PLoS ONE* 5(12):e15234.
35. Maitra S, Yan J (2008) Principle component analysis and partial least squares: Two dimension reduction techniques for regression. *Applying Multivariate Statistical Models* (Casualty Actuarial Society, Quebec City, Quebec, Canada), Vol 79.
36. Tamoi M, Miyazaki T, Fukamizo T, Shigeoka S (2005) The Calvin cycle in cyanobacteria is regulated by CP12 via the NAD(H)/NADP(H) ratio under light/dark conditions. *Plant J* 42(4):504–513.
37. Tikunov Y, et al. (2005) A novel approach for nontargeted data analysis for metabolomics. Large-scale profiling of tomato fruit volatiles. *Plant Physiol* 139(3):1125–1137.
38. Schwarz D, et al. (2011) Metabolic and transcriptomic phenotyping of inorganic carbon acclimation in the Cyanobacterium *Synechococcus elongatus* PCC 7942. *Plant Physiol* 155(4):1640–1655.
39. Pfaff C, Glindemann N, Gruber J, Frentzen M, Sadre R (2014) Chorismate pyruvate-lyase and 4-hydroxy-3-solanesylbenzoate decarboxylase are required for plastoquinone biosynthesis in the cyanobacterium *Synechocystis* sp. PCC6803. *J Biol Chem* 289(5):2675–2686.
40. Sadre R, Pfaff C, Buchkremer S (2012) Plastoquinone-9 biosynthesis in cyanobacteria differs from that in plants and involves a novel 4-hydroxybenzoate solanesyl-transferase. *Biochem J* 442(3):621–629.
41. Ballicora MA, Iglesias AA, Preiss J (2003) ADP-glucose pyrophosphorylase, a regulatory enzyme for bacterial glycogen synthesis. *Microbiol Mol Biol Rev* 67(2):213–225.
42. Gontero B, Maberly SC (2012) An intrinsically disordered protein, CP12: Jack of all trades and master of the Calvin cycle. *Biochem Soc Trans* 40(5):995–999.
43. Howard TP, et al. (2011) Antisense suppression of the small chloroplast protein CP12 in tobacco alters carbon partitioning and severely restricts growth. *Plant Physiol* 157(2):620–631.
44. Kim Y-I, Vinyard DJ, Ananyev GM, Dismukes GC, Golden SS (2012) Oxidized quinones signal onset of darkness directly to the cyanobacterial circadian oscillator. *Proc Natl Acad Sci USA* 109(44):17765–17769.
45. Haydon MJ, Mielczarek O, Robertson FC, Hubbard KE, Webb AAR (2013) Photosynthetic entrainment of the *Arabidopsis thaliana* circadian clock. *Nature* 502(7473):689–692.
46. Streb S, Zeeman SC (2012) Starch metabolism in *Arabidopsis*. *The Arabidopsis Book* 10:e0160.
47. Clerico EM, Ditty JL, Golden SS (2007) Specialized techniques for site-directed mutagenesis in cyanobacteria. *Methods Mol Biol* 362:155–171.
48. Xu Y, Mori T, Johnson CH (2003) Cyanobacterial circadian clockwork: Roles of KaiA, KaiB and the kaiBC promoter in regulating KaiC. *EMBO J* 22(9):2117–2126.
49. Ernst A, Kirschenlohr H, Diez J, Böger P (1984) Glycogen content and nitrogenase activity in *Anabaena variabilis*. *Arch Microbiol* 140(2-3):120–125.
50. ggplot2: Elegant Graphics for Data Analysis (2009) *ggplot2: Elegant Graphics for Data Analysis* (Springer Science and Business Media, New York) Available at link.springer.com/book/10.1007%2F978-0-387-98141-3. Accessed December 1, 2014.
51. R Core Team (2014) R: A Language and Environment for Statistical Computing: Base package (Vienna).
52. R Core Team (2013) R: A Language and Environment for Statistical Computing. 1–3604.
53. Fiehn O, et al. (2008) Quality control for plant metabolomics: Reporting MSI-compliant studies. *Plant J* 53(4):691–704.
54. Xia J, Wishart DS (2011) Metabolomic data processing, analysis, and interpretation using MetaboAnalyst. *Curr Protoc Bioinformatics* Chapter 14:Unit 14.10.
55. Chagoyen M, Pazos F (2011) MBRole: Enrichment analysis of metabolomic data. *Bioinformatics* 27(5):730–731.

Supporting Information

Diamond et al. 10.1073/pnas.1504576112

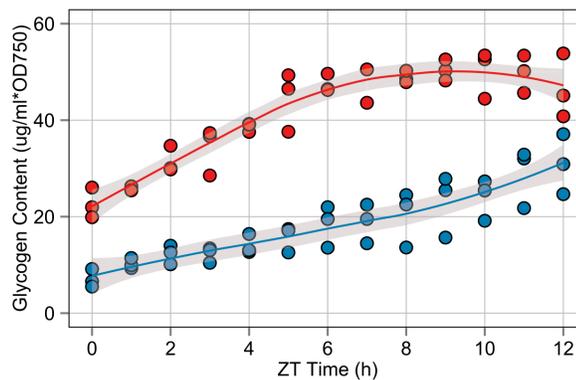


Fig. S1. Glycogen content in WT and $\Delta kaiC$ measured during the 12-h light period as cells entered a light period following a dark period (Fig. 3A). Best-fit curves were calculated for WT (blue line) and $\Delta kaiC$ (red line) cells using LOESS regression. The gray shaded area indicates the 95% CI for the regression line. Glycogen content was normalized to cell density based on OD at 750 nm. Sampling for each strain was conducted in triplicate. ZT, zeitgeber time.

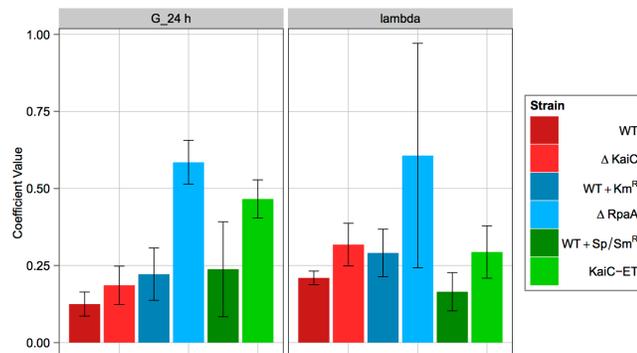


Fig. S2. Graphical representation of coefficients calculated from glycogen degradation rate data in Fig. 4 A–C and presented in the main text. The coefficient being compared is indicated at the top of the graph. Error bars indicate the 95% CI of the coefficient fit.

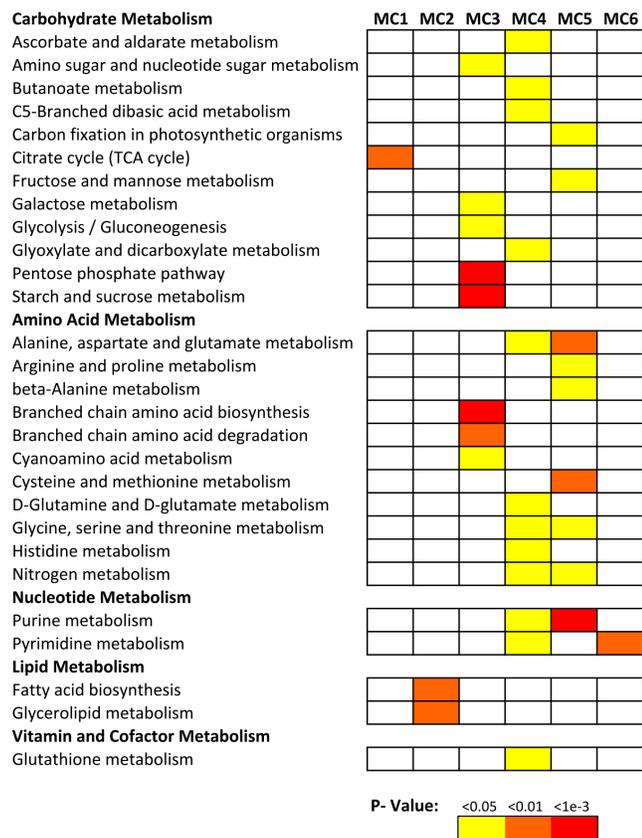
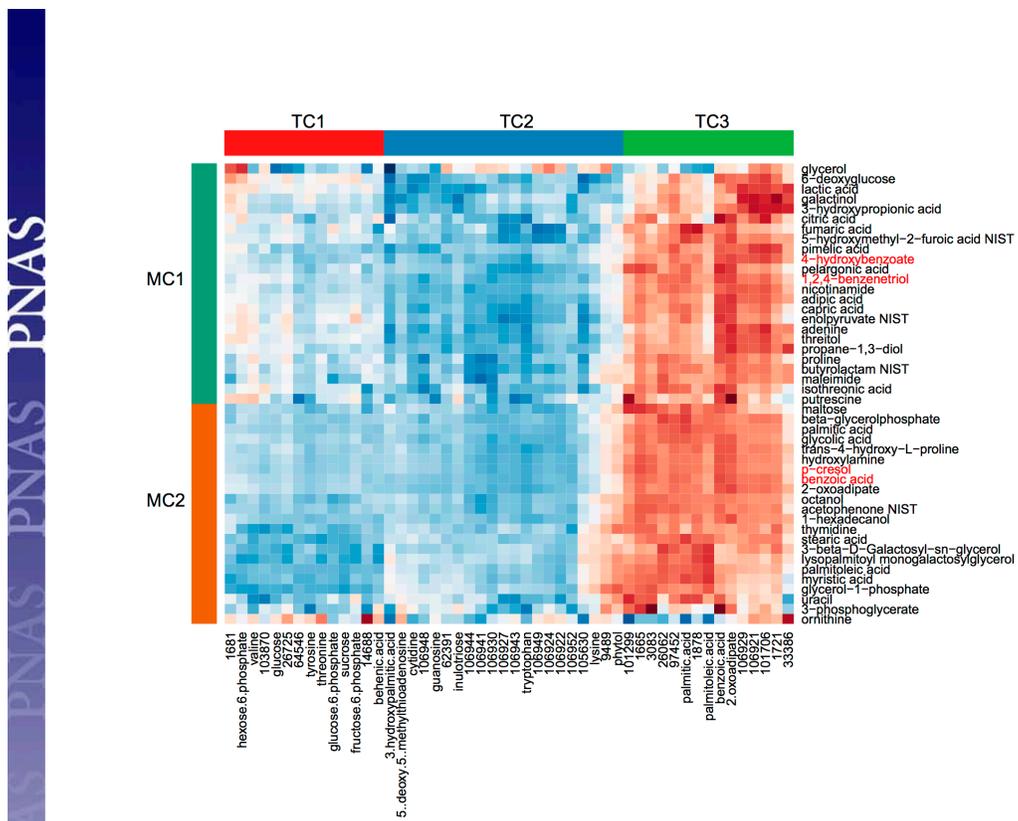


Fig. S3. KEGG functional category analysis of MCs from the heatmap in Fig. 7. Enrichment of compounds in a KEGG category was calculated using the online web server MBRole (1). A colored box indicates the presence of an enriched KEGG category in a given MC. The color of the box represents the FDR-corrected *P* value for that functional category.

1. Chagoyen M, Pazos F (2011) MBRole: Enrichment analysis of metabolomic data. *Bioinformatics* 27(5):730-731.



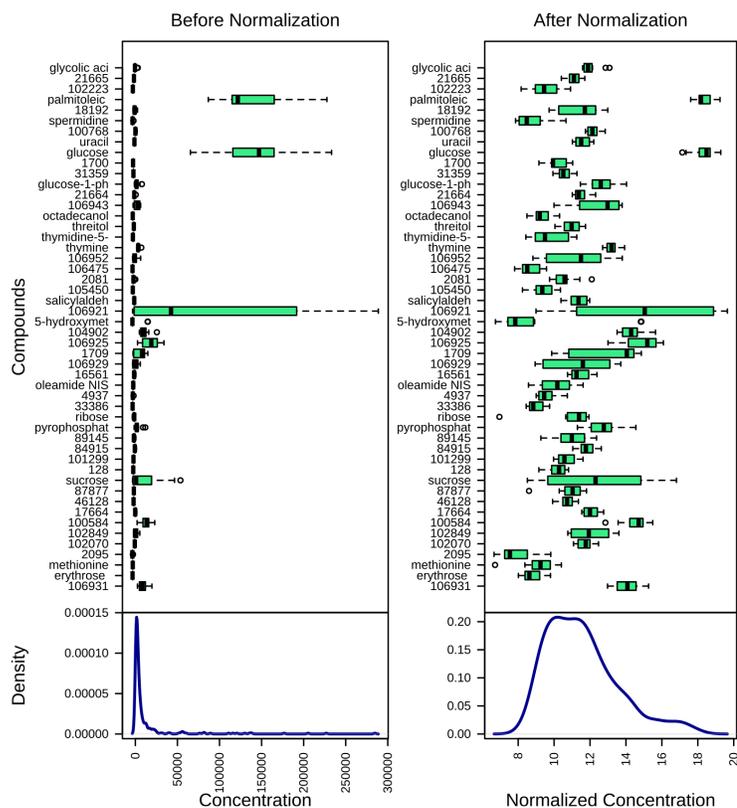


Fig. S6. The effect of \log_2 normalization on raw metabolite abundance before analysis. (Left) Before normalization the overall concentration distribution of the dataset is strongly skewed toward extreme values. (Right) After \log_2 normalization the distribution of compound concentrations much more closely resembles a normal probability distribution and is suitable for parametric statistical analysis such as ANOVA.

Table S1. Metabolites with significant change between 0 and 4 h

Compound	WT fold change*	ΔKaiC fold change*	Significant†
Valine	2.83	2.41	Both
Tyrosine	2.01	1.83	Both
Threonine	1.71	1.59	Both
Sucrose	14.14	22.32	Both
Phytol	0.66	0.71	Both
Palmitoleic acid	0.63	0.72	Both
Lysine	0.60	0.16	KaiC
Hexose-6-phosphate	1.90	1.98	Both
Glucose-6-phosphate	2.12	4.04	Both
Glucose	1.87	1.42	WT
Fructose-6-phosphate	1.83	4.08	KaiC
106943	3.00	1.00	WT
106922	1.85	1.16	WT
106921	2.56	0.39	KaiC
103870	2.29	1.78	WT
101299	0.51	0.97	WT
97452	0.67	0.79	WT
64546	2.32	1.55	WT
33386	1.58	0.85	WT
26725	1.83	1.42	WT
14688	2.10	1.06	WT
9489	0.77	0.57	KaiC
1878	0.52	0.69	WT
1681	1.72	2.33	Both
1665	0.61	0.77	WT

*Fold-change values are reported without log₂ normalization.†Strains named had a significant fold change ($P < 0.05$).**Table S2. Metabolites with significant difference in abundance at 4 h**

Compound	WT_4h*	WT rank†	KaiC_4h*	ΔKaiC rank†	High strain‡
Tryptophan	12.45	90	14.61	27	KaiC
Sucrose	12.76	71	15.44	19	KaiC
Inulotriose	8.54	308	10.03	226	KaiC
Guanosine	12.59	83	13.26	55	KaiC
Glucose-6-phosphate	10.69	182	12.16	98	KaiC
Fructose-6-phosphate	9.73	246	11.15	154	KaiC
Cytidine	9.37	265	10.45	197	KaiC
3-Hydroxypalmitic acid	8.99	290	10.48	193	KaiC
106952	10.60	191	12.84	67	KaiC
106948	10.67	185	12.11	102	KaiC
106944	9.16	281	12.13	100	KaiC
106941	10.65	187	14.31	29	KaiC
106929	12.86	67	9.46	260	WT
106927	12.58	84	13.48	51	KaiC
106921	17.93	3	10.21	219	WT
101706	14.94	23	9.51	256	WT
97452	12.51	87	11.57	129	WT
62391	11.82	117	12.57	77	KaiC
33386	9.72	247	8.89	290	WT
26062	13.39	56	11.58	128	WT
1721	17.53	6	10.58	187	WT

*Log₂ normalized abundance value at 4-h time point.

†Rank of metabolite's abundance at 4 h among the 325 identified metabolites.

‡Strain with the highest overall abundance of the compound at the 4-h time points.

Table S3. Target cluster information

Compound	Time*	Abundance†
Target cluster 1		
Valine	Both	—
Tyrosine	Both	—
Threonine	Both	—
Sucrose	Both	Elevated
Hexose-6-phosphate	Both	—
Glucose-6-phosphate	Both	Elevated
Glucose	WT	—
Fructose-6-phosphate	KaiC	Elevated
Behenic acid	—	—
103870	WT	—
64546	WT	—
26725	WT	—
14688	WT	—
1681	Both	—
Target cluster 2		
Tryptophan	—	Elevated
Phytol	Both	—
Lysine	KaiC	—
Inulotriose	—	Elevated
Guanosine	—	Elevated
Cytidine	—	Elevated
5'-methylthioadenosine	—	—
3-hydroxypalmitic acid	—	Elevated
106952	—	Elevated
106949	—	—
106948	—	Elevated
106944	—	Elevated
106943	WT	—
106941	—	Elevated
106930	—	—
106927	—	Elevated
106924	—	—
106922	WT	—
105630	—	—
62391	—	Elevated
9489	KaiC	—
Target cluster 3		
Palmitoleic acid	Both	—
Palmitic acid	—	—
Benzoic acid	—	—
2-oxoadipate	—	—
106929	—	Depressed
106921	KaiC	Depressed
101706	—	Depressed
101299	WT	—
97452	WT	Depressed
33386	WT	Depressed
26062	—	Depressed
3083	—	—
1878	WT	—
1721	—	Depressed
1665	WT	—

*Strain in which a metabolite changed from 0 to 4 h.

†Significantly elevated or depressed in Δ kaiC relative to WT at 4 h.

Table S4. Cyanobacterial strains used in this study

Strain	Genetic background	Antibiotic*	Source
WT	AMC 06	None	S.S.G. collection
WT ^{KmR}	AMC 06 transformed with pAM1579	Km	S.S.G. collection
WT ^{SpSmR}	AMC 06 transformed with pAM1303	SpSm	S.S.G. collection
$\Delta kaiC$ (AMC704)	$\Delta kaiC$ in-frame deletion in AMC541	Cm	(1)
KaiC-ET	AMC 704 transformed with pAM4685	SpSmCm	(2)
$\Delta rpaA$	AMC06 transformed with pAM4420	Km	S.S.G. collection
AMC541	AMC06 with P _{kaiB} - <i>luc</i> reporter in NS2	Cm	S.S.G. collection

*Antibiotics were not applied to the $\Delta kaiC$ strain. Cm, chloramphenicol; Km, kanamycin; Sm, streptomycin; Sp, spectinomycin.

1. Ditty JL, Canales SR, Anderson BE, Williams SB, Golden SS (2005) Stability of the *Synechococcus elongatus* PCC 7942 circadian clock under directed anti-phase expression of the kai genes. *Microbiology* 151(Pt 8):2605–2613.
2. Paddock ML, Boyd JS, Adin DM, Golden SS (2013) Active output state of the *Synechococcus Kai* circadian oscillator. *Proc Natl Acad Sci USA* 110(40):E3849–E3857.

Other Supporting Information Files

[Dataset S1 \(XLSX\)](#)

[Dataset S2 \(XLSX\)](#)

2.3 Physiology of the Light to Dark Transition



Redox crisis underlies conditional light–dark lethality in cyanobacterial mutants that lack the circadian regulator, RpaA

Spencer Diamond^{a,b,1}, Benjamin E. Rubin^{a,b}, Ryan K. Shultzaberger^c, You Chen^{b,2}, Chase D. Barber^a, and Susan S. Golden^{a,b,3}

^aDivision of Biological Sciences, University of California, San Diego, La Jolla, CA 92093; ^bCenter for Circadian Biology, University of California, San Diego, La Jolla, CA 92093; and ^cKavli Institute for Brain and Mind, University of California, San Diego, La Jolla, CA 92093

Contributed by Susan S. Golden, December 14, 2016 (sent for review August 8, 2016; reviewed by Robert L. Burnap and Louis A. Sherman)

Cyanobacteria evolved a robust circadian clock, which has a profound influence on fitness and metabolism under daily light–dark (LD) cycles. In the model cyanobacterium *Synechococcus elongatus* PCC 7942, a functional clock is not required for diurnal growth, but mutants defective for the response regulator that mediates transcriptional rhythms in the wild-type, regulator of phycobilisome association A (RpaA), cannot be cultured under LD conditions. We found that *rpaA*-null mutants are inviable after several hours in the dark and compared the metabolomes of wild-type and *rpaA*-null strains to identify the source of lethality. Here, we show that the wild-type metabolome is very stable throughout the night, and this stability is lost in the absence of RpaA. Additionally, an *rpaA* mutant accumulates excessive reactive oxygen species (ROS) during the day and is unable to clear it during the night. The *rpaA*-null metabolome indicates that these cells are reductant-starved in the dark, likely because enzymes of the primary nighttime NADPH-producing pathway are direct targets of RpaA. Because NADPH is required for processes that detoxify ROS, conditional LD lethality likely results from inability of the mutant to activate reductant-requiring pathways that detoxify ROS when photosynthesis is not active. We identified second-site mutations and growth conditions that suppress LD lethality in the mutant background that support these conclusions. These results provide a mechanistic explanation as to why *rpaA*-null mutants die in the dark, further connect the clock to metabolism under diurnal growth, and indicate that RpaA likely has important unidentified functions during the day.

in constant light (LL) conditions to distinguish internal circadian regulation from that which is environmentally driven (16, 17). However, diurnal physiology in a natural environment must integrate the two sources of regulation. We recently showed that the circadian clock regulates carbon metabolism in *Synechococcus elongatus* PCC 7942 as cells transition from the dark into the light during diurnal growth (19). Specifically, in the morning the clock represses the activity of the conserved circadian transcriptional regulator regulator of phycobilisome association A (RpaA), which normally activates nighttime metabolic processes (19, 20). This action suppresses primary metabolic processes in the morning, allowing carbon to flow toward secondary metabolic processes when light energy is not limiting (19).

The circadian clock in *S. elongatus* comprises a core oscillator formed by the proteins KaiA, KaiB, and KaiC (16). The oscillator relays timing information to the SasA–RpaA two-component output pathway, in which RpaA is a transcription factor that binds 170 known downstream gene targets (20, 21). RpaA was first identified in another cyanobacterium, *Synechocystis* sp. strain PCC 6803, as an OmpR-type response regulator that influences the ratio of light energy transfer from light-harvesting phycobilisomes to photosystem I (PSI) vs. PSII (22). Thus, it is not surprising that RpaA affects core energy-producing pathways when cells are exposed to light. RpaA protein activity is directly controlled by the

cyanobacteria | metabolism | circadian clock | metabolomics | diurnal

Cyanobacteria are both key agents of global carbon and nitrogen cycles and promising platforms for renewable chemicals, fuels, and nutraceuticals (1–3). Understanding the control mechanisms that govern the flow of carbon and nitrogen through these organisms is crucial for predicting their behavior in natural environments as well as for improving engineering strategies. Although the basic pathways for carbon and nitrogen metabolism, and their regulation, are well understood in heterotrophic bacteria, cyanobacteria exhibit important deviations in these core metabolic pathways (4–7). Additionally, metabolic control mechanisms in cyanobacteria evolved to be compatible with photoautotrophic metabolism and the dramatic shifts that are imposed on those pathways by predictable daily light–dark (LD) cycles. Examples include enzymatic activity that responds to light-dependent cellular redox changes (8–11); the preference for NADPH, the reductant produced by the photochemical reactions, over NADH by many biosynthetic enzymes (12, 13); and a circadian clock that drives 24-h transcriptional rhythms in most genes (14–16).

A daily LD cycle presents a strong metabolic driver for the photosynthetic cyanobacteria, but a circadian clock also imposes daily cycles in transcription and redox regulatory systems (17, 18). Circadian measurements historically have been performed

Significance

The evolution of photosynthetic cyanobacteria under 24-h cycles of light and darkness selected for a robust circadian clock. Understanding how cyanobacteria integrate circadian clock signals with natural light–dark cycles to control metabolism is critical, because these organisms are central to global carbon cycling and hold promise for development of renewable energy. Here we assess how the circadian transcription factor regulator of phycobilisome association A (RpaA) influences metabolism as a cyanobacterium goes through a light-to-dark transition. The data show that RpaA plays a key role in maintaining metabolic stability during the night period. Additionally, RpaA is important in controlling redox balance, which in turn is very important for regulating metabolism at night.

Author contributions: S.D., B.E.R., R.K.S., and S.S.G. designed research; S.D., B.E.R., R.K.S., Y.C., and C.D.B. performed research; S.D., B.E.R., R.K.S., and S.S.G. analyzed data; and S.D., B.E.R., and S.S.G. wrote the paper.

Reviewers: R.L.B., Oklahoma State University; and L.A.S., Purdue University.

The authors declare no conflict of interest.

Dedicated to the memory of Dr. David B. Knaff.

¹Present address: Department of Earth and Planetary Science, University of California, Berkeley, CA 94704.

²Present address: BATI, Inc., San Diego, CA 92121.

³To whom correspondence should be addressed. Email: sgolden@ucsd.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1613078114/-DCSupplemental.

circadian oscillator, which represses the activity of RpaA in the late night/early morning, and subsequently relieves repression throughout the day such that RpaA reaches its peak activity at dusk (19, 20, 23). This temporal activity pattern, and the fact that RpaA directly binds and activates nighttime metabolism genes (20, 24), suggests that it also plays an important role in metabolic control at night. Although the transcriptional targets of RpaA have been identified, and it is clear that LD conditions are deleterious to *rpaA*-null mutants (25), the metabolic and physiological changes that attenuate growth under LD conditions have not been explored.

Under LD conditions, *S. elongatus* performs photosynthesis and carbon fixation during the day via the Calvin–Benson cycle, with excess fixed carbon stored as the branched chain glucose polymer glycogen (19). As cells enter a dark period, glycogen is rapidly degraded via the oxidative pentose phosphate pathway (OPPP), which serves as the primary source of energy and reducing power (NADPH) at night (26, 27). The OPPP shares many reactions with the Calvin–Benson cycle, and the transition from photosynthetic to oxidative metabolism occurs through both transcriptional and redox-regulated steps (28–30). Strict control of cellular redox via the NADPH/NADP⁺ ratio is a common and important mechanism across plant and cyanobacterial species (28). Additionally, RpaA transcriptionally activates genes that code for sugar catabolic and OPPP enzymes at the end of the day, before entering the dark, including *glgP* (glycogen phosphorylase), *gap1* (glyceraldehyde-3-phosphate dehydrogenase 1), *opcA* (OxPP cycle protein A), and the OPPP rate-limiting enzyme *zwf* (glucose-6-phosphate dehydrogenase) (20). In *rpaA* mutants, glycogen degradation is strongly attenuated, which reflects an inability to activate these sugar catabolic pathways (19).

In this study, we investigated whether the LD growth defect in a *rpaA*-null mutant is attributable to specific misregulation of metabolism and physiology as *S. elongatus* transitions into darkness and over the night period. We initially addressed the viability of an *rpaA*-null mutant (hereafter $\Delta rpaA$) over a 12-h dark period. Subsequently, we used untargeted metabolic profiling to investigate how loss of RpaA affects the abundance of primary metabolites at time points after cells enter the dark. Finally, we identified both second-site mutations and physiological growth conditions that suppress LD lethality in the $\Delta rpaA$ mutant and correlated these data with metabolomics, gene expression, and measurements of global oxidative stress. We present a model in which RpaA acts as a critical transcriptional activator of reductant-producing pathways and show that its activity is important to maintain strict metabolic stability at night. This work shows that, even in cyanobacteria that do not carry out obvious nighttime programs such as nitrogen fixation, carbon catabolism and reductant production at night are crucial for homeostasis. Thus, metabolic quiescence is not sustainable under diurnal growth conditions.

Results

Darkness Initiates Pigmentation Changes and Rapid Cell Death in the $\Delta rpaA$ Mutant. Although it is known that $\Delta rpaA$ strains do not grow under LD conditions (19, 25), the nature of the defect has not been characterized. We initially examined changes in cell viability and whole-cell absorbance as cultures entered the dark. WT and $\Delta rpaA$ cultures were sampled immediately before entry into the dark (0 h) and at intervals thereafter. Before dark exposure, the $\Delta rpaA$ strain had significantly elevated absorbance at 440 and 680 nm relative to WT (Fig. 1A), indicating an increase in chlorophyll absorbance. Although this study did not directly address changes in photosynthesis, a similar relative increase in chlorophyll was observed in *Synechocystis* sp. strain PCC 6803 $\Delta rpaA$ mutants (22). The broad differences in pigmentation at 0 h between WT and $\Delta rpaA$ (Fig. 1A) may indicate that the *S. elongatus* $\Delta rpaA$ mutant has altered energy-transfer kinetics

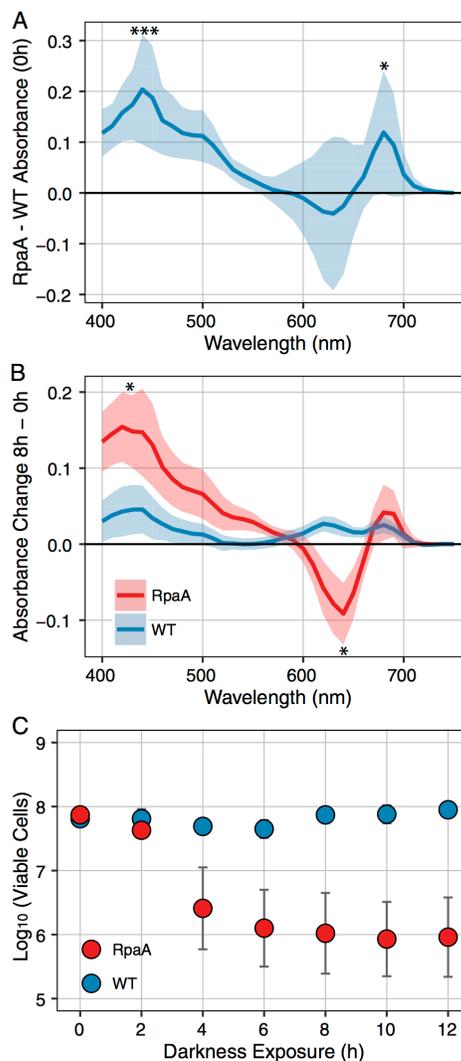


Fig. 1. Absorbance and viability data from WT and the $\Delta rpaA$ mutant. (A) Mean absorbance of WT subtracted from the $\Delta rpaA$ mutant at 0 h, immediately before entering darkness. Shaded area indicates SD of mean, significance of difference between WT and $\Delta rpaA$ calculated by Student's *t* test ($n = 8$). * $P < 0.05$; *** $P < 0.001$. (B) Change in absorbance of WT and $\Delta rpaA$ from 0 h immediately before entering darkness to 8 h of dark exposure. Shaded area indicates SD of mean, significance of difference between 0 and 8 h for each strain calculated by Student's *t* test ($n = 8$). * $P < 0.05$. (C) Mean viable cells counted at time points after WT and the $\Delta rpaA$ mutant entered the dark. Error bars indicate SEM. Significance was calculated by using Student's *t* test ($n = 4$).

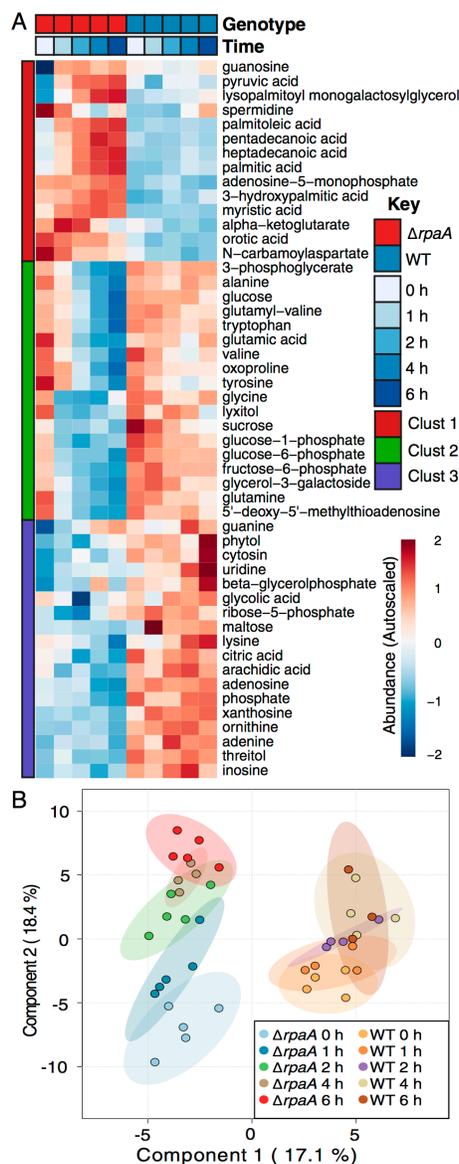


Fig. 2. Summary of metabolic changes in WT and $\Delta rpaA$. (A) Heatmap showing the autoscaled abundances of all metabolites where a significant difference was detected between WT and $\Delta rpaA$ over the time course as analyzed by two-way ANOVA and Tukey's honest significant difference ($n = 4$ for WT; $n = 5$ for $\Delta rpaA$; $P < 0.05$). Autoscaling represents a Z-score difference from the mean value of the metabolite across all time points. (B) Plot of PLS-DA components 1 and 2 for all metabolomics samples. Components 1 and 2 account for 35.5% of the variance in the dataset and are significant

between phycobilisomes, PSI, and PSII, as is true for *Synechocystis* sp. strain PCC 6803. Beginning 1–2 h after entering the dark and reaching a maximal change at 8 h after dark exposure, $\Delta rpaA$ had a significant decrease in absorbance at 630 nm and a further increase in absorbance at 440 nm, whereas WT showed no significant change in its absorbance spectrum (Fig. 1B and Fig. S14). The decrease in absorbance at 630 nm indicates a loss of phycobilisome-specific pigmentation, which is a well-characterized response to stress and macronutrient deprivation (31).

In parallel, separate samples were removed and plated under LL conditions to assess viable cell counts (colony-forming units) at each time point. Samples taken immediately before dark exposure (0 h) showed similar numbers of viable cells in the two strains (Fig. 1C). After 4 h of dark exposure, a large reproducible decrease in cell number was evident for the $\Delta rpaA$ strain, with no corresponding decrease for WT (Fig. 1C); by 8 h of darkness, only ~1% of $\Delta rpaA$ cells were viable (Fig. 1C). Optical density measurements of the sampled cultures also showed that the $\Delta rpaA$ strain did not resume growth during a following light period, even when transitioned back to LL growth conditions (Fig. S1B). These data indicate that $\Delta rpaA$ cells die soon after entering the dark. Together, the speed of the cell death response, the rapid changes in pigment absorbance, and an inability to regain viability in LL support an active mechanism that drives cell death in LD, as opposed to simple failure to thrive under LD conditions.

Temporal Metabolic Changes in the $\Delta rpaA$ Mutant. Previous data showing attenuated glycogen degradation in the $\Delta rpaA$ mutant (19) and RpaA transcriptional regulation of carbon catabolic pathways (20), and our observation of active nutrient deprivation-like bleaching (Fig. 1B), suggest that broad changes in central carbon metabolism likely occur in $\Delta rpaA$ cells after a light-to-dark transition. To characterize metabolic changes, we applied untargeted gas chromatography time-of-flight mass spectrometry (GC-TOF-MS) to samples collected from photobioreactors under LD (12-h light:12-h dark) growth conditions. Samples were collected directly before the dark onset (0 h) and at 1, 2, 4, and 6 h thereafter. A total of 114 known compounds were identified and measured (Dataset S1). Dramatic differences in primary carbon metabolites and metabolites that require NADPH for their biosynthesis were found between WT and $\Delta rpaA$ samples, suggesting that RpaA-mediated reductant production via the OPPP at night is critically important to keep metabolite levels stable. Additionally, the detection of stress-associated metabolites before dark exposure suggests that RpaA also plays a role in mitigating cellular stress during the day.

Elevated polyamines in $\Delta rpaA$ before entering dark indicate stress. Before entering the dark (0 h), polyamines were highly elevated in $\Delta rpaA$ with spermidine and putrescine showing a 48.8- and 4.5-fold increase relative to WT, respectively (Fig. 2A and Fig. S2). Correspondingly, ornithine, which is a known precursor for polyamines (32), was one-third less abundant in $\Delta rpaA$, suggesting mobilization of carbon toward polyamines. Accumulation of polyamines is a known general stress response in cyanobacteria (33), and the observed differences indicate that $\Delta rpaA$ cells may be stressed even before they enter the dark. Thus, in addition to nighttime functions, RpaA may have other, less understood, functions during the day.

WT maintains strict metabolic stability at night, which is lost in $\Delta rpaA$. Over the first half of the dark period, a large number of additional metabolite differences rapidly formed between WT and the $\Delta rpaA$ mutant. Using two-way ANOVA, we identified 50 compounds

predictors of class membership (Materials and Methods). Ellipses indicate the 95% confidence interval (CI) for each sample grouping ($n = 4$ for WT; $n = 5$ for $\Delta rpaA$; 114 metabolites per sample).

with significant differences in abundance patterns over the time course (Fig. 2A and Dataset S1). We also applied a multivariate modeling method, partial least-squares discriminate analysis (PLS-DA), to visualize and statistically test the overall similarity of sample groups, as well as determine which metabolites were more associated with differences in genotype or time (Fig. 2B).

One of the most striking observations from both analysis methods was the overall stability of metabolite levels in WT relative to $\Delta rpaA$ (Fig. 2). We had expected that WT would show significant metabolic changes downstream of the OPPP, because cyanobacteria have significant flux through glycogen degradation and the OPPP at night (19, 34, 35). However, WT maintained a stable metabolic profile, whereas $\Delta rpaA$ exhibited broad metabolic changes. Metabolites in the first and second clusters of the heatmap (Fig. 2A) showed large increases and decreases, respectively, over time in $\Delta rpaA$. In WT, the corresponding metabolites showed very gradual or no change in abundance (Fig. 2A). This effect was also pronounced in the plot of PLS-DA components 1 and 2 (Fig. 2B). Component 1 discriminated well between the two sample genotypes, and component 2 discriminated based on sampling time (Fig. 2B). The $\Delta rpaA$ mutant shows a clear separation across component 2, with temporally close samples more similar to each

other than temporally distant samples (Fig. 2B). This pattern was absent in WT samples, indicating that these samples were globally similar over the time course (Fig. 2B). Thus, ANOVA and PLS-DA both indicated that WT cells maintain a high level of metabolic stability in the early night period, which is lost in $\Delta rpaA$ mutants. **Metabolic changes in $\Delta rpaA$ indicate OPPP depression and NADPH deficit.** Metabolites connected to the OPPP and those that require NADPH for biosynthesis rapidly decrease in $\Delta rpaA$ after dark exposure. The OPPP-connected compounds sucrose, glucose-1-phosphate (G1P), glucose-6-phosphate (G6P), and fructose-6-phosphate (F6P) all showed a precipitous drop in abundance as soon as $\Delta rpaA$ entered the dark (Figs. 2A and 3A). WT also showed decreases in these metabolites over time, which was expected as glycogen stores are used, but the decrease was much more gradual (Fig. 2A). However, WT, unlike $\Delta rpaA$, showed an increase in F6P at the 1-h time point after dark onset. F6P is a known indicator of OPPP activation in cyanobacteria (28), and its stark decrease in $\Delta rpaA$ cells is consistent with highly attenuated OPPP activity (Fig. 2A).

A primary sink of NADPH in cyanobacteria is amino acid biosynthesis, and many amino acids showed strong decreases in the $\Delta rpaA$ mutant (Fig. 2A). The primary nitrogen donors to

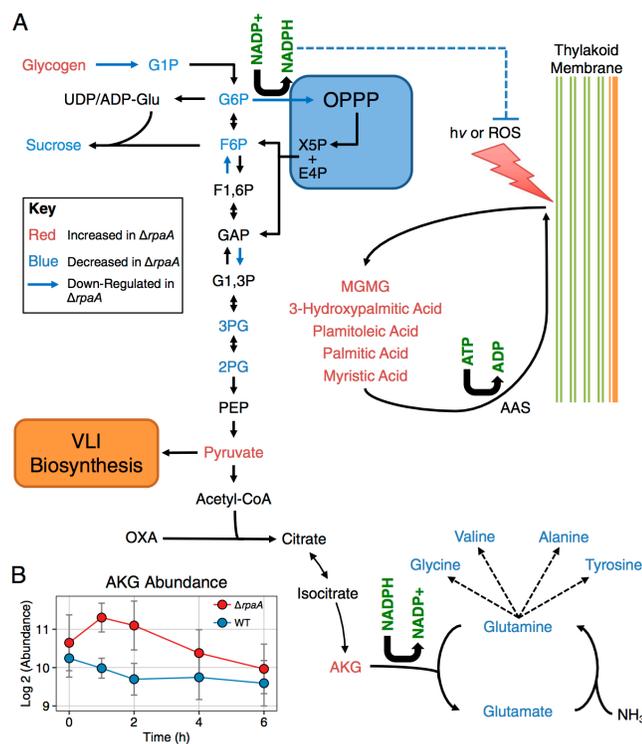


Fig. 3. Metabolic changes in the context of central carbon and nitrogen metabolism. (A) Diagram of relevant reactions in central carbon and nitrogen metabolism. Also included is a model of lipid recycling from photosynthetic membranes (right side). Metabolites are colored based on whether they were elevated (red text) or decreased (blue text) in $\Delta rpaA$ relative to WT at some point during the 6-h time course. Data for RpaA gene regulation were taken from Markson et al. (20). Dotted lines indicate that metabolites are linked, but details are not displayed. (B) Plot of AKG abundance in $\Delta rpaA$ and WT across the metabolic time course. Error bars indicate SD ($n = 4$ for WT; $n = 5$ for $\Delta rpaA$).

amino acid synthesis, glutamine and glutamate, dropped in abundance rapidly when $\Delta rpaA$ cells entered the dark (Figs. 2A and 3A). In turn, $\Delta rpaA$ showed a corresponding increase in α -ketoglutarate (AKG), the precursor metabolite for nitrogen assimilation (Fig. 3B). The conversion of AKG into glutamate in cyanobacteria is catalyzed by ferredoxin-dependent enzymes that use NADPH exclusively in their oxidation/reduction cycle (12, 13). A decrease in amino acid pools with a corresponding increase in AKG is consistent with an NADPH deficit in $\Delta rpaA$ cells. Additionally, elevated AKG can activate a nitrogen-starvation transcriptional response in cyanobacteria (36, 37), and although these cells were not nitrogen-starved, we found that elevated AKG levels in $\Delta rpaA$ were accompanied by this transcriptional response (Fig. S3 and *SI Text*).

Fatty acid accumulation in $\Delta rpaA$ likely causes damage and may result from redox imbalance. Intracellular accumulation of free fatty acids (FFAs) is uncommon in cyanobacteria, because they are activated and recycled into membrane lipids in an ATP-dependent reaction catalyzed by acyl-ACP synthetase (AAS) (38). However, in $\Delta rpaA$, we observed a large increase in the abundance of palmitic acid, palmitoleic acid, and myristic acid at the 4- to 6-h time points after dark exposure (Figs. 2A and 3A). The accumulation of intracellular FFAs in *S. elongatus* is generally toxic and directly causes damage to photosystem complexes, potentially exacerbating redox stress (39, 40). We saw that the accumulation of FFAs in $\Delta rpaA$ temporally coincided with the observed decrease in $\Delta rpaA$ cell viability (Figs. 1C and 2A), which is consistent with accumulation of these compounds contributing to the LD lethality phenotype.

Fatty acid recycling in *S. elongatus* increases dramatically under high light, and functional AAS is important to maintain cell viability under these conditions (40). In the $\Delta rpaA$ mutant, the presence of elevated lysopalmitoyl monogalactosylglycerol and 3-hydroxypalmitic acid indicated that active membrane remodeling and fatty acid recycling are taking place (Figs. 2A and 3A). Although the mechanism that drives membrane remodeling and subsequent lipid accumulation in cyanobacteria is still unclear, de novo synthesis of lipids would be unlikely in $\Delta rpaA$ cells that are reductant-poor. We hypothesize that regulation of membrane lipid turnover and FFA accumulation may be responsive to a change in cellular redox state. This hypothesis is consistent with membrane remodeling occurring under both high light (40) and in $\Delta rpaA$ cells that lack sufficient reductant (NADPH) to control cellular redox state in the dark. We posit that, although both WT and $\Delta rpaA$ may respond to oxidative stress by activation of lipid recycling, only the mutant reaches a triggering threshold of activation under moderate-light growth and does so during the night period, when AAS activity may be limited by ATP availability.

Interventions That Suppress the $\Delta rpaA$ LD Lethality Phenotype Support Reductant Imbalance as a Cause of Cell Death.

Suppression of $\Delta rpaA$ LD lethality by second-site mutagenesis. Older cultures of $\Delta rpaA$ mutants accumulate cells with the ability to grow under normally restrictive LD conditions (Fig. S4A). These clones still maintain fully segregated deletions at the *rpaA* locus; thus, it was surmised that they have accumulated compensatory changes at secondary genetic loci. To investigate the types of mutations that could suppress $\Delta rpaA$ LD lethality, we mutagenized freshly constructed, and still LD-sensitive, $\Delta rpaA$ mutant cells with ethyl methanesulfonate (EMS). Both EMS-exposed and unexposed $\Delta rpaA$ samples were then incubated under a restrictive LD growth condition. Hundreds of colonies with a wide degree of coloration and morphology appeared exclusively on the plate containing EMS-exposed $\Delta rpaA$ cells (Fig. S4B). We isolated 20 colonies, confirmed that all maintained fully segregated deletions at the *rpaA* locus (Fig. S4C), and performed full genome resequencing on each. Comparison of the mutagenized

genomes to both a WT control and the $\Delta rpaA$ parent strain revealed a total of 63 single nucleotide changes across all strains with an average of 3.15 ± 1.2 new mutations per strain. Subsequently, we filtered the mutations (*Materials and Methods*) and identified a subset of 56 that we categorized as “high confidence for biological effect” (Dataset S2).

Of the mutations, 47% occurred in genes that code for metabolic enzymes (Dataset S2). Pathways that synthesize and use amino acids were significantly enriched in these mutations, including valine, leucine, and isoleucine (VLI) biosynthesis [false discovery rate (FDR) = $2.8e-4$]; aminoacyl-tRNA biosynthesis (FDR = 0.01); and global amino acid biosynthesis (FDR = $2.8e-4$).

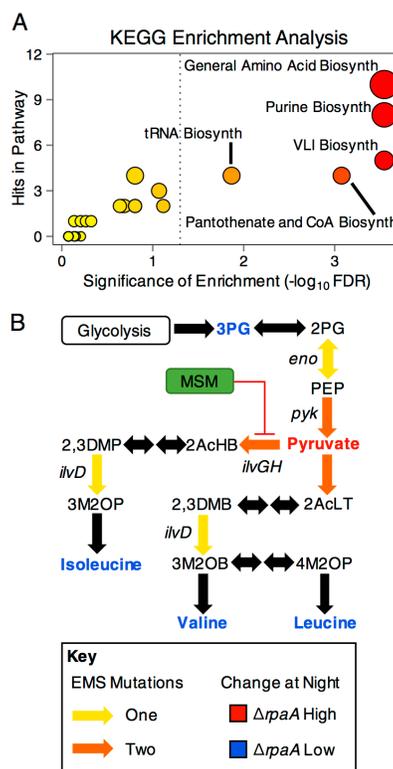


Fig. 4. Summary of enriched KEGG functional categories identified by suppressor mutations and metabolic pathway topology represented by mutated genes. (A) Plot of KEGG metabolic categories that were enriched in the gene set of suppressor mutations. The x axis indicates the number of times a specific KEGG pathway was matched to genes in the set; dots scale from small to large with increasing number of matches, and color of dots scales from yellow to red with increasing significance. Significance was calculated by using the binomial distribution, corrected for multiple testing using the method of Benjamini–Hochberg, and significance cutoff is indicated with a gray dotted line (FDR < 0.05). (B) Subpathway diagram of VLI biosynthesis indicating locations of $\Delta rpaA$ suppressor mutations and average abundance of compounds in the $\Delta rpaA$ strain relative to WT over the metabolomics time course. Genes were named for reactions where a suppressing mutation was identified, and colors are detailed in the key.

(Fig. 4A and Dataset S2). Additionally, multiple independent strains carried mutations affecting the same biochemical step in some of these pathways (Fig. 4B). These mutations affecting the same biochemical step included independent point mutations in *ilvG* and *ilvH* (acetolactate synthase catalytic and regulatory subunits), the rate-limiting enzyme complex in VLI biosynthesis (41), and two in *pyk* (pyruvate kinase), which produces pyruvate, the substrate of *ilvGH* (Fig. 4B) (42). Although mutations were found that could affect a number of metabolic pathways (Dataset S2), the high concentration of mutations in pathways that produce and consume amino acids was interesting, considering the large decreases observed for these compounds in $\Delta rpaA$ at night (Figs. 2A and 4B).

Blocking VLI biosynthesis suppresses LD lethality in $\Delta rpaA$. Because each suppressed $\Delta rpaA$ strain carries multiple EMS generated mutations, we targeted the VLI biosynthetic pathway to test whether manipulation of a single pathway identified through EMS mutagenesis is sufficient for the suppression of LD lethality in $\Delta rpaA$. VLI biosynthesis was chosen because a known herbicide, metsulfuron methyl (MSM), specifically and potently inhibits IlvGH (6, 43), and two suppressor mutations mapped to genes coding for this complex (Fig. 4B). Suppression of LD sensitivity with MSM, if successful, would also imply that the point mutations identified have a negative impact on complex activity.

Treatment of $\Delta rpaA$ with MSM under LD growth suppressed cell death of $\Delta rpaA$ on solid (100 nM MSM) and in liquid (25 μ M MSM) media (Fig. 5A and Fig. S5A). Additionally, although all MSM-treated cultures showed a change in pigmentation (Fig.

5A), we confirmed that when $\Delta rpaA$ entered the dark, treated cells no longer showed phycobilisome-specific pigment bleaching or the activation of a nitrogen-deprivation transcriptional response (Fig. 5B and Fig. S3). Therefore, inhibition of the VLI biosynthetic pathway is sufficient to rescue $\Delta rpaA$ cells from LD lethality, as well as inhibit specific phenotypes associated with cell death in the dark. Additionally, the fact that MSM exerts its suppressive effect by *ilvGH* inhibition suggests that the point mutations identified in *ilvG* and *ilvH* reduce native enzyme complex activity.

Suppression of VLI biosynthesis lowers phycobilisome content during the daytime. $\Delta rpaA$ strains treated with MSM and those that carry VLI pathway mutations had a strong yellow color (Fig. 5A and Fig. S5B). Absorbance scans taken before genomic DNA extraction for sequencing revealed that these mutants had a decreased phycobilisome-to-chlorophyll ratio (630/680 nm) relative to WT (Table S1). To investigate whether pigmentation changes occur during repression of VLI biosynthesis, we incubated WT and $\Delta rpaA$ with 25 μ M MSM in the light for 12 h and compared whole-cell absorbance spectra of treated and untreated cells. Cultures exposed to MSM visibly appeared more yellow. WT treated with MSM had decreases in all three pigment absorption maxima: 440, 630, and 680 nm, with the phycobilisome peak at 630 nm showing the largest difference (Fig. 5C). MSM-treated $\Delta rpaA$ samples also had a significant reduction in absorbance at the 630-nm phycobilisome absorbance peak (Fig. 5C). These results show that treatment with MSM reduces phycobilisome content in both WT and $\Delta rpaA$. Because phycobilisomes are the

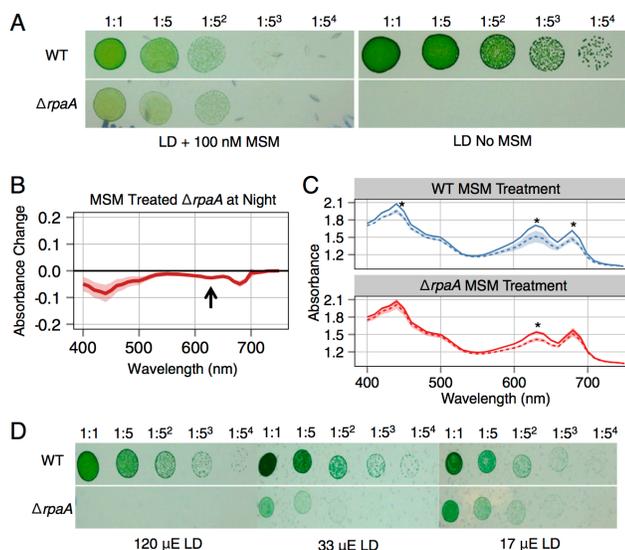


Fig. 5. Summary of data on MSM- and light intensity-mediated suppression of the $\Delta rpaA$ LD lethality phenotype. (A) Representative photo of dilution series of WT and $\Delta rpaA$ cells treated (Left) and not treated (Right) with 100 nM MSM ($n = 6$ biological replications of experiment). Pictured samples were grown in an LD cycle with a light intensity of 120 μ E $m^{-2} \cdot s^{-1}$. (B) Difference in absorbance of $\Delta rpaA$ cells treated with 25 μ M MSM between 0 and 8 h after dark exposure. Shaded region indicates SD of mean, and black arrow points to absorbance at 630 nm highlighting no significant change. Significance was calculated by using Student's *t* test for absorbance at 440, 630, and 680 nm, with no significant change observed ($n = 3$). (C) Mean absorbance values of WT and $\Delta rpaA$ untreated (solid line) and treated (dotted line) with 25 μ M MSM after 12 h in the light. Shaded area indicates SEM. MSM-exposed cells show significantly lower absorption values at 440, 630, and 680 nm, as calculated by a one-sided Student's *t* test ($n = 3$ for MSM treated samples and $n = 4$ for untreated samples). * $P < 0.05$. (D) Representative photo of dilution series of WT and $\Delta rpaA$ cells grown in an LD cycle with decreasing daytime light intensity (indicated below each image; $n = 2$).

primary light-collecting pigment proteins in *S. elongatus*, the MSM-mediated decrease at 630 nm represents a significant change in the ability of cells to collect light and likely alters both photosynthetic output and cellular redox state (44, 45).

Decreasing light intensity suppresses $\Delta rpaA$ LD lethality. Given that reducing the light-collecting ability of cells is one effect of MSM treatment, we tested whether modulating growth light intensity has an effect on the $\Delta rpaA$ LD lethality phenotype. WT and an $\Delta rpaA$ mutant were serially diluted, plated, and incubated under three light intensities in LL and LD. At the highest light intensity tested ($120 \mu\text{E}\cdot\text{M}^{-2}\cdot\text{s}^{-1}$), we observed the expected $\Delta rpaA$ lethality phenotype under LD conditions (Fig. 5D). However, at an intermediate light intensity ($33 \mu\text{E}\cdot\text{M}^{-2}\cdot\text{s}^{-1}$), the $\Delta rpaA$ mutant had slightly improved growth in LD (Fig. 5D), and at the lowest light intensity ($17 \mu\text{E}\cdot\text{M}^{-2}\cdot\text{s}^{-1}$), $\Delta rpaA$ cells grew to almost WT levels (Fig. 5D). Thus, a low-light LD cycle creates a permissive condition where $\Delta rpaA$ cells can survive. These results show that light intensity during the day period is at least one factor that contributes to the $\Delta rpaA$ LD lethality phenotype. Thus, the reduction of phycobilisome pigment associated with MSM treatment may partially contribute to its mechanism for suppressing LD lethality via reducing absorbed light energy during the day.

Redox Stress Is Associated with Cell Death in the $\Delta rpaA$ Mutant. Although $\Delta rpaA$ strains can tolerate high light intensity under LL conditions (Fig. S6A), we observed that $\Delta rpaA$ cells had high levels of metabolites, indicating cellular stress before entering the dark (Fig. 2A and Fig. S2). Additionally, the protective effects of MSM treatment (Fig. 5A) and decreased light intensity (Fig. 5D) suggest that photosynthetically generated reactive oxygen species (ROS) may act as a destructive agent that drives $\Delta rpaA$ LD lethality. *S. elongatus* maintains a strict cellular redox balance and controls ROS using multiple systems, including modulation of phycobilisome abundance, glutathione redox control, and enzymatic ROS scavenging (46, 47). Glutathione biosynthesis is particularly important for modulating redox state and counteracting ROS in cyanobacteria (46, 48, 49). Regeneration of reduced glutathione requires NADPH, and our data suggest that $\Delta rpaA$ cells are NADPH-limited at night (Figs. 2A and 3A). Additionally, metabolites of the glutathione biosynthetic pathway, including oxoproline, glycine, glutamate, and glutamylvaline all showed large decreases in $\Delta rpaA$ after dark transition (Fig. 2A). To determine the influence of redox stress on $\Delta rpaA$ LD phenotypes, we tracked total ROS in WT and $\Delta rpaA$ over a 24-h LD cycle (at high light intensity) using the fluorescent marker 2',7'-dichlorodihydrofluorescein diacetate (H_2DCFDA) (50, 51). Samples were taken every 2 h during the 12-h day period and every hour during the 12-h night period. Additionally, we assessed whether MSM addition impacted the ROS detected in WT and $\Delta rpaA$ cells over the 24-h LD cycle.

ROS levels were similar for all strains at the start of the experiment (Fig. 6). ROS increased gradually in all strains through the first 6 h of the day, then rapidly increased in the $\Delta rpaA$ mutant exclusively, and by the end of the light period (12 h), were 3.5-fold higher than in the WT or MSM-treated $\Delta rpaA$ cultures (Fig. 6). Upon entering the dark, all strains showed a rapid drop in ROS within the first 2 h (Fig. 6). ROS continued to drop in WT and the MSM-treated $\Delta rpaA$ mutant throughout the night, reaching a level similar to the start of the experiment. In contrast, the untreated $\Delta rpaA$ mutant maintained high static ROS levels after the first 2–3 h of darkness (Fig. 6). MSM treatment had no significant effect on WT ROS levels (Fig. S6B).

Comparison of ROS levels with cell viability (Fig. 1C) showed that cell death begins in $\Delta rpaA$ cells around the time ROS levels stabilize. Overall, the $\Delta rpaA$ mutant accumulates high levels of ROS under the restrictive light condition and has trouble clearing ROS over the night period. However, treatment with MSM alleviates the elevated ROS phenotype. We propose that the rapid

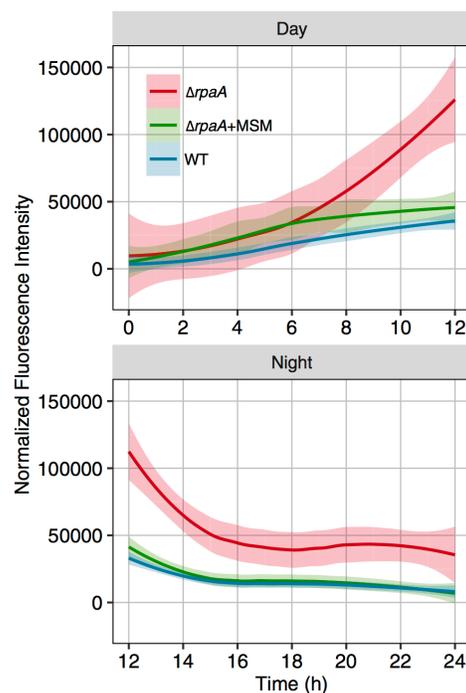


Fig. 6. Plot of H_2DCFDA fluorescence over a 24-h LD cycle indicating total cellular ROS in WT, $\Delta rpaA$, and $\Delta rpaA$ treated with $25 \mu\text{M}$ MSM. Cells were grown at a light intensity in bioreactors empirically determined to support $\Delta rpaA$ growth in LD, and the experiment began after a 12-h dark period for all cells. Curves shown are best-fit lines calculated using LOESS regression to all data points in a given sample; the gray shaded area indicates the 95% CI of the regression line ($n = 21$ data points for day samples; $n = 42$ data points for night samples). Day and night from the same experiment were split to more effectively fit regressions. Places where the CI does not overlap indicate a statistically significant difference in the model.

lethal effect of darkness on a $\Delta rpaA$ mutant results from a failure to clear ROS that accumulates late in the day period. This hypothesis is consistent with the observation that cell death begins in $\Delta rpaA$ cells when ROS levels stop decreasing at night and that the $\Delta rpaA$ metabolome indicates an NADPH deficit. Detoxification of ROS at night places an additional strain on the reductant pool, which likely exacerbates the metabolic imbalances we observe in this strain.

Discussion

This study highlights three functions of RpaA in WT cells that are critical for survival under LD growth conditions: (i) RpaA has daytime functions that are important for limiting ROS buildup; (ii) RpaA activates genes that encode the enzymes of the OPPP, which are critical for NADPH production in the absence of photosynthesis, and inability to generate sufficient reductant at night results in failure to detoxify ROS accumulated during the day and cell death; and (iii) overall, RpaA exerts a strong influence over the control of redox balance, which seems to be critical for maintaining the inherent metabolic stability of WT cells at night. These results are consistent with recent findings in

Synechocystis PCC 6803 that show proper NADPH balance is important for appropriate diurnal regulation of metabolic processes (52), as well as data indicating that the fitness advantage conferred by the *S. elongatus* circadian clock is more likely due to the ability to anticipate a coming dark period rather than a morning period (53).

One of the surprising findings is that events during the day ultimately affect death of the $\Delta rpaA$ mutant in the dark. This outcome was not anticipated because the level of light that is lethal under LD growth was well tolerated by $\Delta rpaA$ cells when they are grown under LL (Fig. S6A). Thus, it was originally hypothesized that events exclusively occurring at night were driving cell death. However, the original identification of RpaA derived from its effect on the association of phycobilisomes with photosystem reaction centers in *Synechocystis* sp. strain PCC 6803. In that strain, *rpaA*-null mutants have an increased efficiency of energy transfer to PSII relative to PSI (22). Our pigment absorbance data collected at 0 h support the hypothesis that RpaA also affects photosynthetic parameters in *S. elongatus* (Fig. 1A). If energy transfer to PSII relative to PSI is also increased in $\Delta rpaA$ *S. elongatus* cells, this alteration could both serve as a source of redox stress via excess excitation energy in PSII as well as a decrease the cells' ability to produce NADPH via PSI-driven reduction. Regardless of the source, the ROS and metabolomics data both indicate that once cells reach the light-to-dark transition, the $\Delta rpaA$ mutant is already under a great deal of cellular redox stress (Figs. 1A, 2A, and 6 and Fig. S2) (33, 46).

We propose that high redox stress generated from photosynthetic activity is a critical component of the $\Delta rpaA$ LD lethality phenotype. Manipulations that reduce light energy absorbed by cells, including reduction of growth light intensity (Fig. 5D) and treatment with MSM (Fig. 5A), rescue $\Delta rpaA$ from death under LD conditions. These results are consistent with overstimulation of photosynthetic pathways as a source of ROS. Although we focused on the EMS suppressor mutations that affect the VLI biosynthetic pathway, via treatment with MSM, strains that carry other mutations in amino acid and aminoacyl-tRNA biosynthesis exhibited similarly depressed phycobilisome absorbance (Fig. S5B and Table S1). As some of the most abundant proteins in cyanobacteria (54), phycobilisome levels are affected significantly by amino acid limitation. Cells with reduced phycobilisome content would collect less light energy and generate less ROS during the day period. Additionally, amino acid biosynthetic pathways consume large amounts of NADPH, and partially blocking these pathways may serve to preserve the limited NADPH pools present in $\Delta rpaA$ at night, which are needed for ROS-scavenging pathways.

The viability of the $\Delta rpaA$ mutant in LL (Fig. S6A), at a light intensity that generates high ROS, suggests that NADPH produced via photosynthesis can drive the ROS detoxification mechanisms necessary to maintain viability while $\Delta rpaA$ cells are in the light; however, when the $\Delta rpaA$ mutant enters the dark under redox stress, it lacks a source of NADPH because of an attenuated ability to degrade glycogen and activate the OPPP (19, 20). Consistent with an inability to activate the OPPP in the dark, $\Delta rpaA$ cells rapidly deplete soluble sugars connected to this pathway (Fig. 2A). Because glycolysis is still active in $\Delta rpaA$ and the OPPP functions as a cycle at night, recycling its inputs (34, 35), we saw a significant increase only in metabolites downstream of glycolysis in the $\Delta rpaA$ mutant, including pyruvate and AKG (Figs. 2A and 3A). Although glycolysis can produce some NADH as reducing power, NADH is a poor electron source for ROS-detoxifying processes in *S. elongatus* (46). AKG buildup normally results in its conversion to glutamine and glutamate via nitrogen assimilation (12, 13). The rapid depletion of glutamine, glutamate, and many other amino acid species in the $\Delta rpaA$ mutant, concurrent with AKG elevation (Figs. 2A and 3B), is consistent with an NADPH deficit that precludes cells from performing

AKG to amino acid biosynthesis. Additionally, the activation of a nitrogen-deprivation transcriptional response (Fig. S3) and phycobilisome degradation (Fig. 1B) indicate that $\Delta rpaA$ cells accumulate AKG to a level that is perceived as C/N imbalance. The EMS mutations in amino acid biosynthetic and utilization pathways slow the mobilization of carbon toward amino acids and lower NADPH consumption, allowing $\Delta rpaA$ to more easily achieve homeostasis with limited NADPH pools. This hypothesis is consistent with the fact that $\Delta rpaA$ cells treated with MSM no longer show nitrogen-deprivation transcriptional or bleaching responses at night (Fig. 5B and Fig. S3).

Protein redox modifications drive important metabolic shifts in cyanobacteria and plant chloroplasts (10, 28, 55, 56). Recent work has shown that redox modifications are pervasive across all metabolic pathways in cyanobacteria and that LD transitions drive global changes in the oxidation state of redox-modified proteins (8, 9, 49). Because de novo transcription is limited to the early night period in *S. elongatus* (57), redox modifications on metabolic enzymes likely play a major role in modulating enzymatic activity and dictating metabolic flux over the dark period. Some of the metabolic changes in the $\Delta rpaA$ mutant after the termination of ROS detoxification may be driven by an inability to further modulate the protein redox state. Specifically, there is evidence that accumulation of lipids can be driven by redox changes and may contribute directly to LD lethality (40, 58).

Lipids are particularly sensitive to oxidative stress, and their turnover is important because oxidized lipid species can further perpetuate oxidative damage (59). Activation of lipid recycling by high light in *S. elongatus* is consistent with a redox-driven mechanism to signal this process. The increase of lysopalmitoyl monogalactosylglycerol and 3-hydroxypalmitic acid in $\Delta rpaA$ indicates that a lipid-recycling response has been activated at night (43). We propose that ROS detoxification terminates in the $\Delta rpaA$ mutant when the limited NADPH pool is exhausted, and subsequent redox stress activates the membrane recycling process (46). We propose that the $\Delta rpaA$ mutant, but not WT, has reached an ROS threshold that would trigger lipid recycling by the end of the day. Although both WT and $\Delta rpaA$ cells should properly express the AAS protein required to activate lipids for recycling (20), AAS is ATP-dependent, and its functionality at night in cyanobacteria would likely be impaired (39). Thus, the activation of lipid recycling in $\Delta rpaA$ occurs when ATP levels are dropping in the dark, and AAS may be unable to deal with FAA load. Indeed, we observed bleaching (Fig. 1B), AKG elevation (Figs. 2A and 3B), fatty acid accumulation (Fig. 2A), and the start of cell death (Fig. 1C) around the time that ROS detoxification terminated (Fig. 6). Thus, the primary mechanisms directly driving cell death likely occur between the 2- to 6-h period after entering the dark, and these mechanisms are likely the result of a strong redox imbalance.

The daytime functions of RpaA are unexpected and should be further explored. In particular, the pigmentation changes during LL growth of the $\Delta rpaA$ mutant (Fig. 1A) suggest that core photosynthetic parameters are altered. Metrics such as photosystem efficiency, capacity, and oxygen evolution in *S. elongatus* have not been explored in the context of circadian rhythms. Additionally, it is interesting that ROS levels in the $\Delta rpaA$ mutant consistently increase at ~6 h after a transition from darkness into light (Fig. 6), because this mutant lacks a clock output mechanism and does not exhibit transcriptional rhythms (20, 21). This timing may represent a point where ROS scavenging resources are naturally exhausted, or it may indicate that other known rhythmic processes, such as 24-h peroxiredoxin rhythms, may be important for modulating ROS (18, 60). The participation of peroxiredoxins in ROS modulation would be consistent with their roles in oxidative stress-mediated signaling (61). This observation hints that Kai-mediated and other rhythmic processes may interact to control the cellular redox state. Overall, the integration of light conditions,

circadian rhythms, and the cellular redox state in the control of cyanobacterial metabolism will be of crucial importance to advance the engineering and understanding of cyanobacteria growing under natural diurnal conditions.

Materials and Methods

The full description of experimental techniques is provided in *SI Materials and Methods*.

Cyanobacterial Strains, Media, and Culture Conditions. All $\Delta rpaA$ mutants were constructed by transformation in a WT *S. elongatus* background with plasmid pAM4420 (25) and were validated by PCR. For all experiments, precultures were first prepared in 100 mL of fresh BG-11 medium as in Diamond et al. (19).

For metabolomics experiments, precultures were used to inoculate Phenometrics ePBR photobioreactors (Version 1.1; Phenometrics Inc.) at an initial density of $OD_{750} = 0.1$ in 400 mL of BG-11 medium without antibiotics. Temperature was maintained at 30 °C; filtered (0.2 μm) air was sparged at a rate of 50 mL/min; and light intensity was either 150 or 500 $\mu\text{E}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ provided from the top of the culture while lights were on. After inoculation, cultures were grown at a constant light intensity of 150 $\mu\text{E}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ until $OD_{750} = 0.3$, then maintained turbidostatically at this density for the duration of the experiment. In the metabolomics experiment, WT circadian rhythms were entrained by growth in a 12:12 LD at a light intensity of 150 $\mu\text{E}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ cycle for 1 d and subsequently at a light intensity of 500 $\mu\text{E}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ for 2 d before release into experimental conditions and sampling. The $\Delta rpaA$ strains were maintained in constant light at the same intensities as for the WT strain before the sampling procedure (Fig. S7).

For absorbance scanning, viable cell counts, quantitative reverse transcription-polymerase chain reaction (qRT-PCR), MSM-treatment absorbance measurements, and oxidative stress measurements precultures were used to inoculate the photobioreactors at an initial density of $OD_{750} = 0.2$ in 400 mL of BG-11 medium without antibiotics. Temperature, airflow rate, and light intensity settings were the same as above. For these experiments, both WT and $\Delta rpaA$ mutants were maintained at a constant light intensity of 150 $\mu\text{E}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ for 1 d. Subsequently, both strains were subjected to growth in a 12:12 LD cycle at a light intensity of 150 $\mu\text{E}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ for 2 d. Light intensity was then increased to 500 $\mu\text{E}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ over the final 12:12 LD period, during which sampling took place.

For viable cell plating, 200 μL of the indicated sample was serially diluted 1:5 in fresh BG-11 medium without antibiotics five times. For LD sensitivity testing, samples were first all diluted to an $OD_{750} = 0.2$, and the same dilution scheme was then followed. Subsequently, 4 μL of each sample was spotted onto solid BG-11 plates without antibiotics. Plates were incubated at 30 °C and 150 $\mu\text{E}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ constant light for 5–6 d. For LD sensitivity testing, samples were plated in duplicate, with one set incubated in constant light for 5–6 d, and a second set under a 12:12 LD cycle for 6–8 d.

Metabolomics and Data Analysis. Strains for metabolomics were grown in Photobioreactors and sampled ($n = 5$ at each time point and for each genotype) as described above (Fig. S7). Metabolite extraction and GC-TOF-MS were conducted by the West Coast Metabolomics Center (WCMC) at the University of California, Davis identically to the methods used in Diamond et al. (19) and Fiehn et al. (62, 63).

Raw metabolite abundance data for known metabolites (Dataset S1) were analyzed by using a combination of the online analysis platform MetaboAnalyst (Version 3.0) (64) and the statistical package R (65). Principal component analysis (PCA) was applied to \log_2 -normalized and autoscaled data to detect outlying samples. Based on PCA, replicate A of the WT sample was removed from the dataset before statistical analysis (Fig. S8A). In addition, WT replicate C time T6 and $\Delta rpaA$ replicate C time T4 were removed before analysis because of problems during sample extraction reported by the WCMC. Statistical analysis of metabolomics data are detailed in *SI Materials and Methods*. Further details and statistical methods are provided in *SI Text*.

Mutagenesis and Identification of $\Delta rpaA$ Suppressing Mutations. EMS mutagenesis of fresh $\Delta rpaA$ mutant cyanobacterial strains was carried out as in Kondo et al. (66). Absorbance scans were taken of all cultures, as detailed in *SI Materials and Methods*, and genomic DNA was extracted by using standard methods (67). Before sequencing genomic DNA, the disruption of *rpaA* in all strains was verified by PCR (Fig. S4C). Genomic library preparation for Illumina short-read sequencing was performed by using the NEBNext DNA library preparation kit (NEB, catalog no. E60405/L) with NEXTflex barcoded cDNA adaptors (BIOO Scientific, catalog no. 514104). Samples were run on an Illumina HiSeq 2500 DNA sequencer at the University of California, Berkeley Q83 Genomics Sequencing Laboratory. Sequencing runs resulted in 50-bp reads with a median coverage depth of 45.7x per sample over the *S. elongatus* genome. Reads from all sequencing methods were mapped against the *S. elongatus* genome (GenBank accession no. NC_007604) and the large plasmid pANL (GenBank accession no. AF441790), and polymorphisms were called by using the program breseq (68). Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis of mutated genes was conducted by using a custom-written R script and the metabolic categories in Dataset S2. Statistical overrepresentation was determined by using the binomial test, and *P* values were corrected by using the method of Benjamini and Hochberg (69).

ACKNOWLEDGMENTS. We thank the members of the Ralph Greenspan Laboratory and the Berkeley Q83 Genomics Sequencing Laboratory for their assistance with library preparation and sequencing; BATJ, Inc. for their collaboration in sequencing additional suppressor mutants; Profs. Stephen Mayfield and Terry Hwa for particularly useful guidance on metabolomics and microbial growth kinetics, respectively; Dr. Mark Paddock for helpful discussion that improved the manuscript; and Anish Pal for assistance with sample collection and strain maintenance. This work was supported by National Science Foundation Grant MCB1244108. S.D. and B.E.R. were supported in part by National Institutes of Health and Molecular Genetics Training Grant T32GM007240.

- Oliver JW, Atsumi S (2014) Metabolic design for cyanobacterial chemical synthesis. *Photosynth Res* 120(3):249–261.
- Ducat DC, Way JC, Silver PA (2011) Engineering cyanobacteria to generate high-value products. *Trends Biotechnol* 29(2):95–103.
- Bryant DA (2003) The beauty in small things revealed. *Proc Natl Acad Sci USA* 100(17):9647–9649.
- Zhang S, Bryant DA (2015) Biochemical validation of the glyoxylate cycle in the cyanobacterium *Chlorogloeopsis fritschii* strain PCC 9212. *J Biol Chem* 290(22):14019–14030.
- Zhang S, Bryant DA (2011) The tricarboxylic acid cycle in cyanobacteria. *Science* 334(6062):1551–1553.
- Kouhen OM-E, Joset F (2002) Biosynthesis of the branched-chain amino acids in the cyanobacterium *Synechocystis* PCC6803: Existence of compensatory pathways. *Curr Microbiol* 45(2):94–98.
- Rubin BE, et al. (2015) The essential gene set of a photosynthetic organism. *Proc Natl Acad Sci USA* 112(48):E6634–E6643.
- Guo J, et al. (2014) Proteome-wide light/dark modulation of thiol oxidation in cyanobacteria revealed by quantitative site-specific redox proteomics. *Mol Cell Proteomics* 13(12):3270–3285.
- Ansong C, Sadler NC, Hill EA, Lewis MP (2014) Characterization of protein redox dynamics induced during light-to-dark transitions and nutrient limitation in cyanobacteria. *Front Microbiol* 5:325.
- Lindahl M, Kieselbach T (2009) Disulphide proteomes and interactions with thio-redoxin on the track towards understanding redox regulation in chloroplasts and cyanobacteria. *J Proteomics* 72(3):416–438.
- Schürmann P (2003) Redox signaling in the chloroplast: The ferredoxin/thioredoxin system. *Antioxid Redox Signal* 5(1):69–78.
- Ohashi Y, et al. (2011) Regulation of nitrate assimilation in cyanobacteria. *J Exp Bot* 62(4):1411–1424.
- Muro-Pastor MI, Reyes JC, Florencio FJ (2005) Ammonium assimilation in cyanobacteria. *Photosynth Res* 83(2):135–150.
- Vijayan V, Zuzow R, O'Shea EK (2009) Oscillations in supercoiling drive circadian gene expression in cyanobacteria. *Proc Natl Acad Sci USA* 106(52):22564–22568.
- Mackey SR, Golden SS, Ditty JL (2011) *The Itty-Bitty Time Machine*. Advances in Genetics (Elsevier, New York) pp 13–53.
- Cohen SE, Golden SS (2015) Circadian rhythms in cyanobacteria. *Microbiol Mol Biol Rev* 79(4):373–385.
- Shultzaberger RK, Boyd JS, Diamond S, Greenspan RJ, Golden SS (2015) Giving time purpose: The *Synechococcus elongatus* Clock in a broader network context. *Annu Rev Genet* 49:485–505.
- Edgar RS, et al. (2012) Peroxiredoxins are conserved markers of circadian rhythms. *Nature* 485(7399):459–464.
- Diamond S, Jun D, Rubin BE, Golden SS (2015) The circadian oscillator in *Synechococcus elongatus* controls metabolite partitioning during diurnal growth. *Proc Natl Acad Sci USA* 112(15):E1916–E1925.
- Markson JS, Piechura JR, Puszynska AM, O'Shea EK (2013) Circadian control of global gene expression by the cyanobacterial master regulator RpaA. *Cell* 155(6):1396–1408.
- Takai N, et al. (2006) A KaiC-associating SasA-RpaA two-component regulatory system as a major circadian timing mediator in cyanobacteria. *Proc Natl Acad Sci USA* 103(32):12109–12114.
- Ashby MK, Mullineaux CW (1999) Cyanobacterial *ycf27* gene products regulate energy transfer from phycobilisomes to photosystems I and II. *FEMS Microbiol Lett* 181(2):253–260.

23. Paddock ML, Boyd JS, Adin DM, Golden SS (2013) Active output state of the *Synechococcus* Kai circadian oscillator. *Proc Natl Acad Sci USA* 110(40):E3849–E3857.
24. Iijima H, et al. (2015) Changes in primary metabolism under light and dark conditions in response to overproduction of a response regulator RpaA in the unicellular cyanobacterium *Synechocystis* sp. PCC 6803. *Front Microbiol* 6:888.
25. Boyd JS, Bordowitz JR, Bree AC, Golden SS (2013) An allele of the *crm* gene blocks cyanobacterial circadian rhythms. *Proc Natl Acad Sci USA* 110(34):13950–13955.
26. Yang C, Hua Q, Shimizu K (2002) Integration of the information from gene expression and metabolic fluxes for the analysis of the regulatory mechanisms in *Synechocystis*. *Appl Microbiol Biotechnol* 58(6):813–822.
27. Osanai T, Azuma M, Tanaka K (2007) Sugar catabolism regulated by light- and nitrogen-status in the cyanobacterium *Synechocystis* sp. PCC 6803. *Photochem Photobiol Sci* 6(5):508–514.
28. Tamoi M, Miyazaki T, Fukamizo T, Shigeoka S (2005) The Calvin cycle in cyanobacteria is regulated by CP12 via the NAD(H)/NADP(H) ratio under light/dark conditions. *Plant J* 42(4):504–513.
29. Wedel N, Soll J (1998) Evolutionary conserved light regulation of Calvin cycle activity by NADPH-mediated reversible phosphoribulokinase/CP12/ glyceraldehyde-3-phosphate dehydrogenase complex dissociation. *Proc Natl Acad Sci USA* 95(16):9699–9704.
30. Waldbauer JR, Rodrigue S, Coleman ML, Chisholm SW (2012) Transcriptome and proteome dynamics of a light-dark synchronized bacterial cell cycle. *PLoS One* 7(8):e43432.
31. Collier JL, Grossman AR (1992) Chlorosis induced by nutrient deprivation in *Synechococcus* sp. strain PCC 7942: Not all bleaching is the same. *J Bacteriol* 174(14):4718–4726.
32. Zhu X, Li Q, Yin C, Fang X, Xu X (2015) Role of spermidine in overwintering of cyanobacteria. *J Bacteriol* 197(14):2325–2334.
33. Incharoensakdi A, Jantaro S, Raksait W, Mäenpää P (2010) Polyamines in cyanobacteria: Biosynthesis, transport and abiotic stress response. *Current Research Topics in Applied Microbiology and Microbial Biotechnology*, ed Méndez-Villas A (Formatex, Badajoz, Spain), pp 23–32.
34. Young JD, Shastri AA, Stephanopoulos G, Morgan JA (2011) Mapping photoautotrophic metabolism with isotopically nonstationary ¹³C flux analysis. *Metab Eng* 13(6):656–665.
35. Nakajima T, et al. (2014) Integrated metabolic flux and omics analysis of *Synechocystis* sp. PCC 6803 under mixotrophic and photoheterotrophic conditions. *Plant Cell Physiol* 55(9):1605–1612.
36. Klotz A, Reinhold E, Doello S, Forchhammer K (2015) Nitrogen starvation acclimation in *Synechococcus elongatus*: Redox control and the role of nitrate reduction as an electron sink. *Life (Basel)* 5(1):888–904.
37. Flores E, Herrero A (2005) Nitrogen assimilation and nitrogen control in cyanobacteria. *Biochem Soc Trans* 33(Pt 1):164–167.
38. Kaczmarek D, Fulda M (2010) Fatty acid activation in cyanobacteria mediated by acyl-acyl carrier protein synthetase enables fatty acid recycling. *Plant Physiol* 152(3):1598–1610.
39. Kato A, et al. (2016) Modulation of the balance of fatty acid production and secretion is crucial for enhancement of growth and productivity of the engineered mutant of the cyanobacterium *Synechococcus elongatus*. *Biotechnol Biofuels* 9:91.
40. Takatani N, et al. (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(8):1608–1615.
41. Shen J, DiTommaso A, Shen M, Lu W, Li Z (2009) Molecular basis for differential metabolic responses to monosulfuron in three nitrogen-fixing cyanobacteria. *Weed Sci* 57(March 2009):133–141.
42. Kanehisa M, Goto S (2000) KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 28(1):27–30.
43. Friedberg D, Seiffers J (1988) Sulfonyleurea-resistant mutants and natural tolerance of cyanobacteria. *Arch Microbiol* 150(3):278–281.
44. Page LE, Liberton M, Pakrasi HB (2012) Phycobilisome antenna truncation reduces photoautotrophic productivity in *Synechocystis* sp. PCC 6803, a cyanobacterium. *Appl Environ Microbiol* 165(2):705–714.
45. Grossman AR, Schaefer MR, Chiang GG, Collier JL (1993) The phycobilisome, a light-harvesting complex responsive to environmental conditions. *Microbiol Rev* 57(3):725–749.
46. Latifi A, Ruiz M, Zhang C-C (2009) Oxidative stress in cyanobacteria. *FEMS Microbiol Rev* 33(2):258–278.
47. Perelman A, Uzan A, Hacohen D, Schwarz R (2003) Oxidative stress in *Synechococcus* sp. strain PCC 7942: Various mechanisms for H₂O₂ detoxification with different physiological roles. *J Bacteriol* 185(12):3654–3660.
48. 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(4):1672–1685.
49. Chardonnet S, et al. (2015) First proteomic study of S-glutathionylation in cyanobacteria. *J Proteome Res* 14(1):59–71.
50. Rastogi RP, Singh SP, Häder D-P, Sinha RP (2010) Detection of reactive oxygen species (ROS) by the oxidant-sensing probe 2',7'-dichlorodihydrofluorescein diacetate in the cyanobacterium *Anabaena variabilis* PCC 7937. *Biochem Biophys Res Commun* 397(3):603–607.
51. Lea-Smith DJ, et al. (2013) Thylakoid terminal oxidases are essential for the cyanobacterium *Synechocystis* sp. PCC 6803 to survive rapidly changing light intensities. *Plant Physiol* 162(1):484–495.
52. Saha R, et al. (2016) Diurnal regulation of cellular processes in the cyanobacterium *Synechocystis* sp. Strain PCC 6803: Insights from transcriptomic, fluxomic, and physiological analyses. *mBio* 7(3):e00464–16.
53. Lambert G, Chew J, Rust MJ (2016) Costs of clock-environment misalignment in individual cyanobacterial cells. *Biophys J* 111(4):883–891.
54. Guerreiro ACL, et al. (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(8):2042–2055.
55. Nikkanen L, Rintamäki E (2014) Thioredoxin-dependent regulatory networks in chloroplasts under fluctuating light conditions. *Philos Trans R Soc Lond B Biol Sci* 369(1640):20130224.
56. Diaz-Troya S, López-Maury L, Sánchez-Riego AM, Roldán M, Florencio FJ (2014) Redox regulation of glycogen biosynthesis in the cyanobacterium *Synechocystis* sp. PCC 6803: Analysis of the AGP and glycogen synthases. *Mol Plant* 7(1):87–100.
57. Hosokawa N, et al. (2011) Circadian transcriptional regulation by the posttranslational oscillator without de novo clock gene expression in *Synechococcus*. *Proc Natl Acad Sci USA* 108(37):15396–15401.
58. Singh SC, Sinha RP, Häder D-P (2002) Role of lipids and fatty acids in stress tolerance in cyanobacteria. *Acta Protozool* 41(4):297–308.
59. Girotti AW (1998) Lipid hydroperoxide generation, turnover, and effector action in biological systems. *J Lipid Res* 39(8):1529–1542.
60. Hoyle NP, O'Neill JS (2015) Oxidation-reduction cycles of peroxiredoxin proteins and nontranscriptional aspects of timekeeping. *Biochemistry* 54(2):184–193.
61. Dietz K-J (2011) Peroxiredoxins in plants and cyanobacteria. *Antioxid Redox Signal* 15(4):1129–1159.
62. Fiehn O, et al. (2010) Plasma metabolomic profiles reflective of glucose homeostasis in non-diabetic and type 2 diabetic obese African-American women. *PLoS One* 5(12):e15234.
63. Fiehn O, et al. (2008) Quality control for plant metabolomics: Reporting MSI-compliant studies. *Plant J* 53(4):691–704.
64. Xia J, Sinelnikov IV, Han B, Wishart DS (2015) MetaboAnalyst 3.0-making metabolomics more meaningful. *Nucleic Acids Res* 43(W1):W252–7.
65. R Core Team (2014) *R: A Language and Environment for Statistical Computing* (R Core Team, Vienna).
66. Kondo T, et al. (1994) Circadian clock mutants of cyanobacteria. *Science* 266(5188):1233–1236.
67. Clerico EM, Ditty JL, Golden SS (2007) Specialized techniques for site-directed mutagenesis in cyanobacteria. *Methods Mol Biol* 362:155–171.
68. Deatherage DE, Barrick JE (2014) Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using breseq. *Methods Mol Biol* 1151:165–188.
69. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J R Stat Soc B* 57:289–300.
70. Bates D, Mächler M, Bolker B, Walker S (2015) Fitting linear mixed-effects models using lme4. *J Stat Softw* 67(1):1–48.

Supporting Information

Diamond et al. 10.1073/pnas.1613078114

SI Text

The $\Delta rpaA$ Mutant Activates a Nitrogen-Starvation Transcriptional Response

In cyanobacteria, elevated AKG is the primary signal for nitrogen deprivation, and activation of this system results in a coordinated transcriptional response with the subsequent degradation of phycobilisome proteins (36, 37). The result is a bleaching process termed chlorosis. The observed AKG elevation in $\Delta rpaA$ mutants temporally precedes a rapid decrease in phycobilisome absorbance at 630 nm in $\Delta rpaA$ mutants that is evident after 8 h in the dark (Figs. 1B and 3B and Fig. S1). However, a variety of stress conditions can cause chlorosis (31). To determine whether elevated AKG is accompanied by a transcriptional response linked to nitrogen deprivation, we tracked transcript levels of *glnN* (glutamine synthase), a primary transcriptional target of this transcriptional response, before and 2 h after WT and the $\Delta rpaA$ mutant entered the dark (37). WT showed a slight decrease in *glnN* transcript levels after the dark transition, whereas the $\Delta rpaA$ mutant showed a statistically significant approximately fourfold increase in *glnN* transcripts (Fig. S3). Combined, these results are consistent with a transcriptionally activated nitrogen starvation response and indicate that AKG reaches high enough levels in $\Delta rpaA$ cells to activate this response in the dark. The activation of this response is normally repressed at night, and its activation may squander already limited cellular resources.

SI Materials and Methods

Cyanobacterial Strains, Media, and Culture Conditions. All $\Delta rpaA$ mutants were constructed by transformation in a WT *S. elongatus* background with plasmid pAM4420 (25) and were validated by PCR. For all experiments, precultures were first prepared in 100 mL of fresh BG-11 medium as described in Diamond et al. (19).

For metabolomics experiments, precultures were used to inoculate Phenometrics ePBR photobioreactors (Version 1.1; Phenometrics Inc.) at an initial density of $OD_{750} = 0.1$ in 400 mL of BG-11 medium without antibiotics. Temperature was maintained at 30 °C; filtered (0.2 μm) air was sparged at a rate of 50 mL/min, and light intensity was either 150 or 500 $\mu\text{E}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ provided from the top of the culture while lights were on. After inoculation, cultures were grown at a constant light intensity of 150 $\mu\text{E}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ until $OD_{750} = 0.3$, then maintained turbidostatically at this density for the duration of the experiment. In the metabolomics experiment, WT circadian rhythms were entrained by growth in a 12:12 LD cycle with the light portion at an intensity of 150 $\mu\text{E}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ for 1 d and subsequently at 500 $\mu\text{E}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ for 2 d before release into experimental conditions and sampling. The $\Delta rpaA$ strains were maintained in constant light at the same intensities as for the WT strain before the sampling procedure (Fig. S7).

For absorbance scanning, viable cell counts, qRT-PCR, MSM-treatment absorbance measurements, and oxidative stress measurements, precultures were used to inoculate the photobioreactors at an initial density of $OD_{750} = 0.2$ in 400 mL of BG-11 medium without antibiotics. Temperature, airflow rate, and light intensity settings were the same as above. For these experiments, both WT and $\Delta rpaA$ mutants were maintained at a constant light intensity of 150 $\mu\text{E}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ for 1 d. Subsequently, both strains were subjected to growth in a LD cycle with the light portion at an intensity of 150 $\mu\text{E}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ for 2 d, and then increased to 500 $\mu\text{E}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ over the final LD period during which sampling took place.

For viable cell plating, 200 μL of the indicated sample was serially diluted 1:5 in fresh BG-11 medium without antibiotics five times. For LD sensitivity testing, samples were first all diluted to an $OD_{750} = 0.2$, and the same dilution scheme was then followed. Subsequently, 4 μL of each sample was spotted onto solid BG-11 plates without antibiotics. Plates were incubated at 30 °C and 150 $\mu\text{E}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ constant light for 5–6 d. For LD sensitivity testing, samples were plated in duplicate, with one set incubated in constant light for 5–6 d and a second set under a LD cycle for 6–8 d.

Whole-Cell Absorbance Spectra Analysis. For all reported absorbance spectra, the absorbance between 400 and 750 nm was determined for 200 μL of the indicated sample by using a Tecan Infinite M200 plate reader. Raw absorbance values were normalized to OD_{750} of each sample. Statistical analyses used the Student's *t* test, with a *P* value < 0.05 considered as significant ($n \geq 3$).

Metabolomics and Data Analysis. Strains for metabolomics sampling were grown in photobioreactors and sampled ($n = 5$ at each time point and for each genotype) as described above (Fig. S7). Metabolite extraction and GC-TOF-MS were conducted by the WCMC at the University of California, Davis identically to the methods used in Diamond et al. (19) and Fiehn et al. (62, 63).

Raw metabolite abundance data for known metabolites (Dataset S1) were analyzed by using a combination of the online analysis platform MetaboAnalyst (Version 3.0) (64) and the statistical package R (65). PCA was applied to \log_2 -normalized and autoscaled data to detect outlying samples. Based on PCA, replicate A of the WT sample was removed from the dataset before statistical analysis (Fig. S8A). Also, WT replicate C time T6 and $\Delta rpaA$ replicate C time T4 were removed before analysis because of problems during sample extraction reported by the WCMC.

To detect metabolites that changed between WT and $\Delta rpaA$ at the initial sampling time point (0 h), we used a *t* test on \log_2 -normalized data (FDR corrected *P* value < 0.05) (69) and required significant metabolites to show a greater than twofold change between genotypes. To detect metabolites that changed between WT and $\Delta rpaA$ over the entire time course, we applied two separate statistical methods to \log_2 -normalized abundance data: two-way ANOVA analysis [Tukey's honest significant difference (HSD); $P < 0.05$] using the MetaboAnalyst platform (Version 3.0) and a linear mixed-effect model via the lme4 package in R (65, 70). For mixed-effect linear models, genotype and time were set as fixed effects, while biological replicate and time of sampling were allowed to contribute random effects. Models produced using genotype and an interaction of genotype and time as fixed effects were compared with a base model without genotype or the interaction effect using ANOVA. Final *P* values generated for each metabolite were corrected by using the method of Benjamini and Hochberg (69) (FDR < 0.05 was considered significant). Both the ANOVA and lme4 methods produced almost identical lists of significant metabolites (Dataset S1); we chose to use the output from two-way ANOVA in our analysis, because of the broader understanding of this method and for simplification of downstream analysis. Hierarchical clustering of significant metabolites for heatmap ordering (Fig. 2A) was performed by using the R package pheatmap, with Euclidian distance and complete linkage for leaf ordering.

PLS-DA modeling was carried out on \log_2 -normalized and autoscaled data with genotype and time as class factors. Using leave-one-out cross-validation indicated that the PLS-DA model

providing the most prediction accuracy used two components (Fig. S8B). Permutation testing was used to confirm that our PLS-DA model had a statistically significant ability to correctly predict class membership of samples based on the top 25 discriminating metabolites relative to a random permuted model ($P = 0.008$; $n = 1,000$) (Fig. S8C).

Metabolite KEGG enrichment analysis was conducted by using a custom-written R script and the metabolic categories in Dataset S2. Statistical overrepresentation was determined by using Fisher's exact test. Because of the difficulty in achieving high statistical significance with the low numbers of tested metabolites, we did not apply a multiple-testing correction to P values, but required more than three metabolites to be present in a pathway with a P value < 0.05 to be considered significant.

qRT-PCR Analysis. For each cyanobacterial sample, 10 mL of culture at an OD_{750} of 0.2–0.4 was collected and immediately placed on ice. The cultures were then centrifuged for 10 min at $4,000 \times g$ and -10°C . Pellets were then frozen at -80°C until extraction. Total RNA was extracted by using the TriZol reagent (Life Technologies) and the Direct-zol RNA MiniPrep Kit (Zymo Research). Briefly, the frozen pellets were thawed on ice and resuspended thoroughly in 1 mL of TriZol reagent. Cell suspensions were then transferred to 1.5-mL microcentrifuge tubes on ice, and cells were lysed by 5–10 cycles of vortexing for 30 s at room temperature and then allowed to sit on ice for 30 s. Cell debris was pelleted by centrifuging at $16,000 \times g$ at room temperature for 5 min. After transferring the supernatant fraction to an RNase-free 2-mL tube, 1 volume of 100% ethanol for every volume of TriZol (typically 1 mL) was added and mixed by pipetting up and down. Total RNA was isolated from the TriZol-ethanol mixture following the manual of the Direct-zol RNA MiniPrep Kit. The RNA quality was checked with agarose gel imaging (1% agarose, $0.5 \times \text{TBE}$, 75 V/60 min), using 10,000 \times SYBR green II RNA gel stain (Lonza). The extracted RNA samples were treated with DNaseI (Thermo Scientific) to remove contaminating genomic DNA. cDNA was synthesized with the SuperScript III First-Strand Synthesis System for RT-PCR (Life Technologies) following the kit manual. For qRT-PCR experiments, standard reactions in triplicate were set up with the Power SYBR Green PCR Master Mix (Life Technologies) and run on a StepOnePlus Real-Time PCR System (Life Technologies) following the instructions of the manufacturer. Significance for a change in *glnN* expression was calculated by using one-way ANOVA and Tukey's HSD ($n = 3$).

Mutagenesis and Identification of $\Delta rpaA$ Suppressor Mutations. EMS mutagenesis of fresh $\Delta rpaA$ mutant cyanobacterial strains was carried out as described in Kondo et al. (66). Mutagenized cultures were resuspended in 5 mL of BG-11 medium with appropriate antibiotics and incubated at 30°C , using 150 rpm shaking, under $30 \mu\text{E}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ constant light for 2 d. Subsequently 300 μL of mutagenized $\Delta rpaA$ cells (as well as untreated controls) were plated on BG-11 plates with appropriate antibiotics. Plates were prepared in duplicate, and one set was incubated at 30°C and $150 \mu\text{E}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ constant light for 15 d, with the second set being incubated at 30°C and a LD cycle with a light intensity of $150 \mu\text{E}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ for 15 d. Colonies (40) that formed on the LD-grown plate that contained EMS-mutagenized $\Delta rpaA$ cells were picked and patched onto BG-11 plates with appropriate antibiotics and grown for a further 10 d at 30°C and a LD cycle with a light intensity of $150 \mu\text{E}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ to confirm suppression of the LD death phenotype. Patches of surviving $\Delta rpaA$ EMS mutants were then transferred to 10 mL BG-11 medium and incubated at 30°C , using 150 rpm shaking, with a LD cycle and light intensity of $120 \mu\text{E}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ for between 11 and 16 d. Absorbance scans were taken of all cultures as described above, and genomic DNA

was extracted by using standard methods (67). Before sequencing genomic DNA, the disruption of *rpaA* in all strains was verified by PCR (Fig. S4C).

Genomic library preparation for Illumina short-read sequencing was performed by using the NEBNext DNA library preparation kit (NEB, catalog no. E6040S/L) with NEXTflex barcoded cDNA adaptors (BIOO Scientific, catalog no. 514104). Samples were run on an Illumina HiSeq2500 DNA sequencer at the University of California, Berkeley QB3 Genomics Sequencing Laboratory. Sequencing runs resulted in 50-bp reads with a median coverage depth of $45.7 \times$ per sample over the *S. elongatus* genome. However, for samples 1.1, 1.25, and 2.3, (Dataset S2), genome resequencing was performed at Bio Applied Technologies Joint, Inc., by using the Ion PGM system (Life Technologies) following the standard workflow illustrated in the manuals of the Life Technologies kits. Sequencing libraries were prepared with the Ion Xpress Plus Fragment Library Kit. In brief, 1 μg of total gDNA was sheared into desired fragment size (200–400 bp) via enzymatic digestion and then ligated to specified sequencing adaptors and/or barcode adaptors. The sequencing templates were amplified from constructed libraries by using the Ion OneTouch 2 System and the Ion PGM Hi-Q OT2 Kit. Amplified templates were processed with the Ion PGM Hi-Q Sequencing Kit and loaded on an Ion 316 Chip Kit (Version 2) for sequencing on a Ion PGM system with 500 sequencing flows (200 bp reads). Reads from all sequencing methods were mapped against the *S. elongatus* genome (GenBank accession no. NC_007604) and the large plasmid pANL (GenBank accession no. AF441790), and polymorphisms were called by using the program breseq (68). Polymorphisms were filtered for “high confidence” mutations by removing all mutations that resulted in a high-use codon coding for the same amino acid. All mutations in noncoding regions were included in the high-confidence set.

KEGG enrichment analysis of mutated genes was conducted by using a custom-written R script and the metabolic categories in Dataset S2. Statistical overrepresentation was determined by using the binomial test, and P values were corrected by using the method of Benjamini and Hochberg (69).

MSM Treatment. Initially by testing decreasing concentrations of MSM (Sigma-Aldrich catalog no. N12482) on WT and $\Delta rpaA$ cells, we were able to determine that concentrations of 100 nM and 25 μM , in solid and liquid media, respectively, were non-lethal to WT and $\Delta rpaA$, yet resulted in slightly suppressed growth and suppression of LD lethality. For all experiments using MSM, a stock solution was filter-sterilized (0.2 μm), and subsequently added to sterile medium to reach the desired working concentration. For experiments on solid medium, MSM was directly added to plate agar before plating and growing cells as described above. For experiments in photobioreactors, cells were grown as described above, and sterile MSM was added to a concentration of 25 μM at the beginning of a 12-h light period that preceded the transition into darkness for $\Delta rpaA$ cultures.

Quantification of ROS. ROS were quantified by using the fluorescent marker H_2DCFDA (Life Technologies catalog no. D399). Briefly, 2 mL of photobioreactor-grown (see above) culture was collected and split into 1-mL aliquots. H_2DCFDA was added to one sample at a final concentration of 5 μM . Tubes were protected from light and shaken at 30°C for 30 min. After incubation, 200 μL of each tube was added to a separate well in a 96-well plate. Fluorescence was quantified at an excitation of 480 nm and an emission of 520 nm on a Tecan Infinite M200 plate reader with the gain manually set to 120. Fluorescence data were normalized to OD_{750} of each sample, and untreated-sample background fluorescence was then subtracted from treated-sample fluorescence values.

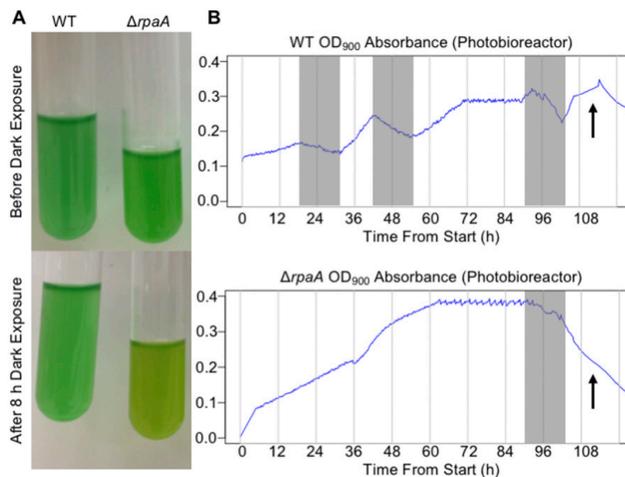


Fig. S1. Summary of phenotypic effects of darkness on WT and the $\Delta rpaA$ mutant. (A) Representative photographs of WT and $\Delta rpaA$ before and after 8 h of dark exposure. Chlorotic bleaching is evident in the $\Delta rpaA$ strain after incubation in darkness (A, Lower). (B) Representative data collected from the photobioreactor optical density sensor (900 nm) over the course of an experiment where WT and the $\Delta rpaA$ mutant were exposed to darkness. Time is given from the inoculation of photobioreactors, and gray bars indicate 12-h periods of darkness. Turbidostatic growth of both cultures can be observed at ~72–84 h from the start of the experiment. Both cultures show a decrease in optical density in the final dark period; however, the WT culture resumes growth in the following light period (B, Upper, black arrow), whereas the optical density of the $\Delta rpaA$ mutant continues to decrease despite being in a light period (B, Lower, black arrow).

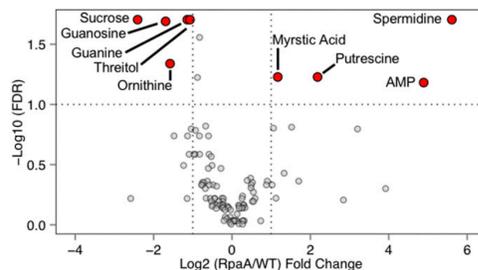


Fig. S2. Comparison of metabolite abundance between WT and $\Delta rpaA$ before entering the dark. This volcano plot of metabolites highlights metabolites that show a significant difference in abundance between WT and $\Delta rpaA$ at 0 h (named red points) vs. those without a detectable difference (gray points). Dotted lines indicate required thresholds for significance. Metabolites on the right and left sides of the plot were elevated and decreased in $\Delta rpaA$ relative to WT, respectively. Significance was calculated using Student's *t* test ($n = 4$ for WT and $n = 5$ for $\Delta rpaA$), and correction for multiple testing used the method of Benjamini–Hochberg.

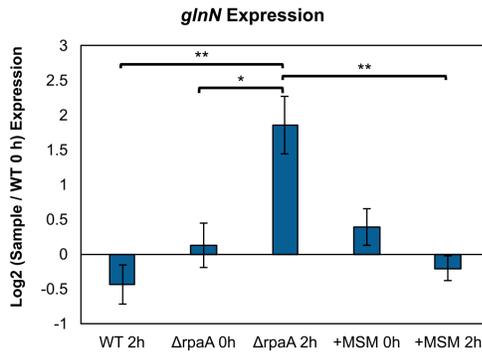


Fig. 53. Changes in expression of *glnN*. Relative expression levels of the *glnN* transcript at 0 h (before entering darkness) and 2 h after exposure to darkness as measured by qRT-PCR. Log₂ values were calculated relative to WT at 0 h. +MSM sample indicates $\Delta rpaA$ treated with 25 μ M MSM for a 12-h period in the light before entering the dark. Error bars indicate SEM. Significance calculated using one-way ANOVA and Tukey's HSD ($n = 3$). * $P < 0.05$; ** $P < 0.01$. Additionally, there was no significant difference between $\Delta rpaA$ treated with MSM, a WT control treated with MSM, and WT untreated with MSM at any time point, indicating that transcript abundance of *glnN* was not generally affected by MSM treatment.

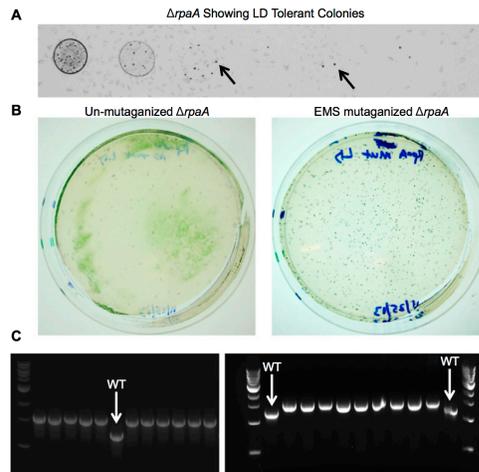


Fig. 54. Supporting data for EMS mutagenesis of the $\Delta rpaA$ mutant. (A) A representative photo of an older $\Delta rpaA$ culture that was plated as a serial dilution and grown in an LD cycle. Black arrows indicate $\Delta rpaA$ colonies showing robust growth even under normally restrictive LD conditions. (B) Photographs of plates of unmutagenized and EMS-mutagenized $\Delta rpaA$ cultures after incubation in an LD cycle for 15 d. The photographs show that hundreds of colonies form on the EMS-mutagenized $\Delta rpaA$ plate (B, Right), but not on the plate that received unmutagenized cells (B, Left). (C) PCR amplification of the *rpaA* locus from $\Delta rpaA$ cells carrying second-site mutations and from WT controls. The recombination at the *rpaA* locus with the pAM4420 vector to produce the $\Delta rpaA$ strain results in an expected amplified fragment of 1.8 kb. This region is larger than the 1.2-kb amplicon expected from WT cells. Comparison with amplification from WT cells (white arrows) shows that all $\Delta rpaA$ strains tested in this study produce the expected increased amplicon size for a strain that carries the $\Delta rpaA$ mutation. Additionally, no WT-size bands are present in the $\Delta rpaA$ strains, indicating that the mutation is fully segregated.

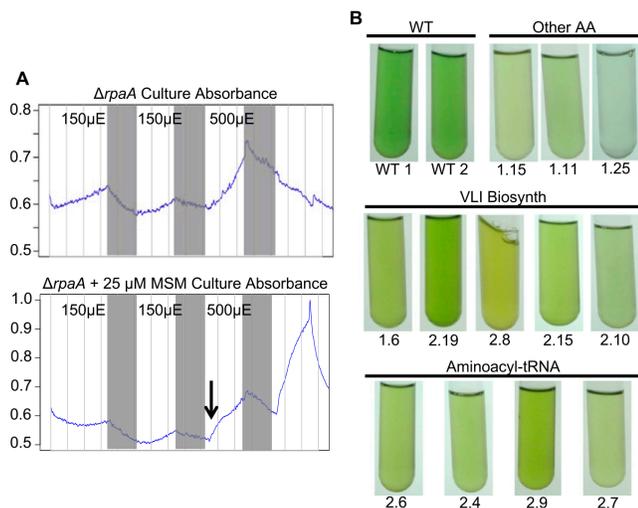


Fig. S5. Supporting data for potential mechanisms that suppress LD lethality in the $\Delta rpaA$ mutant. (A) Representative data collected from the photobioreactor optical density sensor (900 nm; y axis) over the course of an experiment where $\Delta rpaA$ mutants were exposed to darkness and not treated (Upper) or treated (Lower) with MSM. Time (x axis) starts at inoculation of photobioreactors, gray bars indicate 12-h periods of darkness, and light intensity during the light periods is noted with black text. The black arrow in the bottom panel indicates when 25 μM MSM was added to one culture. The $\Delta rpaA$ mutant receiving MSM was able to continue growth even after an LD cycle (Lower) that was lethal to the $\Delta rpaA$ mutant not receiving MSM (Upper). (B) Photographs of EMS-mutagenized $\Delta rpaA$ strains before genomic DNA extraction. Strains are organized by the type of mutation they were found to carry, and only strains with mutations that affected amino acid metabolism in some way are included. The general metabolic pathway affected is indicated above each photograph panel and the specific EMS mutant number is indicated below each test tube (Dataset S2). The photos highlight the altered pigmentation that is present in these strains relative to WT cells.

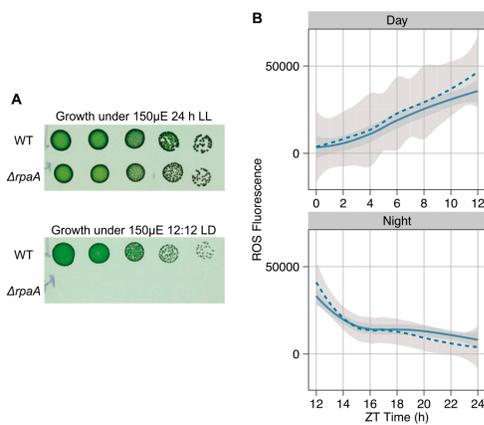


Fig. S6. Controls for LD plating and ROS experiments. (A) Serial dilutions of WT and $\Delta rpaA$ grown at a light intensity of $150 \mu\text{E}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ using 24-h LL (Upper) and 12 h:12 h LD (Lower) light regimes for 7 d. These data show the ability of $\Delta rpaA$ to tolerate high light conditions as long as a 24-h constant light regime is used. (B) Plot of H_2DCFDA fluorescence over a 24-h LD cycle indicating total cellular ROS in WT untreated (solid line) and treated (dashed line) with 25 μM MSM. Curves are best-fit lines calculated by using LOESS regression to all data points in a given sample; the gray-shaded area indicates the 95% CI of the regression line ($n = 21$ data points for day samples; $n = 42$ data points for night samples). The data show that treatment of WT with 25 μM MSM does not significantly affect levels of ROS over the 24-h LD cycle.

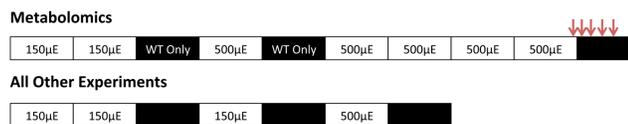


Fig. S7. Diagram of the growth and sampling scheme used for experiments conducted in photobioreactors. Each box indicates a 12-h period with white boxes corresponding to periods of light and black boxes corresponding to periods of darkness. Where indicated, only the WT strain was exposed to darkness. Light intensity during each light period is indicated in black text within each white box. Times when samples were taken during the metabolomics experiment are indicated by red arrows. Red arrows correspond to time points ZT12 (0 h), ZT13 (1 h), ZT14 (2 h), ZT16 (4 h), and ZT18 (6 h), from left to right.

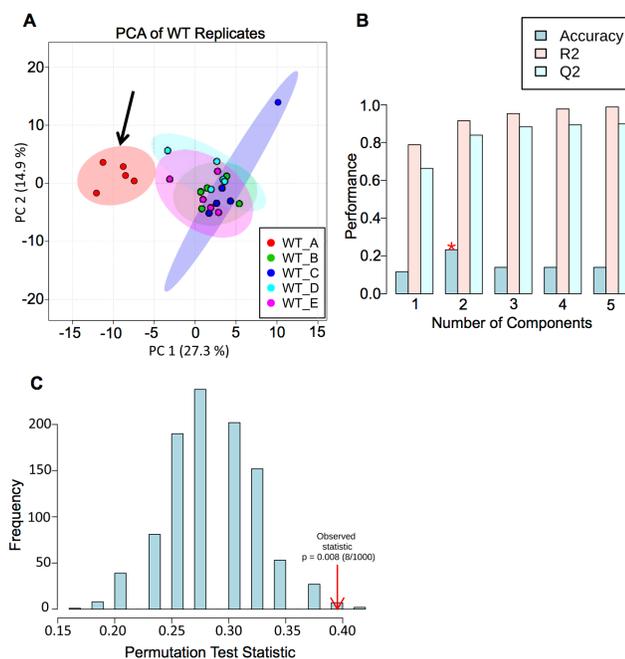


Fig. S8. Supporting data for metabolomics statistical analysis. (A) Plot of PCA components 1 and 2 for \log_2 autoscaled metabolite abundance data for all WT samples. Ellipses indicate the 95% CI for each sample group. The black arrow indicates all time points collected for WT biological replicate A. The statistically significant separation of the WT_A replicate indicates that this sample is an outlier relative to the other WT samples collected. (B) Plot showing result from LOOCV performed on the PLS-DA model. The red star indicates the accuracy of the model is highest when it includes only two components. Due to a high degree of variability in the data, which is typical of metabolomics datasets, we chose prediction accuracy as a metric to select a number of components over other metrics such as model fit (Q2). (C) Plot showing the results of accuracy permutation testing on the PLS-DA model. The red arrow indicates the test statistic. The data indicate that our PLS-DA model is significantly better at predicting class membership than a random model ($P < 0.01$; $n = 1,000$ permutations).

Table S1. Absorbance ratio (630/680 nm) of select $\Delta rpaA$ suppressor mutants

EMS mutant	Gene affected*	Color	Ratio	Significant [†]
WT controls				
WT_1	None	Green	0.939	No
WT_2	None	Green	0.944	
WT_3	None	Green	0.970	
Purine biosynthesis				
1_9	<i>guaA</i> (0189)	Light green	0.914	No
1_10	<i>guaA</i> (0189)	Light green	1.004	
1_3	<i>guaB</i> (1831)	Very light green	1.056	
1_1	<i>guaB</i> (1831)	Light green	0.979	
2_3	<i>guaA</i> (0189)	Yellow green	0.929	
BCAA biosynthesis				
1_6	<i>pyk</i> (0098)	Yellow green	0.847	Yes ($P < 0.01$)
2_19	<i>AHAS</i> (0139)	Yellow green	0.853	
2_8	<i>ilvD</i> (0626)	Strong yellow	0.915	
2_10	<i>eno</i> (0639)	Light green	0.874	
2_15	<i>ilvH</i> (2434)	Yellow green	0.892	
tRNA-related				
2_4	<i>ileS</i> (2437)	Yellow green	0.807	Yes ($P < 0.05$)
2_6	<i>glyQ</i> (2457)	Yellow green	0.889	
2_7	<i>tRNA-Arg</i> (R0011)	Pale yellow green	0.890	
2_9	<i>leuS</i> (1920)	Strong yellow green	0.854	
Other amino acid biosynthetic				
1_11	<i>trpB</i> (2143)	Pale green	0.931	No
1_15	<i>thrA</i> (2090)	Pale yellow	0.861	
1_25	<i>hisD</i> (1519)	Pale yellow/clear	1.049	

*SynPCC7942_# of gene given in parentheses next to gene name.

[†]Average of ratio from a functional gene group was different from WT average by Student's t test.

Other Supporting Information Files

[Dataset S1 \(XLSX\)](#)

[Dataset S2 \(XLSX\)](#)

2.4 Acknowledgments

Chapter 2, in part, is made up of reprints of two published manuscripts. The first is: Diamond S, Jun D, Rubin BE, and Golden SS (2015) The circadian oscillator in *Synechococcus elongatus* controls metabolite partitioning during diurnal growth. *Proc Natl Acad Sci* 112(15). The dissertation author was the tertiary author of this paper. The second is: Diamond S, Rubin BE, Shultzaberger RK, Chen Y, Barber DB, Golden SS. (2017) Redox crisis underlies conditional light-dark lethality in cyanobacterial mutants that lack the circadian regulator, RpaA. *Proc Natl Acad Sci* 114(4). The dissertation author was the secondary author of this paper.

CHAPTER 3: Associating Genotype to Phenotype with RB-TnSeq

3.1 Chapter Summary

Chapter 3 describes the development and use of RB-TnSeq in *S. elongatus* to achieve improved functional understanding of genes and intergenic regions throughout the genome. Section **3.2 RB-TnSeq Development and Essential Genes in *S. elongatus*** describes the creation of a pooled library of approximately 250,000 barcoded transposon mutants. The tracking of these mutants by sequencing led to identification of the first essential gene set in a photosynthetic organism, and was also used to make individual and global conclusions about the importance of intergenic regions throughout the genome. This work is presented here in the form of a *PNAS* paper on which I was first author. The data on gene essentiality was next used in the generation of a whole-genome metabolic model presented in section **3.3 RB-TnSeq Guided Metabolic Modeling** as a *PNAS* paper on which I was co-first author. This model has improved accuracy over those previously available for *S. elongatus* and revealed unique facets of the organism's biology.

The library is also a powerful tool for identifying genes that are beneficial or detrimental under different growth conditions. Identification of key genes involved in biofilm formation and amoeba grazing using RB-TnSeq are presented in **3.4 RB-TnSeq Screens: Biofilm Formation** and **3.5 RB-TnSeq Screens: Amoeba Grazing**. These sections include information from manuscripts that are

in preparation for publication on which I will be second and third author, respectively. Finally, section **3.6 RB-TnSeq Screens: Analysis Across Conditions** contains a brief description of meta-analysis of RB-TnSeq data across multiple conditions and organisms for wholesale gene functional annotation. This section cites, but does not include, a manuscript under review at *Nature* on which I'm a middle author and which currently resides in the *bioRxiv* preprint server. Together, these genome-wide datasets of the essential and conditionally relevant loci of *S. elongatus* improve our understanding of the organism's basic physiology, aid efforts to develop *S. elongatus* as a bioproduction platform, and serve as the only resource of their kind for a photosynthetic organism.

3.2 RB-TnSeq Development and Essential Genes in *S. elongatus*



The essential gene set of a photosynthetic organism

Benjamin E. Rubin^a, Kelly M. Wetmore^b, Morgan N. Price^b, Spencer Diamond^a, Ryan K. Shultzaberger^c, Laura C. Lowe^a, Genevieve Curtin^a, Adam P. Arkin^{b,d}, Adam Deutschbauer^b, and Susan S. Golden^{a,1}

^aDivision of Biological Sciences, University of California, San Diego, La Jolla, CA 92093; ^bPhysical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720; ^cKavli Institute for Brain and Mind, University of California, San Diego, La Jolla, CA 92093; and ^dDepartment of Bioengineering, University of California, Berkeley, CA 94720

Contributed by Susan S. Golden, September 29, 2015 (sent for review July 16, 2015; reviewed by Caroline S. Harwood and William B. Whitman)

Synechococcus elongatus PCC 7942 is a model organism used for studying photosynthesis and the circadian clock, and it is being developed for the production of fuel, industrial chemicals, and pharmaceuticals. To identify a comprehensive set of genes and intergenic regions that impacts fitness in *S. elongatus*, we created a pooled library of ~250,000 transposon mutants and used sequencing to identify the insertion locations. By analyzing the distribution and survival of these mutants, we identified 718 of the organism's 2,723 genes as essential for survival under laboratory conditions. The validity of the essential gene set is supported by its tight overlap with well-conserved genes and its enrichment for core biological processes. The differences noted between our dataset and these predictors of essentiality, however, have led to surprising biological insights. One such finding is that genes in a large portion of the TCA cycle are dispensable, suggesting that *S. elongatus* does not require a cyclic TCA process. Furthermore, the density of the transposon mutant library enabled individual and global statements about the essentiality of noncoding RNAs, regulatory elements, and other intergenic regions. In this way, a group I intron located in tRNA^{Leu}, which has been used extensively for phylogenetic studies, was shown here to be essential for the survival of *S. elongatus*. Our survey of essentiality for every locus in the *S. elongatus* genome serves as a powerful resource for understanding the organism's physiology and defines the essential gene set required for the growth of a photosynthetic organism.

RB-TnSeq | transposon mutagenesis | Tn-seq | cyanobacteria | photosynthesis

Determining the sets of genes necessary for survival of diverse organisms has helped to identify the fundamental processes that sustain life across an array of environments (1). This research has also served as the starting point for efforts by synthetic biologists to design organisms from scratch (2, 3). Despite the importance of essential gene sets, they have traditionally been challenging to gather because of the difficulty of observing mutations that result in lethal phenotypes. More recently, the pairing of transposon mutagenesis with next generation sequencing, referred to collectively as transposon sequencing (Tn-seq), has resulted in a dramatic advance in the identification of essential gene sets (4–7). The key characteristic of Tn-seq is the use of high-throughput sequencing to screen for the fitness of every transposon mutant in a pooled population to measure each mutation's impact on survival. These data can be used to quantitatively ascertain the effect of loss-of-function mutations at any given locus, intragenic or intergenic, in the conditions under which the library is grown (8). Essential gene sets for 42 diverse organisms distributed across all three domains have now been defined, largely through the use of Tn-seq (9). A recently developed variation on Tn-seq, random barcode transposon site sequencing (RB-TnSeq) (10), further minimizes the library preparation and sequencing costs of whole-genome mutant screens.

Despite the proliferation of genome-wide essentiality screens, a complete essential gene set has yet to be defined for a photosynthetic organism. A collection of phenotyped *Arabidopsis thaliana* mutants has been created but extends to only one-tenth of *Arabidopsis* genes (11). In algae, efforts are underway to produce a Tn-seq-like system in *Chlamydomonas reinhardtii*; however, the

mutant library currently lacks sufficient saturation to determine gene essentiality (12). To date, the essential genes for photoautotrophs have only been estimated by indirect means, such as by comparative genomics (13). The absence of experimentally determined essential gene sets in photosynthetic organisms, despite their importance to the environment and industrial production, is largely because of the difficulty and time required for genetic modification of these organisms.

Cyanobacteria comprise an extensively studied and ecologically important photosynthetic phylum. They are responsible for a large portion of marine primary production and have played a foundational role in research to decipher the molecular components of photosynthesis (14, 15). *Synechococcus elongatus* PCC 7942 is a particularly well-studied member of this phylum because of its genetic tractability and streamlined genome (16). As a result, it has been developed as a model photosynthetic organism and a production platform for a number of fuel products and high-value chemicals (17). Despite the importance of *S. elongatus* for understanding photosynthesis and industrial production, 40% of its genes have no functional annotation, and only a small portion of those that do have been studied experimentally.

Here, we use RB-TnSeq, a method that pairs high-density transposon mutagenesis and pooled mutant screens, to probe the *S. elongatus* genome for essential genes and noncoding regions. We categorized 96% of 2,723 genes in *S. elongatus* as either essential (lethal when mutated), beneficial (growth defect when mutated), or nonessential (no phenotype when mutated) under standard laboratory conditions. Furthermore, we determined the genome-wide essentiality of noncoding RNAs (ncRNAs), regulatory regions, and intergenic regions. Our investigation has produced an extensive analysis of the loci essential for the growth of a

Significance

Cyanobacteria are model organisms for photosynthesis in the laboratory, are key producers of the chemical energy that drives life, and are being developed as biofuel and chemical producers for industry. Despite the importance of these organisms for environmental and biotechnological applications, only a small percentage of cyanobacterial genes and intergenic regions have been experimentally evaluated for their impact on the organisms' survival. Here, we present experimental analysis of the complete set of genomic regions necessary for survival in a cyanobacterium achieved by screening for the fitness of hundreds of thousands of mutants. In addition to improving our fundamental understanding of Cyanobacteria, this research more broadly provides a snapshot of the essential genes and intergenic regions necessary to live the photosynthetic lifestyle.

Author contributions: B.E.R., K.M.W., S.D., A.D., and S.S.G. designed research; B.E.R., K.M.W., M.N.P., R.K.S., L.C.L., and G.C. performed research; K.M.W., M.N.P., A.P.A., and A.D. contributed new reagents/analytic tools; B.E.R., M.N.P., S.D., R.K.S., and S.S.G. analyzed data; and B.E.R. and S.S.G. wrote the paper.

Reviewers: C.S.H., University of Washington; and W.B.W., The University of Georgia.

The authors declare no conflict of interest.

See Commentary on page 14747.

¹To whom correspondence should be addressed. Email: sgolden@ucsd.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1519220112/-DCSupplemental.

photosynthetic organism and developed a powerful genomic tool that can be used for additional screens under a wide array of ecologically and industrially relevant growth conditions.

Results

Transposon Library Creation and Insertion Site Mapping. Our RB-TnSeq library in *S. elongatus* was constructed by mutagenesis with a Tn5-derived transposon delivered by conjugation. The transposon contains a kanamycin resistance cassette for selection of mutants. As an addition to the traditional Tn-Seq approach, in RB-TnSeq, a 20-bp random DNA barcode is also inserted with the resistance marker. These unique barcodes, after being linked to the surrounding sequences, serve as identifier tags for each insertion's location and simplify downstream genome-wide screens using BarSeq (10, 18). To achieve the efficiency of transposition necessary to create a high-density insertion library and minimize contamination by *Escherichia coli* DNA in sequencing reactions, several improvements were made to traditional *S. elongatus* conjugation protocols (16, 19), including increasing the light intensity during conjugation approximately fourfold, decreasing the conjugation time, and using an additional outgrowth step (*Materials and Methods* and *Table S1*). In total, ~375,000 individual transposon mutants were pooled to create the final library. The pooled library was sequenced before storage or outgrowth to map the location of each transposon insertion as well as its random DNA barcode (Tn-Seq). We identified 246,913 mutants with unique insertion locations that were supported by at least two sequencing reads. Insertion locations showed a relatively even distribution, with an average density of one insertion mutation present in the population for every 11 bp of the 2.7-Mbp genome (Fig. 1). The locations of all transposon insertions are presented in *Dataset S1*. The associated barcodes that could be mapped with high confidence are presented in *Dataset S2*.

S. elongatus maintains three to six copies of its genome (20, 21). Therefore, mutants containing a transposon insertion in an

essential region can acquire a kanamycin resistance insertion on one copy of the chromosome and retain viability by maintaining at least one copy of the essential WT allele. Indeed, removal of the selective agent from a transposon library in *Methanococcus maripaludis* has been previously shown to cause heterozygous mutants to lose their resistance-encoding insertions (22). To test the possibility that the pooled *S. elongatus* library is harboring heterozygous mutants, we performed an outgrowth of an aliquot of the mutant library in the absence of kanamycin alongside a control aliquot of the library containing kanamycin. Before and after this outgrowth, the abundances of the mutants comprising these library aliquots were assayed by sequencing only their DNA barcodes (BarSeq), which had been previously associated with insertion sites. The kanamycin and no kanamycin libraries had minimal divergence over seven to eight generations ($R^2 = 0.89$) (Fig. S1). These data suggest that heterozygosity had largely been resolved before analysis of the library and should have minimal impact on the conclusions drawn from this library.

Determining Gene Essentiality. To use the distribution of transposons in the library to make conclusions about essential genomic regions, it was necessary to first rule out potential sources of bias in our transposon insertion data. Polar effects, in which a transposon disrupts expression of downstream genes in a transcript, seem to have some influence but are not pervasive in the data (*SI Results*). Although previous studies have shown increased transposon insertion density around the origin of replication (23), our library did not contain such skewing (Fig. 1). Another concern was bias toward insertions occurring at specific sequence motifs; however, there was not strong enrichment for specific sequences around the insertion site (24) (Fig. S2). There was, however, a positive bias for insertion into guanine-cytosine (GC) rich regions. Thus, during the determination of gene essentiality, insertion frequency was normalized to GC content (Fig. S3).

To identify essential genes, we determined the number of insertions present in the library that mapped to each *S. elongatus* gene. Insertions within essential genes were expected to be underrepresented in the library, because such mutants should not be viable. To create a comparable measure of insertion density for genes, an insertion index, we normalized for the GC% bias of insertions (Fig. S3A) and divided the number of insertions in each gene by its length to get an insertion density. We also removed 25 genes from consideration that were either too short or too similar to other genes to measure confidently, and excluded the beginning and end 10% of every gene from analysis, because the extremes of otherwise essential genes can be permissive of insertions (25). In this way, we calculated an insertion index for 2,698 of 2,723 genes in *S. elongatus* (details are in *Materials and Methods*). The index had a bimodal distribution, with a group of putative essential genes with zero or very few insertions and a group of putative nonessential genes that could tolerate insertions (Fig. 2A). Using methods developed previously (5), we found a subset of genes that was four times more likely to belong to the distribution of genes with low insertion indexes and categorized them as essential genes. Genes that were four times more likely to be part of the set of genes with high insertion indexes were classified as nonessential genes, whereas those genes that fell in between these cutoffs were put in the ambiguous category. This initial survey of essentiality allowed the categorization of 1,889 nonessential genes, 764 likely essential genes, and 45 ambiguous genes.

These essentiality calls were further refined by thawing an aliquot of the library, growing it for an additional six generations, and assaying the abundance of its constitutive mutants by BarSeq. The results of this outgrowth were made more generalizable by conducting it in four commonly used laboratory conditions (*Materials and Methods*). Mutant abundance before and after outgrowth was used to determine a fitness score for each gene in the library over approximately six generations (Fig. 2B). Thus, in addition to both essential and nonessential genes, we were able to

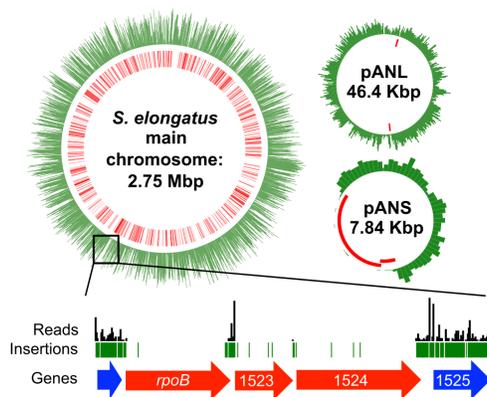


Fig. 1. The distribution of transposon mutations in the library overlaid across the *S. elongatus* main chromosome and two plasmids. *Upper*, the number of transposons (in 1,000-bp bins for the main chromosome and 100-bp bins for the plasmids) is represented by the length of the green bars in the outer circles. The locations of essential genes are shown in red in the inner circles. *Lower* shows a blown-up view of a region with underrepresentation of transposon insertions that encodes subunits of RNA polymerase. Lengths of black vertical bars represent numbers of sequence reads, and green bars indicate positions of insertions. Essential genes are in red, and nonessential genes are in blue (numbers represent SynPCC7942 gene numbers from the Joint Genome Institute annotation).

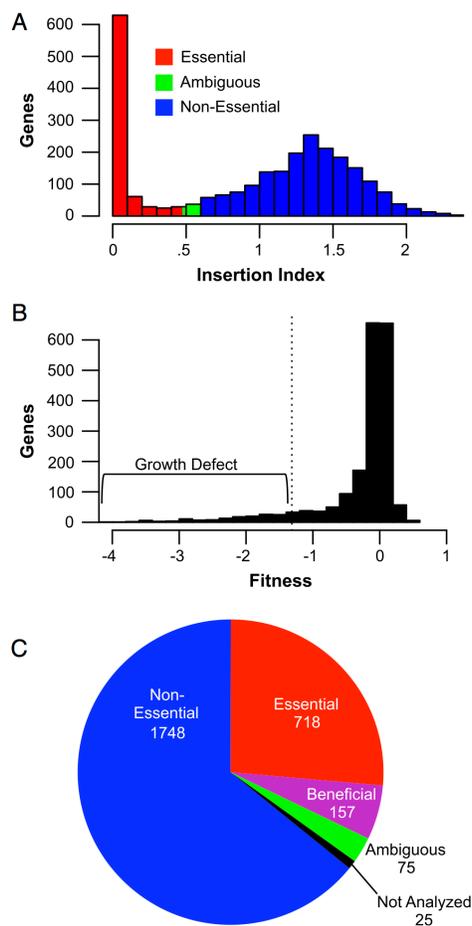


Fig. 2. The determination of gene essentiality. (A) The distribution of insertion indexes of all analyzed genes immediately after creation of the library, which was used to determine gene essentiality. The y axis indicates the number of genes, with the insertion index shown on the x axis. (B) The distribution of fitness for each gene after six generations used to refine the essentiality measurements and assign genes that are beneficial (growth defect when mutated). The y axis indicates the number of genes, with the fitness score shown on the x axis. The cutoff for beneficial genes that have significant growth defects when mutated is denoted by a dotted vertical line. Each gene's fitness is averaged from four growth samples in control conditions and normalized to zero, which represents a neutral fitness contribution. (C) The number of genes in the genome that are nonessential, essential, beneficial, ambiguous, or not analyzed.

identify 157 genes where insertions reduced average fitness across the four common laboratory conditions (average fitness ≤ -1.32) over the period of the outgrowth [$P < 0.01$ and false discovery rate (FDR) < 0.1 ; t test]. These genes were assigned to the new category of beneficial genes, which cause a growth defect when they are mutated in standard laboratory conditions. Interestingly, we identified no deleterious genes for which insertion

mutations conferred a growth advantage to *S. elongatus* under the conditions tested.

Data from these outgrowths were also used to make our essentiality calls more stringent. Genes were moved to the ambiguous bin when the data from the prefreeze characterization of the library and the outgrowth were conflicting (*Materials and Methods*). In this way, the final essentiality calls were made, in which 1,748 *S. elongatus* genes were called as nonessential, 718 were categorized as essential, 157 were binned as beneficial, and 75 were considered ambiguous (Fig. 2C and Dataset S3).

Comparisons to Other Essentiality Measures. The essential gene set experimentally derived in this study was compared with indirect measurements of gene importance to both provide support for our experimental results and identify potentially informative disagreements. One indirect assessment of gene importance was provided by gene conservation among different species of cyanobacteria. We compared 682 genes conserved across 13 diverse cyanobacterial genomes (26) with the set of essential genes identified in our study. Sixty percent of these conserved genes were also part of our essential gene set (Fig. 3A and Dataset S3), which represents a significant enrichment over random chance ($P < 0.001$; Fischer's exact test) and thus, a strong correlation between essentiality and conservation. The analysis was repeated with two other cyanobacterial conserved gene sets, which had very similar size overlaps with our essential gene set (27, 28), providing validation of the essentiality calls made using the library.

The genes that fall outside the overlap of essential and conserved genes are also of interest; 312 *S. elongatus* genes that we identified as essential but are not in the conserved gene set illustrate the limitation of determining gene importance by conservation and the necessity of using experimental approaches, such as RB-TnSeq, to determine essentiality. Conversely, 276 genes that are conserved but not essential may be important under environmental conditions that were not tested in this study.

The essential gene set was also probed for any enrichment in particular functional categories. The set was highly enriched for genes involved in synthesis of proteins, nucleic acids, and small molecules as well as lipid metabolism ($P < 0.05$ and FDR < 0.05) (Fig. 3B and Dataset S3). Very similar enrichment patterns have previously been observed in the *E. coli* set of essential genes (29), with the notable exception of energy metabolism, which is significantly enriched among *S. elongatus* essential genes and significantly underrepresented in the essential genes of *E. coli*. This discrepancy may be explained by the necessity of photosynthesis and carbon fixation in *S. elongatus*, which is extremely limited in the types of metabolism that it can perform; in contrast, *E. coli* can be grown on a wide variety of carbon sources. The enrichment in the *S. elongatus* essential gene set for conserved genes and core functional groups as well as its tight correlation with *E. coli* essential genes offer significant support to the validity of our essentiality calls.

A much broader measure of gene functionality is the mere presence or absence of a functional annotation. *S. elongatus* genes were divided into those that are annotated with functional predictions and those that are not (hypothetical genes). Hypothetical genes make up 40% of the genome; however, in the essential gene set, the portion of unannotated genes is only 15%. This difference likely represents the conservation of essential genes in other well-studied organisms as well as a bias toward studying genes that can be linked to measurable phenotypes. Despite their underrepresentation in the essential gene set, there are 109 genes called as essential that have no functional annotation in *S. elongatus* (Fig. 3A and Dataset S3). Among these genes, 21 are conserved throughout cyanobacteria (26), and 10 are conserved throughout the Greencut2 dataset of green plant and algae conserved genes (30) (Dataset S3). These unstudied but indispensable genes and specifically, those that are broadly conserved represent important targets for future research.

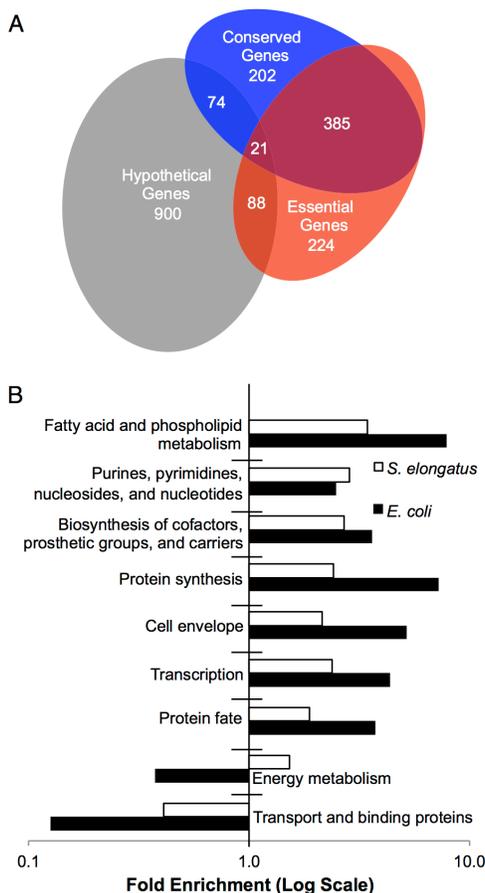


Fig. 3. Comparing the essential gene set with other predictors of gene importance. (A) The numbers of essential, conserved (26), and hypothetical genes that are overlapping and unique. (B) Fold enrichment for functional categories (TIGR function roles) that were significantly enriched or underrepresented in the essential gene sets of *E. coli* (29) (black bars) and *S. elongatus* (white bars).

Essentiality of Energy Metabolism.

Carbon metabolism. To further validate our essential gene set and explore any inconsistencies that it may have with predictions of gene importance based on conservation and functionality, we examined central carbon metabolism. This area is ideal for the verification of our essentiality predictions, because it is at the core of all life and must be leveraged for the development of cyanobacteria as a bioproduction platform. Any differences between our expectations for gene importance in these pathways and the experimental evidence from the transposon library suggest either incorrect calls of essentiality or interesting and unexpected biological findings.

We first examined genes in the following pathways of central carbon metabolism for their essentiality: the pentose phosphate pathway, glycolysis, the Calvin–Benson cycle, and the TCA cycle.

For each pathway, we identified the *S. elongatus* genes and any functional redundancy of these genes using BioCyc (31). Our expectation was that genes in these essential pathways of metabolism that are not functionally redundant and are conserved among cyanobacteria are likely to be essential. Of a total of 27 conserved nonredundant genes in these pathways, 22 agreed with this expectation and were called as likely essential from our library (Fig. 4A and Table S2), significantly more than could be expected by chance ($P < 0.001$; Fischer's exact test). Of five nonessential "disagreements" to this expectation, four were in the pentose phosphate pathway. One of these, transaldolase (*tal*; SynPCC7942_2297), acts in the nonoxidative phase of the pentose phosphate pathway. To validate its nonessentiality, we regenerated an insertion mutant in the *tal* gene to show that it is not required for growth under standard laboratory conditions (Fig. S4). The other three disagreements in the pentose phosphate pathway encode 6-phosphogluconolactonase (*pgl*; SynPCC7942_0529), glucose-6-phosphate 1-dehydrogenase (*zwf*; SynPCC7942_2334), and 6-phosphogluconate dehydrogenase (*gnd*; SynPCC7942_0039). These proteins make up the oxidative branch of the pathway, where reducing equivalents are produced in the form of NADPH. This finding is supported by previous literature, which has shown that both *zwf* and *gnd* mutants are viable (32, 33), although both mutants have decreased growth in light–dark cycles. This defect is likely because the cell relies on the oxidative branch of the pentose phosphate pathway for reducing equivalents when cells are in the dark and photosynthesis is inactive, whereas this pathway would be dispensable under the constant light of standard laboratory conditions.

The only other nonredundant member of central metabolism that is conserved but not essential is fumarate hydratase (*fumC*; SynPCC7942_1007). This finding was unexpected, because the absence of *fumC* would likely block cyclic flow through the TCA cycle. Furthermore, *fumC* is thought to be important for the recycling of fumarate in another freshwater cyanobacterium, *Synechocystis* sp. PCC 6803 (34). To validate this finding, we regenerated an insertion mutant of *fumC*. In accordance with our library-based call, we were able to obtain a fully segregated mutant (Fig. 4B) that grew at a statistically indistinguishable rate from the WT (Fig. 4C). Furthermore, the nonconserved enzyme directly upstream of *fumC*, succinate dehydrogenase (*sdhB*; SynPCC7942_1533), is also nonessential (Fig. 4D). The dispensability of these enzymes suggests that a complete TCA cycle is not required in *S. elongatus* under standard laboratory conditions.

Photosynthesis. Because *S. elongatus* serves as a model for photosynthesis, we examined the essentiality of some of the central components of the photosynthetic lifestyle. To provide a broad overview of core genes in the green lineage, we produced Table S3 of the *S. elongatus* genes called as essential here that are also present in the greencut2 dataset (30), which contains genes conserved among plants and green algae that are not present in nonphotosynthetic organisms (Table S3). These data provide a synopsis of some of the most conserved and important components of photoautotrophism. However, central components of photosynthesis that are not ubiquitous are not included in Table S3. As an example, carboxysome components are not contained in Table S3 because of their absence in most plants and algae, although many of them are essential for survival in *S. elongatus* according to the literature (35) and the essential gene set (Dataset S3).

We also examined the main complexes of photosynthetic light reactions for their essentiality in our dataset. We could not analyze the photosystem II core reaction center genes using our library, because the high sequence identity within the paralogous *psbA* (SynPCC7942_0424, SynPCC7942_0893, and SynPCC7942_1389) and *psbD* (SynPCC7942_0655 and SynPCC7942_1637) genes complicated transposon mapping; however, previous work has shown that the *psbA* genes and *psbDII* (SynPCC7942_1637) are not individually necessary (36, 37). The genes encoding the cytochrome *b₅₅₉* complex, *psbE* (SynPCC7942_1177) and *psbF* (SynPCC7942_1176), and the internal antenna proteins, *psbB* (SynPCC7942_0697) and *psbC* (SynPCC7942_0656), of photosystem

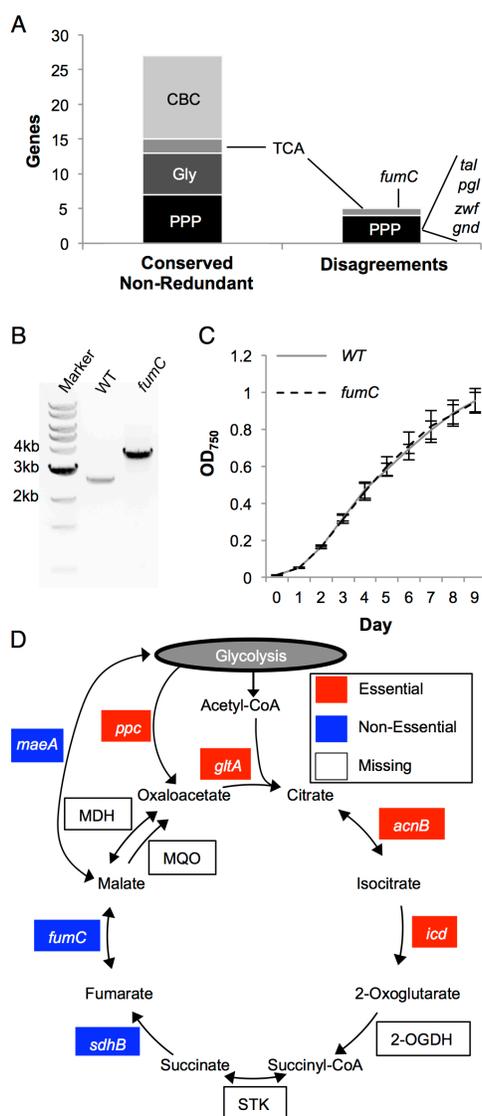


Fig. 4. Gene essentiality in central metabolism. (A) The number of genes that are conserved and nonredundant members of the Calvin-Benson cycle (CBC), the TCA cycle, glycolysis (Gly), and the pentose phosphate pathway (PPP). In the disagreements column, conserved nonredundant members of these pathways that are not essential are shown. (B) Genotypic characterization of the recreated *fumC* mutant. Lane 1, standard 1-kb ladder (New England Biolabs); lane 2, amplification of WT DNA with primers surrounding the *fumC* gene; lane 3, amplification of *fumC* mutant (8542-06), in which a 1.3-kb insertion is present, with the same primers. Each band is representative of three colonies tested. (C) Growth curves of the WT and *fumC* mutant strain. The error bars indicate the SDs for three independent replicates.

II were all classified as essential. These data are in agreement with studies in *Synechocystis* sp. PCC 6803, where cytochrome *b₅₅₉*, PsbB, and PsbC are required for photoautotrophic growth (38–40). Of the remaining 16 supportive and stabilizing proteins in photosystem II, only 4 [*psbH* (SynPCC7942_0225), *psbM* (synPCC7942_0699), *psbL* (SynPCC7942_1175), and *psbV* (SynPCC7942_2010)] were classified as essential. In the cytochrome *b_{6/f}* complex, the genes that encode the four large core subunits [*petA* (SynPCC7942_1231), *petB* (SynPCC7942_2331), *petC* (SynPCC7942_1232), and *petD* (SynPCC7942_2332)] were all called as essential. In accordance with the literature, the smaller subunits *petG* (SynPCC7942_1479) and *petN* (SynPCC7942_0475) were classified as essential (41), whereas *petM* (SynPCC7942_2426) was ambiguous. In photosystem I, the core genes, *psaA* (SynPCC7942_2049) and *psaB* (SynPCC7942_2048), were classified as essential along with two of three proteins that make the docking site for ferredoxin: *psaC* (SynPCC7942_0535) and *psaD* (SynPCC7942_1002). Only one of five remaining supporting genes, *psaI* (SynPCC7942_1249), was called as essential. Overall, as expected, the genes at the core of the photosynthetic light reactions were largely classified as essential, whereas genes with a more supportive role were largely predicted by the library to be nonessential or beneficial.

Beyond Coding Sequence.

Essentiality of ncRNAs. The saturation of the library is such that it was possible to do an extensive analysis of ncRNAs. There are currently three ncRNA loci in the National Center for Biotechnology Information (NCBI) *S. elongatus* genome annotation (NC_007604.1) that do not encode ribosomal or tRNAs. These widely conserved ncRNAs are *ssrA* (SynPCC7942_R0017), which mediates tagging of polypeptides for degradation, *mpB* (SynPCC7942_R0036), a member of the RNase P complex involved in tRNA processing, and *ffs* (SynPCC7942_R0047), the RNA component of the signal recognition particle (SRA), a ribonucleoprotein that targets proteins to the plasma membrane. We classified all three of these ncRNAs as essential in *S. elongatus*, which corresponds to findings in *E. coli* that *mpB* and *ffs* are essential (42, 43). Although not essential in *E. coli*, *ssrA* is essential in a number of other species (44). These three ncRNAs are included in the 718-gene essential gene set (Dataset S3).

Recently, 1,579 putative ncRNAs beyond those in the current NCBI annotation were identified in *S. elongatus* by RNA sequencing (45). To address the importance of these ncRNAs to the survival of the organism, we used the same approach taken for determining the gene essentiality of the previously annotated genes. Those ncRNAs that overlap each other or are too small to confidently predict essentiality were eliminated from the analysis. In addition, the ncRNAs encoded within genes that had been characterized by the library as essential, beneficial, ambiguous, or unanalyzed were not considered. This elimination was made, because ncRNAs in this set may be falsely called as essential when the gene surrounding or overlapping them is the true essential element. For the remaining 847 putative ncRNAs, we calculated insertion density and normalized it for GC bias to create an insertion index (Fig. S3B). The insertion indexes of these recently discovered ncRNAs, unlike the NCBI annotated genes described in Fig. 2A, did not contain a clear “essential peak” of ncRNA with low insertion indexes (Fig. 5A). The ncRNA insertion distribution was very similar to the previously analyzed nonessential genes, with a larger variance, presumably because of the short average length of the ncRNAs. Therefore, under standard laboratory conditions, these recently identified ncRNAs

(D) Essentiality in the TCA cycle. For enzymes that are present in *S. elongatus*, their names are shown: *acnB* (SynPCC7942_0903), *icd* (SynPCC7942_1719), *sdhB* (SynPCC7942_1533), *fumC* (SynPCC7942_1007), *gltA* (SynPCC7942_0612), *maeA* (SynPCC7942_1297), and *ppc* (SynPCC7942_2252). Abbreviations for enzymes that are missing are shown in white boxes: MDH, malate dehydrogenase; MQO, malate:quinone oxidoreductase; 2-OGDH, 2-oxoglutarate dehydrogenase; STK, succinate thiokinase.

have little effect on survival of the organism relative to the largely protein-coding set of genes, which were previously annotated.

Although most of the analyzed ncRNAs are nonessential, we identified 35 ncRNAs with normalized insertion densities below the essentiality cutoff as determined for annotated genes (Dataset S4). We manually examined the transposon insertion coverage around each of these 35 ncRNAs to ensure that the ncRNA did not fall in an area where transposons were underrepresented, such as regulatory regions for essential genes, or nonessential genes with below-average transposon numbers. Of 35 ncRNAs with normalized insertion densities below the cutoff, 10 were both visually and statistically considered to be underrepresented for insertions ($P < 0.01$ and $FDR < 0.05$; Poisson distribution) and called as likely essentials. Overall, we identified 13 likely essential ncRNAs that are not tRNAs or ribosomal: 10 from the recently discovered ncRNAs (45) and 3 with loci that had previously been annotated.

The 10 ncRNAs from the recently discovered set were searched against known ncRNA families using the RNA families database (RFAM) (46). One of the likely essential ncRNAs, ncRNA136, was identified as a putative group I intron (Fig. 5B). These introns are inserted into some cyanobacterial tRNA^{Leu} genes (47, 48) and have been shown to catalyze their own splicing out of pre-tRNA^{Leu} transcripts in vitro (49). In *S. elongatus*, ncRNA136 interrupts tRNA^{Leu} (UAA). There are four other uninterrupted tRNA^{Leu}s in *S. elongatus* with anticodons that were determined by tRNAscan-SE (50). Taking wobble into account, the anticodons of these four tRNAs cover five of six possible leucine codons. The ncRNA136 identified here as likely essential represents the fifth and final tRNA^{Leu} anticodon necessary to complement all six leucine codons. Therefore, this group I intron is likely essential to *S. elongatus*, because proper splicing of this nonredundant tRNA^{Leu} (UAA) cannot occur when it is mutated. The essentiality of this ncRNA was supported by our failure to regenerate insertion loss-of-function mutants for ncRNA136 in parallel with successful generation of mutants for both surrounding genes (Fig. 5C). In conclusion, there is no evidence for nonribosomal, non-tRNA ncRNAs having global importance close to that of protein-coding genes, but there is a smaller set of 13 likely essential ncRNAs, including ncRNA136, a group I intron.

Essential regulatory regions. To characterize the essential regulatory regions of *S. elongatus*, we examined insertion frequencies upstream from the predicted start codon of every essential gene. It might be expected that insertions in the regulatory regions of essential genes would have a lesser effect if the promoter for the transposon's antibiotic resistance gene lay in the same direction as the essential gene. However, we found that the average insertion frequency in the 100 bp upstream of essential genes was very similar for insertions in the same or opposite orientation as the essential gene (0.044 and 0.050, respectively). Therefore, we ignored directionality of upstream transposon insertions and analyzed them as a group. To define the average regulatory region for essential genes, we compared the insertion density upstream of the translation start site for essential genes with that of nonessential genes. We found that the region from the start codon to 52 bp upstream had a significantly lower transposon insertion rate in essential genes relative to nonessential genes ($P < 0.05$ and $FDR < 0.05$; Poisson distribution) (Fig. 6A). This region is large enough to encompass the Shine-Dalgarno sequence and basal promoter (51, 52).

It is of note that, even in the regulatory positions with the lowest average insertion density, the upstream regions of essential genes are still reasonably permissive of insertions. The density of insertions at this low point is one insertion mutant every 15 bp compared with a genome-wide average of one insertion mutant every 11 bp. Therefore, it is likely that many essential genes can still be transcribed sufficiently to support cell survival, even with transposon mutations directly upstream of the start codon.

To further explore the essential genes with insertions directly upstream of their start codons, essential genes were examined individually. Of those 557 essential genes that contained an insertion within either 200 bp upstream of their translation start

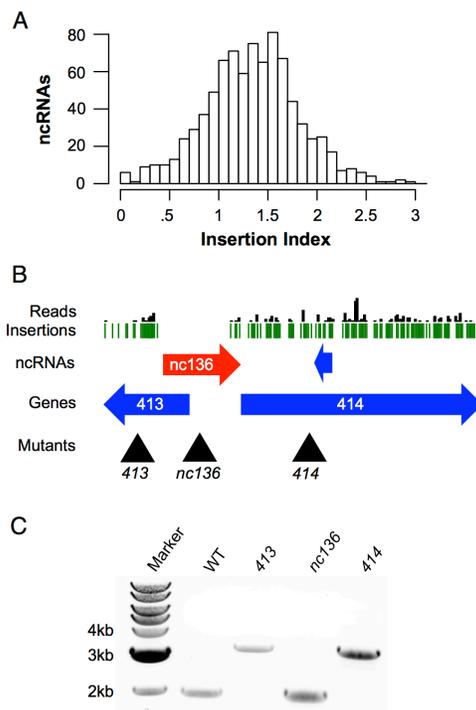


Fig. 5. Essentiality of ncRNAs. (A) The distribution of insertion indexes for the recently discovered ncRNAs (45). Axes are the same as in Fig. 2A. (B) The insertion distribution in and around the group I intron: ncRNA136. Lengths of black vertical bars represent numbers of sequence reads, and green bars indicate positions of insertions. The nonessential genes surrounding the essential ncRNA136 (red arrow) are shown as blue arrows. Black triangles indicate the locations of insertion mutations used to support the essentiality of ncRNA136. (C) Genotypic characterization of the failure to create a mutant of ncRNA136. Lane 1, standard 1-kb ladder (New England Biolabs); lane 2, amplification of WT DNA with primers surrounding ncRNA136 and both flanking genes; lane 3, amplification with the same primers of the region, in which the gene that flanks the ncRNA136 on the left, SynPCC7942_0413 (2E11-E-C4), carries a 1.3-kb insertion; lane 4, amplification with the same primers of a putative transformant, in which interruption of ncRNA136 (2E11-E-N7) was attempted, but the 1.3-kb insertion is absent; lane 5, amplification with the same primers of the region, in which the gene that flanks the ncRNA136 on the right, SynPCC7942_0414 (2E11-E-N11), carries a 1.3-kb insertion. Each band is representative of genotyping of three colonies.

site or before the closest upstream gene, 382 were able to sustain transposon insertions within 20 bp of the translation start site (Fig. 6B). Only 138 of the essential genes had no upstream insertions or genes within 40 bp of the start codon, and these upstream regions were categorized as likely essential ($P < 0.01$ and $FDR < 0.05$). For all essential genes, the regulatory regions and the length for which they are uninterrupted by a transposon insertion are presented in Dataset S5. The small number of essential upstream regions that we identified and the prevalence of insertions near start codons suggest that sufficient transcription can occur in the absence of typical regulatory elements or that the polymerase is able to read through the transposon cassette.

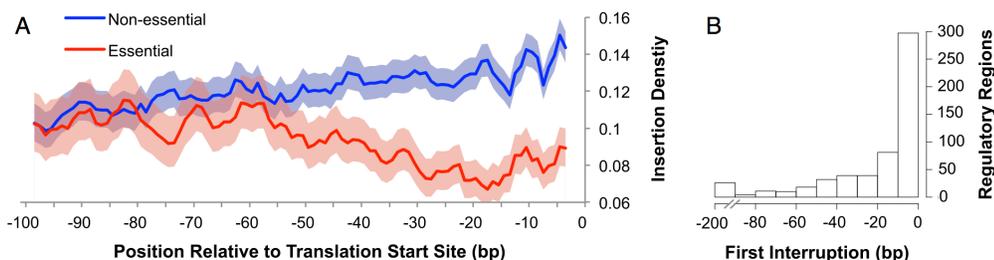


Fig. 6. Essential regulatory regions. (A) Transposon insertion density (insertions per base pair) on the y axis is plotted against the distance upstream from the translational start sites of essential genes (red) and nonessential genes (blue). Shading around the solid lines represents 95% confidence intervals (Poisson distribution). (B) For 557 essential genes that had an insertion before the nearest upstream gene and within 200 bp of the translation start site, the location of the closest insertion relative to the start site is shown.

Essential intergenic regions. To ensure that no essential regions had been missed by our survey for essentiality in genes, ncRNAs, and regulatory regions, we performed an unbiased analysis of insertion mutants present in the library to identify essential intergenic regions that we had not otherwise analyzed. Specifically, we searched the genome for regions of at least 100 bp for which there are no insertion mutants in the library. Forty-eight such regions were identified, with a maximum uninterrupted region of 222 bp and an average size of 130 bp (Dataset S6). Many of these regions, however, have very low GC%, and because insertion rate is GC%-dependent (Fig. S3), we only called the regions with GC% above 40% as “high-likelihood essentials.” There are 15 of these high-likelihood essential intergenic regions in the genome. Therefore, there are no large previously undetected essential regions; however, a small number of short likely essential regions could be detected, which may be regulatory regions or previously undiscovered ncRNAs.

Discussion

The density of the transposon library created for this study, with one insertion per 11 bp on average, enabled a rich and complete investigation of the genes and intergenic regions that are necessary for the photosynthetic lifestyle. In summary, we identified 718 putative essential genes, 13 likely essential non-tRNA, nonribosomal ncRNAs, 138 potential essential regulatory regions, and 15 other likely essential intergenic regions. The RB-TnSeq approach greatly extends the utility of the library, because it enables fast and inexpensive resequencing of the barcoded transposons in the population after an outgrowth period under standard laboratory conditions and can be used similarly to query the fitness contributions of each locus under additional growth conditions in the future.

There are certain limitations to the essentiality information determined here. Although we identified genes that are essential to the organism when individually mutated, they do not represent a minimal gene set. Essential processes for which there are redundant genes will not be discovered using an approach based on single mutants. In *S. elongatus*, however, this complication is of lesser concern than in most other cyanobacteria because of its small genome size, which at a streamlined 2.7 Mbp, harbors little redundancy. In addition, the findings of essentiality reported here apply only to the specific laboratory conditions used and are likely to be different for a subset of genes under other growth conditions. Finally, because ncRNAs, regulatory regions, and other intergenic regions are much smaller, on average, than protein-coding genes, the essentiality calls for these regions are inherently of lower confidence than those made for protein-coding genes. Therefore, conclusions of essentiality for non-coding loci and to a lesser extent, protein-coding genes must be validated by targeted mutation before definitive statements can be made about their essentiality.

TCA Cycle in Cyanobacteria. The ability to compare our essentiality results and predictions of gene importance based on conservation and function yielded fresh insights into fundamental *S. elongatus* biological processes. An example is the finding that two of the genes of the TCA cycle, including the widely conserved *fumC*, are dispensable in *S. elongatus*. The nature of the TCA cycle in cyanobacteria has been a subject of frequent debate. Until recently, it was assumed that the TCA cycle in cyanobacteria is incomplete because of the absence of the enzyme 2-oxoglutarate dehydrogenase (53, 54). More recent research, however, has closed the cyanobacterial TCA cycle with a number of bypasses, such as the 2-oxoglutarate decarboxylase pathway (55), the GABA shunt (56), and the glyoxylate cycle (57). The search for bypasses around missing elements of the TCA cycle presumes that having a complete cycle is important. Here, we found that the TCA cycle enzymes *sdhB* and *fumC* are nonessential in *S. elongatus*. Furthermore, no functionally annotated genes in *S. elongatus* account for the function of the TCA cycle enzymes malate dehydrogenase, malate:quinone oxidoreductase, succinate thiokinase, or 2-oxoglutarate dehydrogenase, and the 2-oxoglutarate decarboxylase bypass seems to be absent (Fig. 4D) (28). In agreement with the finding that these enzymes are nonessential or missing, the metabolites whose synthesis that they catalyze, with the exception of oxaloacetate, are not required for essential biosynthetic pathways in *Synechocystis* sp. PCC 6803 (58). Oxaloacetate is required for aspartate biosynthesis but can be produced without cyclic flux through the TCA cycle by phosphoenolpyruvate carboxylase (*ppc*; SynPCC7942_2252), shown here to be essential. Therefore, a large portion of the TCA cycle between 2-oxoglutarate and oxaloacetate seems to be nonessential for *S. elongatus* survival (Fig. 4D).

In light of these data, the traditional complete TCA cycle should be reconsidered in *S. elongatus*. The relevant pathways of the TCA cycle for the organism may resemble more closely the metabolism of certain obligate autotrophs, where cyclic flow through the TCA cycle is replaced by two separate branches that produce the metabolic precursors succinyl-CoA and 2-oxoglutarate independently (59). In *S. elongatus*, however, none of the necessary enzymes for the succinyl-CoA branch are functionally annotated (31) other than *fumC*, which is nonessential. This branch of the TCA cycle is likely dispensable because of the succinyl-CoA-independent pathway for heme biosynthesis in photosynthetic organisms (60). Therefore, the traditional understandings of the TCA cycle should be reassessed in *S. elongatus*, with consideration that its importance is likely not a result of its completeness or its role as an energy generator but in its provision of a few important precursor metabolites, such as oxaloacetate and 2-oxoglutarate, which likely require only short linear portions of the TCA cycle. This interpretation is compatible with the organism's strict photosynthetic metabolism, where the degradation of carbon for energy using cyclic flux through the TCA cycle would

be counterproductive at times when the organism is spending its energy to fix CO₂.

Beyond Coding Sequences. Although cyanobacterial ncRNAs have been studied extensively in silico, little is known about their individual importance in vivo. Mutants of *yrf1* (nc549) have been shown to be sensitive to several stresses in the closely related *S. elongatus* PCC 6301 (61). Another iron stress-dependent ncRNA, *IsrR* (nc468), regulates photosynthesis in *Synechocystis* sp. PCC 6803 (62). However, none of the nonribosomal, non-tRNA ncRNAs have been shown previously to be essential for survival under standard laboratory conditions in *S. elongatus*. In our study, we revealed 13 likely nonribosomal, non-tRNA essential ncRNAs: 10 from the recently discovered set of *S. elongatus* ncRNAs (45) and 3 that were previously annotated. One is a group I intron (ncRNA136), which catalyzes its own splicing out of the surrounding tRNA^{Leu} that carries the nonredundant UAA anticodon. Because the corresponding UUA codon is found 7,908 times in the *S. elongatus* genome (63), the inability of the pre-tRNA^{Leu} to correctly splice when this group I intron is mutated likely explains its essentiality. Although the tRNA^{Leu} (UAA) group I introns have been well-reported in cyanobacteria (48, 49, 64–67), there has been no previous work showing their importance in vivo. The other likely essential and so far unexplored ncRNAs discovered in this study represent interesting targets for additional research.

Library Stability. Although we examined the outgrowth of the library for genes whose loss improves growth, we found none. This finding is in contrast to other Tn-seq studies on various microbial species, in which some mutants outcompeted the rest of the library under standard laboratory conditions (5, 68). It is to be expected that, in novel environments to which a microbe has not adapted, there will be loss-of-function mutants that increase fitness (69). The lack of beneficial mutations found in this study likely speaks to the unusual culturing practice used for cyanobacteria: because inoculation from a frozen sample is a lengthy process, WT cultures are repeatedly passaged on benchtops and not inoculated from a freezer stock before each use. Thus, most *S. elongatus* cultures have been selected for laboratory conditions for years, and it is unsurprising that the laboratory-evolved genotype has no detrimental genes in these conditions. This caveat suggests that the strains on which experimentation is performed are no longer representative of the strains found in nature. For the purposes of the RB-TnSeq library created here, however, the absence of detrimental genes and the relatively small number of beneficial genes mean that the library loses little of its diversity over each generation (Fig. 2B). After reviving the library from frozen stocks and growing it for seven to eight generations, 93% of the mutant strains barcoded before freezing could still be found in the population. This robustness enables its use in screens under conditions of interest outside of standard laboratory conditions.

Future Uses of the Library. Essential and beneficial genes make up only about 32% of 2,723 genes in *S. elongatus*. Many of the remaining genes are likely important for specific biological conditions not experienced in standard laboratory conditions. We are currently exposing the library to an array of alternative conditions to determine genes specifically important for the survival of the organism under variations, such as high osmolarity and oxidative stress. We are also probing the library with targeted conditions to elucidate specific questions in cyanobacterial biology, such as the set of genes important for resistance to amoeba and the survival of light–dark cycles. This library can additionally be used for screens of phenotypes other than fitness if mutants with the phenotype of interest can be identified from the population and sequenced separately. With the use of RB-TnSeq, every additional screen requires minimal time, library preparation, and sequencing. This process is compared with previous screening techniques in Cyanobacteria, in which thousands of mutants had to be maintained

and phenotyped individually (70, 71). Finally, although we have maintained the pooled nature of the library for this study, it can be easily arrayed into individual clones when viewing the mutants under noncompetitive conditions is advantageous or when the phenotype of interest cannot be screened for in a pooled library. Using these approaches, the RB-TnSeq library, used here to delve into essentiality, will be a valuable tool for improving our understanding of *S. elongatus*, cyanobacteria, and photosynthetic organisms.

Materials and Methods

Strains and Culture Conditions. The library and individual insertion mutants were constructed in WT *S. elongatus* PCC 7942 stored in our laboratory as AMC06. All cultures were grown at 30 °C. Liquid cultures were shaken at 150 rpm (Thermo Fisher MaxQ 2000 Orbital Shaker) and grown in 100-mL flasks unless otherwise noted.

Mutant Library Creation. The conjugal recipient *S. elongatus* was inoculated and grown in BG-11 liquid medium (72) in light levels of 174–199 μmol photons·m⁻²·s⁻¹ for 3 d. For the *E. coli* donor, we used the diaminopimelic acid (DAP) auxotrophic strain APA766 that carried the library of barcoded Tn5 elements (pKMW7) (10). The donor *E. coli* was grown overnight in LB broth with 60 μg/mL DAP and 50 μg/mL kanamycin. *E. coli* cells were washed two times to remove kanamycin and resuspended in LB. The washed *E. coli* were mixed with *S. elongatus* at a 1:1 donor cell:recipient cell ratio on 0.45 μM nitrocellulose filters (Millipore) overlaid on LB agar plates with 60 μg/mL DAP. The conjugation reaction was performed for 7 h under 100–140 μmol photons·m⁻²·s⁻¹ of illumination. For selection of exconjugants, the filters were transferred to BG-11 kanamycin agar plates. To minimize *E. coli* contamination of the library, after 8 d of growth under 100–140 μmol photons·m⁻²·s⁻¹, the colonies on the filters were stamped onto new BG-11 kanamycin agar plates by pressing the filters face down onto the new plates. After 3 more d of growth with the same illumination, we scraped and flushed the colonies into BG-11 kanamycin liquid medium. At this point, cells were collected for DNA extraction (for Tn-Seq), and the remainder was frozen at –80 °C in 1-mL aliquots after adding 80 μL DMSO.

Library Preparation and DNA Sequencing for Tn-seq. To determine transposon insertion sites and link them to random DNA barcodes within each insertion, we created an Illumina-compatible sequencing library as described previously (10). Briefly, genomic DNA was extracted by phenol-chloroform extraction (16), sheared to 300 bp (Covaris), size-selected (Ampure SPRI), end-repaired, A-tailed, and ligated with adapters. Amplification of transposon insertions and flanking DNA was conducted using the transposon-specific primer, *Nspacer_barseq_universal* (ATGATACGGCGACCCAGAGATCTACACTCTTCCCTACACGACGCTCTCCGATCTNNNNNGATGCTCCACGAGGTCT), and the adaptor-specific primer, *P7_MOD_TS_index 12* primer (CAAGCAGAAGACGGCATAACGATTACAAGGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT) (10). For PCR, 100-μL reaction volumes were used with JumpStart Taq DNA Polymerase (Sigma) and the following thermocycler protocol: 94 °C for 2 min; 25 cycles of 94 °C for 30 s, 65 °C for 20 s, and 72 °C for 30 s; and a final extension at 72 °C for 10 min. The amplicons were then purified with AMPure XP Beads (Beckman Coulter), quantified on an Agilent Bioanalyzer with a DNA1000 Chip, and sequenced on a single lane of HiSeq2500 (Illumina) in rapid run mode.

Analysis of Tn-seq Data. Tn-seq reads were analyzed as described previously (10). Briefly, for each sequencing read, we identified the flanking sequence around the transposon and used BLAT (73) to map it to the *S. elongatus* genome. The part of the sequencing read internal to the transposon was used to link each transposon's unique barcode to its location within the genome. We identified 20,401,559 reads with insertions that mapped to the genome.

Library Outgrowths. Two sets of outgrowth experiments were conducted to examine the library for segregation and growth under control conditions. In both cases, library aliquots were thawed in a 37 °C water bath for 2 min and diluted 1:300 into BG-11 kanamycin liquid medium. The cultures were allowed to recover at 30 μmol photons·m⁻²·s⁻¹ without shaking for 1 d, moved back to standard shaking conditions, and allowed to grow for 4 d under 70 μmol photons·m⁻²·s⁻¹, at which point we collected cells for DNA extraction as the time 0 point of the outgrowth.

Aliquots of this culture were reinoculated into fresh BG-11 at OD₇₅₀ of 0.025. For segregation testing of the library, the culture was grown in parallel in the presence and absence of 5 μg/mL kanamycin for approximately

seven generations under $199 \mu\text{mol photons}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$, at which point both cultures were sampled for DNA extraction as the end point of the outgrowth. Approximately halfway through the growth period, the culture was reinoculated into fresh BG-11 liquid at OD_{750} of 0.025 to prevent the culture from reaching stationary phase.

For testing growth under standard laboratory conditions, cultures were grown in four conditions: on solid BG-11 kanamycin under $116 \mu\text{mol photons}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$, in liquid BG-11 kanamycin under $199 \mu\text{mol photons}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$, in liquid BG-11 kanamycin under $60 \mu\text{mol photons}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$, and in a Phenometrics ePBR v1.1 Photobioreactor (Phenometrics Inc.) maintained at a constant OD_{750} of 0.1 under $500 \mu\text{mol photons}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$. All liquid cultures were collected for BarSeq after six to eight generations. The growth on solid BG-11 kanamycin was conducted by spreading $100 \mu\text{L}$ library culture, diluted to have an OD_{750} of 0.086, onto the agar. The colonies were collected for BarSeq after 3 d of growth. The photobioreactor was inoculated with 400 mL library culture at an OD_{750} of 0.05 and bubbled at $50 \text{ mL}/\text{min}$ of $0.2\text{-}\mu\text{m}$ filtered air.

BarSeq. To use barcodes to quantify the survival of each mutant in the population, we first isolated genomic DNA through phenol-chloroform extraction (16). The procedure for sample preparation, sequencing, and preliminary analysis was described previously (10). Briefly, amplification of the barcode was done using 1 of 96 indexed forward primers for later multiplexing, BarSeq_P2_ITXXX (CAAGCAGAAGACGGCATACGAGATXXXXXXGTGACTGGAGTTTCAGAGCTGTGCTCTCCGATCTGTATGCCACGAGGTCTCT), and a common reverse primer, BarSeq_P1 (AATGATACGGCGACCCAGAGATCTACACTCTTTCCCTACACGACGCTCTCCGATCTNNNNNCGACTCGCAGCGTACG). For PCR, $50\text{-}\mu\text{L}$ reaction volumes were used with Q5 DNA polymerase, Q5 GC Enhancer (New England Biolabs), and the following thermocycler conditions: 98°C for 4 min; 25 cycles of 30 s at 98°C , 30 s at 55°C , and 30 s at 72°C ; and a final extension at 72°C for 5 min. The PCR products were then combined, purified with the DNA Clean & Concentrator Kit (Zymo Research), quantified using the Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific), and sequenced using Illumina HiSeq 2500. Barcodes were mapped to their previously identified positions in the genome using an R script. The fitness of each transposon mutant strain is its \log_2 change in abundance between the beginning of the experiment and the end:

$$\text{strain fitness} = \log_2 \left(\frac{(n_{\text{end}} + e_{\text{end}})}{(n_{\text{begin}} + e_{\text{begin}})} \right) + C,$$

where n_{end} and n_{begin} are the read counts of the strain's barcode from the end and the beginning of the experiment, e_{end} and e_{begin} are small constants that prevent infinite fitness values, and C is a normalization constant. The fitness of each gene is the weighted average of the fitness of strains within the central 10–90% of each gene. The normalization constant is chosen so that the peak of the gene fitness values is at zero (10).

Essentiality Analysis.

Genes. To determine gene essentiality, a normalized insertion index was created for the initial Tn-seq of the library and statistically analyzed for genes with underrepresentation of insertions. The JGI gene annotation was used for this mapping (chromosomes are stored under the GenBank accession nos. CP000100.1, CP000101.1, and S89470.1). The first step of creating the normalized insertion index was the elimination of genes from analysis that were shorter than 70 bp. The likelihood of the central 80% of a 70-bp gene having zero insertions by chance is $P < 0.01$ as calculated by the Poisson distribution. We used BLAT (73) to identify parts of genes that are nearly identical to other parts of the genome. Genes with any nearly identical parts were excluded from analysis. For the remaining genes, we divided the insertions in the middle 80% of each by the length of the middle 80% to create an insertion density for each gene. The insertion densities for all genes were then plotted against their GC%. A linear trend line was fitted to this plot and used to normalize gene insertion density by GC content (Fig. S3A). This normalized insertion density for each gene was given the label of insertion index. Finally, a preliminary essentiality measure was determined using an approach described previously (5). Briefly, γ -distributions were fit to the

essential and nonessential peaks in insertion index, and \log_2 likelihood ratios were calculated from these distributions. Genes with \log_2 likelihood ratios below -2 were called as essential genes, and those with \log_2 likelihood ratios above 2 were called as nonessential genes; genes that fell between these \log_2 likelihood ratios were called as ambiguous. Scripts were adapted from the Bio::Tradis pipeline (github.com/sanger-pathogens/Bio-Tradis) (74).

To improve the accuracy and precision of these essentiality calls, data from the outgrowths under standard laboratory conditions were used. A t test was used to find genes that had significantly lower or higher fitness under the four control conditions (Library Outgrowths), and a false discovery rate was determined for each P value. Genes for which insertions reduced gene fitness below -1.32 ($P < 0.01$ and $\text{FDR} < 0.1$; t test) were called as beneficial. Furthermore, to make our calls of essentiality more stringent, genes that were previously called as essential but were not significantly different from the mean gene fitness from the outgrowth were added to the ambiguous group. Conversely, those genes that had previously been called as non-essential but were not present in the outgrowth were also added to the ambiguous group.

ncRNAs. To determine essential ncRNAs, we calculated an insertion index using the same procedure as that used for previously annotated genes. We again only counted insertions in the middle 80% of ncRNAs and eliminated ncRNAs that overlapped each other. A length of 50 bp was set as the lower limit, because by using that cutoff, we would expect only one ncRNA that had zero insertions by chance as determined by the Poisson distribution. As had been done for essential genes, we again corrected for GC content (Fig. S3B) but in this case, also discarded ncRNAs with GC content below 35% because of the lower GC% of ncRNAs. We also eliminated ncRNAs that were overlapping essential, beneficial, ambiguous, or uncategorized genes, because it would be difficult to know whether essentiality of these ncRNAs was because of the ncRNAs themselves or the surrounding genes. The same cutoff for essentiality that had been applied to the gene insertion indexes was again applied to the ncRNAs. Those that fell on the essential side of this cutoff were visually examined to determine if they fell in likely essential regulatory regions or other areas of below-average insertion density. If they did not, they were categorized as likely essential ncRNAs.

Essential regulatory regions. Essentiality of the region upstream of the translation start site was determined for essential genes using the distance upstream of the start site for which there was no transposon mutants and no upstream gene. The minimum uninterrupted region necessary to be considered as a likely essential regulatory region was determined to be 40 bp, because a Poisson distribution predicted only one false positive using this cutoff.

Targeted Mutants: Transformation, Genotyping, and Growth Assays. Plasmids for targeted insertional mutation were taken from the unigene set, an existing insertion mutant library for *S. elongatus* (70, 75). Transformation of *S. elongatus* was achieved using standard protocols (16). Genotyping was done using colony PCR with Taq DNA Polymerase (NEB). Growth assays were done in liquid culture under $199 \mu\text{mol photons}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ of illumination.

ACKNOWLEDGMENTS. We thank B. Irvine for his input into metrics for essentiality; Drs. R. Simkovsky and A. Taton for thoughtful advice on experimental design and data analysis; A. Pal for assistance in conjugating the library; Drs. J. Bristow and L. Pennacchio for ideas and assistance at the conception of the project; Dr. D. Welkie for edits; and Dr. R. Steuer for consultation on the TCA cycle figure. This research was supported by National Science Foundation Grant MCB1244108 (to S.S.G.) and NIH Cell and Molecular Genetics Training Grant T32GM00724. BarSeq mutant fitness data were supported by Laboratory-Directed Research and Development Funding from Lawrence Berkeley National Laboratory provided by the Director, Office of Science of the US Department of Energy Contract DE-AC02-05CH11231 and a Community Science Project from the Joint Genome Institute (to A.P.A. and A.D.). The work conducted by the US Department of Energy Joint Genome Institute, a Department of Energy Office of Science User Facility, is supported by Office of Science of the US Department of Energy Contract DE-AC02-05CH11231.

- Juhas M, Eberl L, Glass JI (2011) Essence of life: Essential genes of minimal genomes. *Trends Cell Biol* 21(10):562–568.
- Hutchison CA, et al. (1999) Global transposon mutagenesis and a minimal Mycoplasma genome. *Science* 286(5447):2165–2169.
- Glass JL, et al. (2006) Essential genes of a minimal bacterium. *Proc Natl Acad Sci USA* 103(2):425–430.
- van Opijnen T, Bodi KL, Camilli A (2009) Tn-seq: High-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. *Nat Methods* 6(10):767–772.
- Langridge GC, et al. (2009) Simultaneous assay of every Salmonella Typhi gene using one million transposon mutants. *Genome Res* 19(12):2308–2316.
- Gawronski JD, Wong SMS, Giannoukos G, Ward DV, Akerley BJ (2009) Tracking insertion mutants within libraries by deep sequencing and a genome-wide screen for Haemophilus genes required in the lung. *Proc Natl Acad Sci USA* 106(38):16422–16427.
- Goodman AL, Wu M, Gordon JI (2011) Identifying microbial fitness determinants by insertion sequencing using genome-wide transposon mutant libraries. *Nat Protoc* 6(12):1969–1980.
- Christen B, et al. (2011) The essential genome of a bacterium. *Mol Syst Biol* 7(1):528.

9. Luo H, Lin Y, Gao F, Zhang CT, Zhang R (2014) DEG 10, an update of the database of essential genes that includes both protein-coding genes and noncoding genomic elements. *Nucleic Acids Res* 42(Database issue):D574–D580.
10. Wetmore KM, et al. (2015) Rapid quantification of mutant fitness in diverse bacteria by sequencing randomly bar-coded transposons. *mBio* 6(3):e00306-15.
11. Lloyd J, Meinke D (2012) A comprehensive dataset of genes with a loss-of-function mutant phenotype in Arabidopsis. *Plant Physiol* 158(3):1115–1129.
12. Zhang R, et al. (2014) High-throughput genotyping of green algal mutants reveals random distribution of mutagenic insertion sites and endonucleolytic cleavage of transforming DNA. *Plant Cell* 26(4):1398–1409.
13. Merchant SS, et al. (2007) The Chlamydomonas genome reveals the evolution of key animal and plant functions. *Science* 318(5848):245–250.
14. Zouni A, et al. (2001) Crystal structure of photosystem II from *Synechococcus elongatus* at 3.8 Å resolution. *Nature* 409(6821):739–743.
15. Jordan P, et al. (2001) Three-dimensional structure of cyanobacterial photosystem I at 2.5 Å resolution. *Nature* 411(6840):909–917.
16. Clerico EM, Ditty JL, Golden SS (2007) Specialized techniques for site-directed mutagenesis in cyanobacteria. *Methods Mol Biol* 362(2007):155–171.
17. Angermayr SA, Gorchs Rovira A, Hellingwerf KJ (2015) Metabolic engineering of cyanobacteria for the synthesis of commodity products. *Trends Biotechnol* 33(6):352–361.
18. Smith AM, et al. (2010) Highly-multiplexed barcode sequencing: An efficient method for parallel analysis of pooled samples. *Nucleic Acids Res* 38(13):e142.
19. Andersson CR, et al. (2000) Application of bioluminescence to the study of circadian rhythms in cyanobacteria. *Methods Enzymol* 305:527–542.
20. Griese M, Lange C, Soppa J (2011) Ploidy in cyanobacteria. *FEMS Microbiol Lett* 323(2):124–131.
21. Chen AH, Afonso B, Silver PA, Savage DF (2012) Spatial and temporal organization of chromosome duplication and segregation in the cyanobacterium *Synechococcus elongatus* PCC 7942. *PLoS One* 7(10):e47837.
22. Sarmiento F, Mrázek J, Whitman WB (2013) Genome-scale analysis of gene function in the hydrogenotrophic methanogenic archaeon *Methanococcus marisplacidus*. *Proc Natl Acad Sci USA* 110(12):4726–4731.
23. Chao MC, et al. (2013) High-resolution definition of the *Vibrio cholerae* essential gene set with hidden Markov model-based analyses of transposon-insertion sequencing data. *Nucleic Acids Res* 41(19):9033–9048.
24. Crooks GE, Hon G, Chandonia J-M, Brenner SE (2004) WebLogo: A sequence logo generator. *Genome Res* 14(6):1188–1190.
25. Griffin JE, et al. (2011) High-resolution phenotypic profiling defines genes essential for mycobacterial growth and cholesterol catabolism. *PLoS Pathog* 7(9):e1002251.
26. Shi T, Falkowski PG (2008) Genome evolution in cyanobacteria: The stable core and the variable shell. *Proc Natl Acad Sci USA* 105(7):2510–2515.
27. Simm S, Keller M, Selymes M, Schleiff E (2015) The composition of the global and feature specific cyanobacterial core-genomes. *Front Microbiol* 6:219.
28. Beck C, Knoop H, Axmann IM, Steuer R (2012) The diversity of cyanobacterial metabolism: Genome analysis of multiple phototrophic microorganisms. *BMC Genomics* 13(1):56.
29. Baba T, et al. (2006) Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: The Keio collection. *Mol Syst Biol* 2(2006):2006.0008.
30. Karpowicz SJ, Prochnik SE, Grossman AR, Merchant SS (2011) The GreenCut2 resource, a phylogenomically derived inventory of proteins specific to the plant lineage. *J Biol Chem* 286(24):21427–21439.
31. Caspi R, et al. (2014) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Res* 42(Database issue):D459–D471.
32. Scanlan DJ, Sundaram S, Newman J, Mann NH, Carr NG (1995) Characterization of a zwf mutant of *Synechococcus* sp. strain PCC 7942. *J Bacteriol* 177(9):2550–2553.
33. Broedel SE, Jr, Wolf RE, Jr (1990) Genetic tagging, cloning, and DNA sequence of the *Synechococcus* sp. strain PCC 7942 gene (gnd) encoding 6-phosphogluconate dehydrogenase. *J Bacteriol* 172(7):4023–4031.
34. Knoop H, Zilligies Y, Lockau W, Steuer R (2010) The metabolic network of *Synechocystis* sp. PCC 6803: Systemic properties of autotrophic growth. *Plant Physiol* 154(1):410–422.
35. Rae BD, Long BM, Badger MR, Price GD (2012) Structural determinants of the outer shell of β -carboxysomes in *Synechococcus elongatus* PCC 7942: Roles for CcmK2, K3-K4, CcmO, and CcmL. *PLoS One* 7(8):e43871.
36. Golden SS, Bruslan J, Haselkorn R (1986) Expression of a family of psbA genes encoding a photosystem II polypeptide in the cyanobacterium *Anacystis nidulans* R2. *EMBO J* 5(11):2789–2798.
37. Bustos SA, Golden SS (1991) Expression of the psbDII gene in *Synechococcus* sp. strain PCC 7942 requires sequences downstream of the transcription start site. *J Bacteriol* 173(23):7525–7533.
38. Pakrasi HB, Williams JG, Arntzen CJ (1988) Targeted mutagenesis of the psbE and psbF genes blocks photosynthetic electron transport: Evidence for a functional role of cytochrome b559 in photosystem II. *EMBO J* 7(2):325–332.
39. Eaton-Rye JJ, Vermaas WF (1991) Oligonucleotide-directed mutagenesis of psbB, the gene encoding CP47, employing a deletion mutant strain of the cyanobacterium *Synechocystis* sp. PCC 6803. *Plant Mol Biol* 17(6):1165–1177.
40. Carpenter SD, Charite J, Eggers B, Vermaas WF (1990) The psbC start codon in *Synechocystis* sp. PCC 6803. *FEBS Lett* 260(1):135–137.
41. Schneider D, Volkmer T, Rögner M (2007) PetG and PetN, but not PetL, are essential subunits of the cytochrome b6f complex from *Synechocystis* PCC 6803. *Res Microbiol* 158(1):45–50.
42. Brown S, Fournier MJ (1984) The 4.5 S RNA gene of *Escherichia coli* is essential for cell growth. *J Mol Biol* 178(3):533–550.
43. Waugh DS, Pace NR (1990) Complementation of an RNase P RNA (rnpB) gene deletion in *Escherichia coli* by homologous genes from distantly related eubacteria. *J Bacteriol* 172(11):6316–6322.
44. Karzai AW, Roche ED, Sauer RT (2000) The SsrA-SmpB system for protein tagging, directed degradation and ribosome rescue. *Nat Struct Biol* 7(6):449–455.
45. Vijayan V, Jain IH, O'Shea EK (2011) A high resolution map of a cyanobacterial transcriptome. *Genome Biol* 12(5):R47.
46. Nawrocki EP, et al. (2015) Rfam 12.0: Updates to the RNA families database. *Nucleic Acids Res* 43(Database issue):D130–D137.
47. Xu MQ, Kathe SD, Goodrich-Blair H, Nierzwicki-Bauer SA, Shub DA (1990) Bacterial origin of a chloroplast intron: Conserved self-splicing group I introns in cyanobacteria. *Science* 250(4987):1566–1570.
48. Kuhsel MG, Strickland R, Palmer JD (1990) An ancient group I intron shared by eubacteria and chloroplasts. *Science* 250(4987):1570–1573.
49. Sugita M, et al. (1995) Genes encoding the group I intron-containing tRNA(Leu) and subunit L of NADH dehydrogenase from the cyanobacterium *Synechococcus* PCC 6301. *DNA Res* 2(2):71–76.
50. Lowe TM, Eddy SR (1997) tRNAScan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 25(5):955–964.
51. Shultzaberger RK, Bucheimer RE, Rudd KE, Schneider TD (2001) Anatomy of *Escherichia coli* ribosome binding sites. *J Mol Biol* 313(1):215–228.
52. Shultzaberger RK, Chen Z, Lewis KA, Schneider TD (2007) Anatomy of *Escherichia coli* sigma70 promoters. *Nucleic Acids Res* 35(3):771–788.
53. Smith AJ, London J, Stanier RY (1967) Biochemical basis of obligate autotrophy in blue-green algae and thiobacilli. *J Bacteriol* 94(4):972–983.
54. Pearce J, Leach CK, Carr NG (1969) The incomplete tricarboxylic acid cycle in the blue-green alga *Anabaena variabilis*. *J Gen Microbiol* 55(3):371–378.
55. Zhang S, Bryant DA (2011) The tricarboxylic acid cycle in cyanobacteria. *Science* 334(6062):1551–1553.
56. Xiong W, Brune D, Vermaas WFJ (2014) The γ -aminobutyric acid shunt contributes to closing the tricarboxylic acid cycle in *Synechocystis* sp. PCC 6803. *Mol Microbiol* 93(4):786–796.
57. Zhang S, Bryant DA (2015) Biochemical validation of the glyoxylate cycle in the cyanobacterium *Chlorogloeopsis fritschii* Strain PCC 9212. *J Biol Chem* 290(22):14019–14030.
58. Knoop H, et al. (2013) Flux balance analysis of cyanobacterial metabolism: The metabolic network of *Synechocystis* sp. PCC 6803. *PLoS Comput Biol* 9(6):e1003081.
59. Wood AP, Aurikko JP, Kelly DP (2004) A challenge for 21st century molecular biology and biochemistry: What are the causes of obligate autotrophy and methanotrophy? *FEMS Microbiol Rev* 28(3):335–352.
60. Obornik M, Green BR (2005) Mosaic origin of the heme biosynthesis pathway in photosynthetic eukaryotes. *Mol Biol Evol* 22(12):2343–2353.
61. Nakamura T, Naito K, Yokota N, Sugita C, Sugita M (2007) A cyanobacterial non-coding RNA, Yfr1, is required for growth under multiple stress conditions. *Plant Cell Physiol* 48(9):1309–1318.
62. Dühring U, Axmann IM, Hess WR, Wilde A (2006) An internal antisense RNA regulates expression of the photosynthesis gene *isiA*. *Proc Natl Acad Sci USA* 103(18):7054–7058.
63. Nakamura Y, Gojohori T, Ikemura T (2000) Codon usage tabulated from international DNA sequence databases: Status for the year 2000. *Nucleic Acids Res* 28(11):292.
64. Paquin B, Kathe SD, Nierzwicki-Bauer SA, Shub DA (1997) Origin and evolution of group I introns in cyanobacterial tRNA genes. *J Bacteriol* 179(21):6798–6806.
65. Rudi K, Jakobsen KS (1999) Complex evolutionary patterns of tRNA Leu(UAA) group I introns in the cyanobacterial radiation [corrected]. *J Bacteriol* 181(11):3445–3451.
66. Axmann IM, Hertel S, Wiegand A, Dörrich AK, Wilde A (2014) Diversity of KaiC-based timing systems in marine Cyanobacteria. *Mar Genomics* 14:3–16.
67. Costa J-L, Paulsrud P, Lindblad P (2002) The cyanobacterial tRNA(Leu)(UAA) intron: Evolutionary patterns in a genetic marker. *Mol Biol Evol* 19(6):850–857.
68. Deutschbauer A, et al. (2014) Towards an informative mutant phenotype for every bacterial gene. *J Bacteriol* 196(20):3643–3655.
69. Hottes AK, et al. (2013) Bacterial adaptation through loss of function. *PLoS Genet* 9(7):e1003617.
70. 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*, eds Burnap R, Vermaas W (Springer, Dordrecht, The Netherlands), pp 119–137.
71. Simkovsky R, et al. (2012) Impairment of O-antigen production confers resistance to grazing in a model amoeba-cyanobacterium predator-prey system. *Proc Natl Acad Sci USA* 109(41):16678–16683.
72. Allen MM (1968) Simple conditions for growth of unicellular blue-green algae on plates. *J Phycol* 4(1):1–4.
73. Kent WJ (2002) BLAT—the BLAST-like alignment tool. *Genome Res* 12(4):656–664.
74. Barquist L, et al. (2013) A comparison of dense transposon insertion libraries in the *Salmonella* serovars Typhi and Typhimurium. *Nucleic Acids Res* 41(8):4549–4564.
75. Holtman CK, et al. (2005) High-throughput functional analysis of the *Synechococcus elongatus* PCC 7942 genome. *DNA Res* 12(2):103–115.

Supporting Information

Rubin et al. 10.1073/pnas.1519220112

SI Results

As an additional measure of bias in our calls of essential genes, we assayed these calls for polar effects, in which transposon insertions disrupt expression of downstream genes. We reasoned that, if there were strong polar effects, mutations in a nonessential gene would be lethal if they stopped expression of the downstream essential gene in the transcript; thus, we would not expect to find an operon with a nonessential gene upstream of an essential gene.

To check for this possibility, we examined all adjacent genes within annotated operons (45) for essentiality. There were 87 instances of nonessential genes upstream from essential genes and 125 instances of essential genes upstream of nonessential genes. The similar frequency of these two arrangements indicates that there are not pervasive polar effects of the transposon, and conclusions of essentiality within an operon can be made for individual genes.

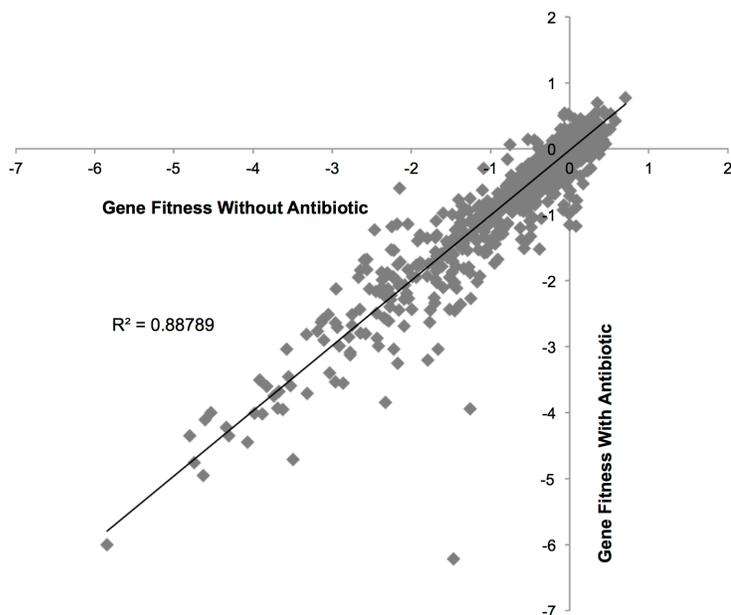


Fig. S1. Divergence of antibiotic and no antibiotic library outgrowths. Each gene's fitness score is plotted after outgrowth with kanamycin (*y* axis) or without kanamycin (*x* axis) for seven generations. Negative values represent genes that had negative effects when mutated (beneficial genes), and positive values represent genes that had positive effects when mutated (disadvantageous genes).

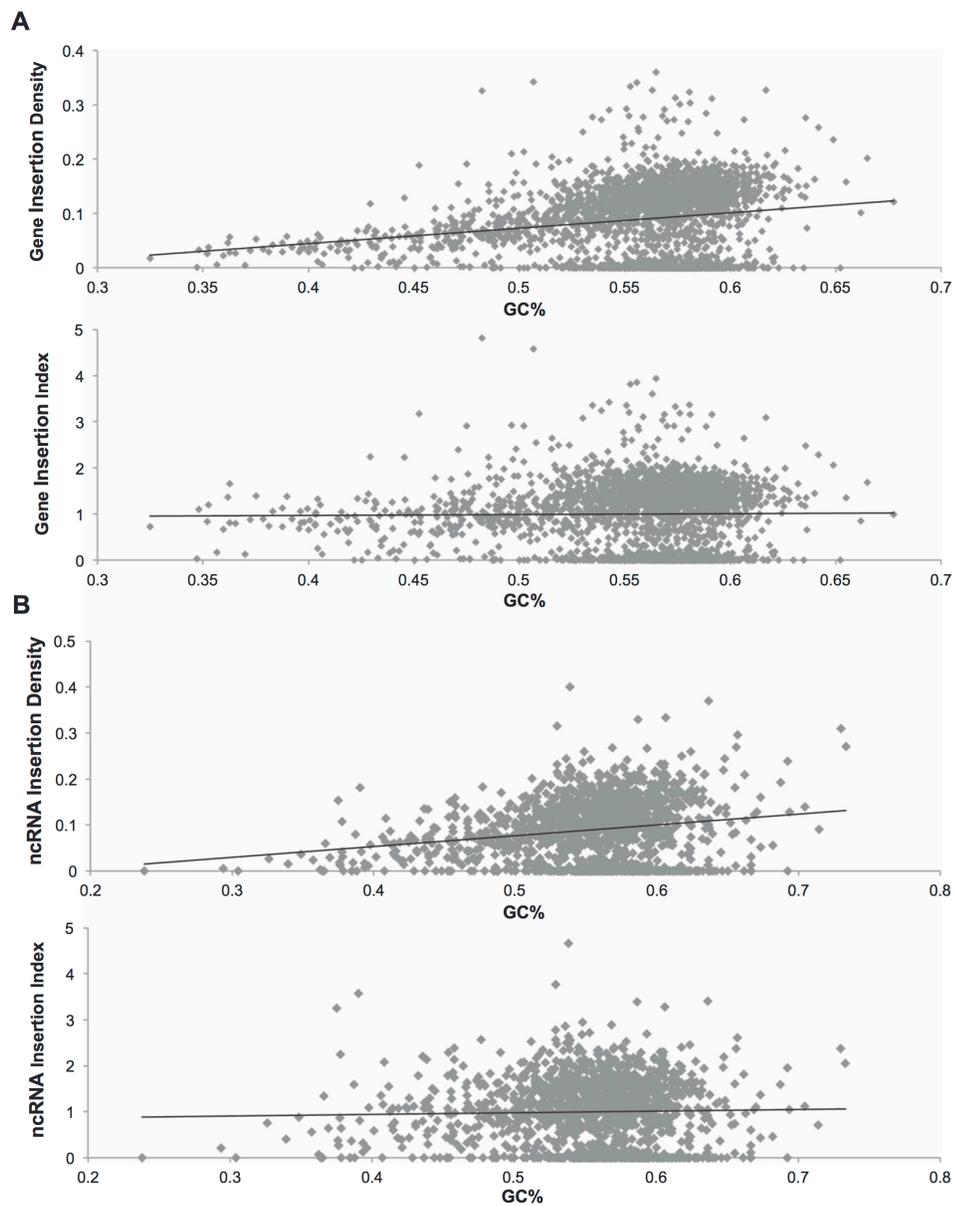


Fig. S3. Normalizing for GC bias of insertion density. *A, Upper* shows the insertion density (y axis) for each gene compared with its GC% (x axis) before normalization. These values were normalized by dividing each insertion density by the trend line value for its GC%. *A, Lower* shows the insertion density normalized for GC% (insertion index). *B, Upper* shows the insertion density for each of the recently annotated ncRNAs (45) compared with its GC% before normalization. *B, Lower* shows the insertion density normalized for GC% (insertion index).

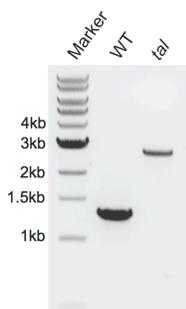


Fig. S4. Genotypic characterization of the recreated *taI* mutant. Lane 1, standard 1-kb ladder (New England Biolabs); lane 2, amplification of WT DNA with primers surrounding the *taI* gene; lane 3, amplification of *taI* mutant (UGS-3-C-11), which carries a 1.3-kb insertion, with the same primers. Each band is representative of three colonies tested.

Table S1. The major variations of techniques used for conjugation

Conjugation light intensity ($\mu\text{mol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$)	Ratio (mL:mL) of <i>Synechococcus elongatus</i> : <i>Escherichia coli</i>	Conjugation medium	Conjugation time (h)	Efficiency*
10–40	5:1	Filter (directly transferred to selective plate)	6–8	1.1×10^{-8} ($n = 1$)
10–40	2:1	Filter	6–8	1.5×10^{-8} ($n = 1$)
10–40	10:1	Filter	24	2.2×10^{-8} ($n = 1$)
10–40	10:1	Filter	6–8	3.3×10^{-8} ($n = 1$)
10–40	10:1	Agar transferred by wash to selective plate	24	4.4×10^{-8} ($n = 2$)
10–40	10:1	Agar transferred by wash to selective plate	6–8	4.7×10^{-8} ($n = 2$)
100–140	5:1	Filter	24	2.1×10^{-7} ($n = 1$)
100–140	5:1	Filter	6–8	2.1×10^{-7} ($n = 1$)
100–140	2:1	Filter	24	2.8×10^{-7} ($n = 1$)
10–40	1:1	Filter	6–8	3.0×10^{-7} ($n = 2$)
100–140	2:1	Filter	6–8	3.3×10^{-7} ($n = 1$)
10–40	1:1	Filter	24	5.0×10^{-7} ($n = 2$)
10–40	1:1	Agar with selective agent underlayed	24	7.8×10^{-7} ($n = 1$)
10–40	1:1	Filter	48	1.1×10^{-6} ($n = 1$)
100–140	1:1	Filter	24	2.0×10^{-6} ($n = 1$)
100–140	1:1	Filter	6–8	3.9×10^{-6} ($n = 4$) [†]

*Colony forming units (CFU) of conjugated mutants per CFUs of *S. elongatus* culture.

[†]The conditions used for generation of the RB-TnSeq library.

Table S2. Essential genes in central carbon metabolism

<i>Synechococcus elongatus</i> identification	Name	Pathway*	Essentiality	Conserved [†]	Redundant [‡]
SynPCC7942_2297	tal	Pentose phosphate pathway	Nonessential	1	0
SynPCC7942_0538	tktA	Pentose phosphate pathway	Essential	1	0
SynPCC7942_0604	cbbE	Pentose phosphate pathway	Essential	1	0
SynPCC7942_0584	rpiA	Pentose phosphate pathway	Essential	1	0
SynPCC7942_0529	pgl	Pentose phosphate pathway	Nonessential	1	0
SynPCC7942_2334	zwf	Pentose phosphate pathway	Nonessential	1	0
SynPCC7942_0039	gnd	Pentose phosphate pathway	Nonessential	1	0
SynPCC7942_0098	pyk	Glycolysis	Essential	1	0
SynPCC7942_0639	eno	Glycolysis	Essential	1	0
SynPCC7942_0469	pgmM	Glycolysis	Essential	1	1
SynPCC7942_1116	pgk	Glycolysis	Essential	1	0
SynPCC7942_1939	gap3	Glycolysis	Nonessential	0	1
SynPCC7942_0245	gap1	Glycolysis	Nonessential	0	1
SynPCC7942_1261	tpiA	Glycolysis	Essential	1	0
SynPCC7942_1443	cbbA	Glycolysis	Essential	1	0
SynPCC7942_0592	prfA	Glycolysis	Nonessential	0	0
SynPCC7942_2029	pgi	Glycolysis	Essential	1	0
SynPCC7942_0781	ppsA	Glycolysis	Nonessential	0	0
SynPCC7942_2335	fbp	Glycolysis	Nonessential	0	1
SynPCC7942_0505	fbpl	Glycolysis	Essential	1	1
SynPCC7942_0485	pgam1	Glycolysis	Essential	1	1
SynPCC7942_1516	pgam2	Glycolysis	Essential	0	1
SynPCC7942_2078	pgam3	Glycolysis	Nonessential	0	1
SynPCC7942_0612	glcA	TCA cycle	Essential	1	0
SynPCC7942_0903	acnB	TCA cycle	Essential	0	0
SynPCC7942_1007	fumC	TCA cycle	Nonessential	1	0
SynPCC7942_1719	icd	TCA cycle	Essential	0	0
SynPCC7942_1533	sdhB	TCA cycle	Nonessential	0	0
SynPCC7942_1427	cbbS	Calvin-Benson-Bassham cycle	Essential	1	0
SynPCC7942_1426	cbbL	Calvin-Benson-Bassham cycle	Essential	1	0
SynPCC7942_1116	pgk	Calvin-Benson-Bassham cycle	Essential	1	0
SynPCC7942_1742	gap2	Calvin-Benson-Bassham cycle	Essential	1	0
SynPCC7942_1261	tpiA	Calvin-Benson-Bassham cycle	Essential	1	0
SynPCC7942_1443	cbbA	Calvin-Benson-Bassham cycle	Essential	1	0
SynPCC7942_2335	fbp	Calvin-Benson-Bassham cycle	Nonessential	0	1
SynPCC7942_0538	tktA	Calvin-Benson-Bassham cycle	Essential	1	0
SynPCC7942_0505	fbpl	Calvin-Benson-Bassham cycle	Essential	1	1
SynPCC7942_0604	cbbE	Calvin-Benson-Bassham cycle	Essential	1	0
SynPCC7942_0977	prk	Calvin-Benson-Bassham cycle	Essential	0	0
SynPCC7942_0584	rpiA	Calvin-Benson-Bassham cycle	Essential	1	0

*Metabolic pathway as taken from the BioCyc database (31).

[†]Refers to conservation among all 13 of the cyanobacterial genomes studied in the work by Shi and Falkowski (26).

[‡]Functional redundancy as shown in the BioCyc database (31).

Table S3. Essential genes for the photosynthetic lifestyle

<i>Synechococcus elongatus</i> identification	Name	Description	Cyano conserved*
SynPCC7942_0027	bioF	8-Amino-7-oxononanoate synthase	No
SynPCC7942_0203	ribD	5-Amino-6-(5-phosphoribosylamino)uracil reductase...	Yes
SynPCC7942_0222		Hypothetical protein	No
SynPCC7942_0226		Sec-independent protein translocase Tata	No
SynPCC7942_0257		Protein of unknown function DUF92, transmembrane	Yes
SynPCC7942_0297	ftsH	FtsH peptidase homolog, chloroplast; metallopeptidase	No
SynPCC7942_0320	galE	UDP-galactose 4-epimerase	No
SynPCC7942_0322	ycf44	c-Type cytochrome biogenesis protein	Yes
SynPCC7942_0323	ccdA	Cytochrome c biogenesis protein-like	Yes
SynPCC7942_0330	atp1	Hypothetical protein	Yes
SynPCC7942_0333	atpG	F0F1 ATP synthase subunit B'	Yes
SynPCC7942_0439	chlM	Mg-protoporphyrin IX methyl transferase	Yes
SynPCC7942_0450	nifU	Putative NifU-like protein	Yes
SynPCC7942_0485	gpmB	Phosphoglycerate mutase	Yes
SynPCC7942_0492	ribF	Bifunctional riboflavin kinase/FMN adenyltransferase	Yes
SynPCC7942_0537	fabF	3-Oxoacyl-(acyl carrier protein) synthase II	Yes
SynPCC7942_0560		ATPase	No
SynPCC7942_0584	rpiA	Ribose-5-phosphate isomerase A	Yes
SynPCC7942_0632	rplJ	50S ribosomal protein L10	Yes
SynPCC7942_0684	fabG	3-oxoacyl-[acyl-carrier-protein] reductase	Yes
SynPCC7942_0694	rps1	30S ribosomal protein S1	Yes
SynPCC7942_0775		Hypothetical protein	No
SynPCC7942_0853	dapL1	L,L-diaminopimelate aminotransferase	No
SynPCC7942_0884	EF-Tu	Elongation factor Tu	Yes
SynPCC7942_0894	aroK	Shikimate kinase	Yes
SynPCC7942_0897	minE	Cell division topological specificity factor MinE	Yes
SynPCC7942_0912	dnaX	DNA polymerase III, τ -subunit	Yes
SynPCC7942_0928		Outer envelope membrane protein	No
SynPCC7942_0942	ftsH	FtsH peptidase homolog, chloroplast, metallo peptidase	No
SynPCC7942_0959	obgE	GTPase ObgE	Yes
SynPCC7942_0977	prk	Phosphoribulokinase	No
SynPCC7942_0978	petH	Ferredoxin-NADP oxidoreductase	Yes
SynPCC7942_0988	ycf54	Conserved hypothetical protein YCF54	No
SynPCC7942_1002	psaD	Photosystem I reaction center subunit II	Yes
SynPCC7942_1068	pdhC	Branched-chain α -keto acid dehydrogenase subunit E2	Yes
SynPCC7942_1083		Probable glycosyltransferase	Yes
SynPCC7942_1086	hemE	Uroporphyrinogen decarboxylase	Yes
SynPCC7942_1198	phdD	Dihydrolipoamide dehydrogenase	Yes
SynPCC7942_1232	petC	Cytochrome b6-f complex iron-sulfur subunit	Yes
SynPCC7942_1274	ycf37	TPR repeat	Yes
SynPCC7942_1351		HAD-superfamily hydrolase subfamily IA, variant 3	No
SynPCC7942_1359		Coenzyme F420 hydrogenase	No
SynPCC7942_1427	rbcS	Ribulose 1,5-bisphosphate carboxylase small subunit	Yes
SynPCC7942_1457	plsC	1-Acyl-sn-glycerol-3-phosphate acyltransferase	Yes
SynPCC7942_1497		Hypothetical protein	No
SynPCC7942_1499	petF	Ferredoxin (2Fe-2S)	No
SynPCC7942_1520	rpsT	30S ribosomal protein S20	Yes
SynPCC7942_1590		Hypothetical protein	No
SynPCC7942_1673		Hypothetical protein	Yes
SynPCC7942_1734	ftcC	Ferredoxin-thioredoxin reductase catalytic chain	No
SynPCC7942_1737	sufD	Iron-regulated ABC transporter permease protein SufD	Yes
SynPCC7942_1830	trxA	Thioredoxin	Yes
SynPCC7942_1883	ycf53	Conserved hypothetical protein YCF53	No
SynPCC7942_1907	acsF	Mg-protoporphyrin IX monomethyl ester (oxidative) cyclase	No
SynPCC7942_1959	proS	Prolyl-tRNA synthetase	Yes
SynPCC7942_1983	pds	ζ -Carotene desaturase/three-step phytoene desaturase	No
SynPCC7942_2017		Hypothetical protein	Yes
SynPCC7942_2062	crtl	Lycopene cyclase (Crtl-type)	No
SynPCC7942_2084	chlG	Bacteriochlorophyll/chlorophyll a synthase	Yes
SynPCC7942_2089		Thioredoxin domain 2	Yes
SynPCC7942_2113	kpr5	Ribose-phosphate pyrophosphokinase	Yes
SynPCC7942_2117	gatA	Aspartyl/glutamyl-tRNA amidotransferase subunit A	Yes

Table S3. Cont.

<i>Synechococcus elongatus</i> identification	Name	Description	Cyano conserved*
SynPCC7942_2136	dapB	Dihydrodipicolinate reductase	Yes
SynPCC7942_2274	chiD	Protoporphyrin IX magnesium-chelatase	Yes
SynPCC7942_2312	murG	UDP- <i>N</i> -acetylglucosamine- <i>N</i> -acetylmuramyl-(pentapeptide)...	Yes
SynPCC7942_2330	ctpA	C-terminal processing peptidase-2; serine peptidase	No
SynPCC7942_2359	nha3	Na ⁺ /H ⁺ antiporter	No
SynPCC7942_2415	lysS	Lysyl-tRNA synthetase	Yes
SynPCC7942_2503	por	Protochlorophyllide oxidoreductase	Yes
SynPCC7942_2524	tig	Trigger factor	Yes
SynPCC7942_2531	tsf	Elongation factor Ts	Yes
SynPCC7942_2537	clpPIII	ATP-dependent Clp protease proteolytic subunit	Yes
SynPCC7942_2538	clpR	ATP-dependent Clp protease-like protein	Yes
SynPCC7942_2581	petF	Ferredoxin (2Fe-2S)	Yes

5. *elongatus* essential genes conserved in the green lineage (plants and green algae) but not in nonphotosynthetic organisms as determined using the greencut2 dataset (30). GTPase, guanosine triphosphate hydrolase.

*Refers to conservation among all 13 of the cyanobacterial genomes studied in the work by Shi and Falkowski (26).

Other Supporting Information Files

[Dataset S1 \(XLSX\)](#)

[Dataset S2 \(XLSX\)](#)

[Dataset S3 \(XLSX\)](#)

[Dataset S4 \(XLSX\)](#)

[Dataset S5 \(XLSX\)](#)

[Dataset S6 \(XLSX\)](#)

3.3 RB-TnSeq Guided Metabolic Modeling



Unique attributes of cyanobacterial metabolism revealed by improved genome-scale metabolic modeling and essential gene analysis

Jared T. Broddrick^{a,b,1}, Benjamin E. Rubin^{b,c,1}, David G. Welkie^{c,1}, Niu Du^{d,e}, Nathan Mih^f, Spencer Diamond^{b,2}, Jenny J. Lee^b, Susan S. Golden^{b,c,3}, and Bernhard O. Palsson^{a,f}

^aDepartment of Bioengineering, University of California, San Diego, La Jolla, CA 92093; ^bDivision of Biological Sciences, University of California, San Diego, La Jolla, CA 92093; ^cCenter for Circadian Biology, University of California, San Diego, La Jolla, CA 92093; ^dScripps Institution of Oceanography, University of California, San Diego, La Jolla, CA 92093; ^eMicrobial and Environmental Genomics, J. Craig Venter Institute, La Jolla, CA 92037; and ^fBioinformatics and Systems Biology Graduate Program, University of California, San Diego, La Jolla, CA 92093

Contributed by Susan S. Golden, October 25, 2016 (sent for review August 12, 2016; reviewed by Shota Atsumi and Thomas K. Wood)

The model cyanobacterium, *Synechococcus elongatus* PCC 7942, is a genetically tractable obligate phototroph that is being developed for the bioproduction of high-value chemicals. Genome-scale models (GEMs) have been successfully used to assess and engineer cellular metabolism; however, GEMs of phototrophic metabolism have been limited by the lack of experimental datasets for model validation and the challenges of incorporating photon uptake. Here, we develop a GEM of metabolism in *S. elongatus* using random barcode transposon site sequencing (RB-TnSeq) essential gene and physiological data specific to photoautotrophic metabolism. The model explicitly describes photon absorption and accounts for shading, resulting in the characteristic linear growth curve of photoautotrophs. GEM predictions of gene essentiality were compared with data obtained from recent dense-transposon mutagenesis experiments. This dataset allowed major improvements to the accuracy of the model. Furthermore, discrepancies between GEM predictions and the in vivo dataset revealed biological characteristics, such as the importance of a truncated, linear TCA pathway, low flux toward amino acid synthesis from photorespiration, and knowledge gaps within nucleotide metabolism. Coupling of strong experimental support and photoautotrophic modeling methods thus resulted in a highly accurate model of *S. elongatus* metabolism that highlights previously unknown areas of *S. elongatus* biology.

cyanobacteria | constraint-based modeling | TCA cycle | photosynthesis | *Synechococcus elongatus*

The unicellular cyanobacterium *Synechococcus elongatus* PCC 7942 is being developed as a photosynthetic bioproduction platform for an array of industrial products (1–3). This model strain is attractive for this purpose because of its genetic tractability (4) and its reliance on mainly CO₂, H₂O, and light for metabolism, reducing the environmental and economic costs of cultivation. For low-cost, high-volume products, such as biofuels, however, one of the biggest challenges is attaining profitable product yields (5, 6). Genome-scale models (GEMs) of metabolism provide a valuable tool for increasing product titers by optimizing yield in silico and then, reproducing the changes in vivo (7). For instance, GEMs were used to select the optimal synthetic pathway for 3-hydroxypropanoate biosynthesis in *Saccharomyces cerevisiae* (8). In *Escherichia coli*, GEM optimization was used to realize heterologous production of 1,4-butanediol synthesis and increase titers three orders of magnitude (9). Although there have been numerous modeling efforts in *Synechocystis* sp. PCC 6803 (here in referred to as PCC 6803), this organism is highly divergent from *S. elongatus*, where limited modeling has been done (10).

This deficit can partially be explained by the lack of in vivo validation datasets, such as ¹³C metabolic flux analysis (MFA), for obligate phototrophs (11). Development of metabolic models of *S. elongatus* with strong experimental support is necessary to

exploit the organism as a bioproduction platform and advance models of obligate phototrophic metabolism.

A metabolic network reconstruction is a representation of all metabolic reactions, the enzymes responsible for their catalysis, and the genes that encode them. Genome-scale reconstructions have a proven record of contextualizing organism-specific information and facilitating the characterization and engineering of cellular metabolism (12, 13). When complete, the reconstruction enables quantitative prediction of metabolic phenotypes represented as reaction fluxes. The overall predictive power of a GEM is naturally dependent on its quality (14). Essentiality datasets have been successfully used to increase the accuracy of GEMs (15). We recently determined genome-wide gene essentiality by screening ~250,000 pooled mutants for their survival under standard laboratory conditions with continuous light via random barcode transposon site sequencing (RB-TnSeq) (16). This dataset facilitated the generation and testing of a high-quality genome-scale reconstruction through comparison of the model outputs and in vivo phenotypes at the genome scale. Inconsistencies between model predictions and in vivo data can highlight parts of *S. elongatus* metabolism where current understanding is inadequate (17).

Another key characteristic of an accurate GEM is the application of constraints that place physical, chemical, and biological limitations on a culture and generate biologically relevant phenotypic predictions. Incorporating light, a dominant constraint

Significance

Genome-scale models of metabolism are important tools for metabolic engineering and production strain development. We present an experimentally validated and manually curated model of metabolism in *Synechococcus elongatus* PCC 7942 that (i) leads to discovery of unique metabolic characteristics, such as the importance of a truncated, linear TCA pathway, (ii) highlights poorly understood areas of metabolism as exemplified by knowledge gaps in nucleotide salvage, and (iii) accurately quantifies light input and self-shading. We now have a metabolic model that can be used as a basis for metabolic design in *S. elongatus*.

Author contributions: J.T.B., B.E.R., D.G.W., S.D., S.S.G., and B.O.P. designed research; J.T.B., B.E.R., D.G.W., N.D., N.M., and J.J.L. performed research; J.T.B., B.E.R., D.G.W., N.D., N.M., and J.J.L. analyzed data; and J.T.B., B.E.R., D.G.W., S.S.G., and B.O.P. wrote the paper.

Reviewers: S.A., University of California, Davis; and T.K.W., Pennsylvania State University. The authors declare no conflict of interest.

¹J.T.B., B.E.R., and D.G.W. contributed equally to this work.

²Present address: Department of Earth and Planetary Science, University of California, Berkeley, CA 94704.

³To whom correspondence should be addressed. Email: sgolden@ucsd.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1613446113/-DCSupplemental.

on phototrophic growth, into a GEM remains a challenge (18). Photon uptake is typically fixed based on experimental results, an approach that allows retrospective analysis but not predictive modeling (19). Therefore, no current model inputs light quantity, quality, and shading, resulting in a linear growth curve characteristic of photoautotrophic batch culture.

Here, we present a comprehensive GEM of obligate phototrophic metabolism. We performed a complete reannotation and reconstruction of metabolic genes in the *S. elongatus* genome and developed an approach to incorporate light absorption that factors in the effects of cell shading. In addition, GEM predictions have been compared with and improved by essentiality data (16). These comparisons are also used to reveal unique attributes of the organism's metabolism. The result is a comprehensive metabolic model of *S. elongatus* metabolism.

Results

Genome-Scale Reconstruction of Phototrophic Metabolism. A metabolic reconstruction is a knowledge base that places biochemical, genetic, and genomic information into a structured framework. The reconstruction contains the functional annotation of the genome and defines the organism's metabolic capability: the substrates that it can use and the reactions that it can perform. To properly define the metabolic capability of *S. elongatus*, we curated the genome annotation, leveraged state of the art in silico methods, incorporated comprehensive in vivo essentiality data, and included a detailed reconstruction of light harvesting. The resulting model is an organized collection of the extensive data available for *S. elongatus* in a format that enables accurate predictions of phototrophic metabolism (Fig. S1).

Manual curation of *S. elongatus* genome annotation. Because the metabolic capability reflected in a reconstructed network is dependent on the functional annotation of the organism, we reannotated the metabolic genes using amino acid as well as protein functional domain homology-based methods. This functional reannotation enabled the conversion of an initial draft reconstruction to a completed GEM (20) (*SI Materials and Methods*). Of 2,723 genes in the *S. elongatus* genome, 785 (29%) were included in the final version of the model, and 118 of these genes (15%) had updated functional annotations. This GEM is named *iJB785* (model files are in [Dataset S1](#), and the Excel file is in [Dataset S2](#)) following convention (21).

Protein structure-guided reconstruction. Amino acid and protein domain-based annotations often do not provide sufficient detail to assign enzyme function. Therefore, enzymes are often incorrectly annotated as functionally equivalent (isozymes). In *S. elongatus*, this challenge resulted in instances where multiple genes that are essential in vivo were assigned to the same reaction; however, if the enzymes were truly compensatory, none should be essential. The importance of enzyme structure in catalytic activity suggests that structural homology modeling, which uses in silico-derived 3D analysis of a target protein based on the crystal structure of a similar enzyme, may provide additional insight into protein function.

Protein structure data have recently been applied in the global analysis of GEMs (22). We set out to apply structural modeling to protein annotation by determining the functional difference between four annotated phosphoglycerate mutases (PGMs; Synpcc7942_2078, Synpcc7942_1516, Synpcc7942_0485, EC 5.4.2.11; and Synpcc7942_0469, EC 5.4.1.12) in the *S. elongatus* genome, three of which are essential in vivo. Previous work in *S. elongatus* suggested that multiple PGMs work in concert to regulate metabolic flux during shifts in CO₂ availability (23). However, it seemed unlikely that three of four PGMs would be essential for regulation in a stable CO₂ environment. To test the hypothesis of divergent functions between the *S. elongatus* PGMs, structural homology models were generated and compared with published control crystal structures (*SI Materials and Methods*, [Table S1](#)).

Based on the structural comparison, it was possible to ascribe a more detailed function to each of the annotated PGMs. The Synpcc7942_0469 protein is structurally distinct from the three other PGMs and was annotated as the primary glycolytic PGM in *S. elongatus* based on its canonical PGM structure and the fact that it is essential. Previous work indicated that two PGMs are required to regulate central carbon flux during a transition from high to low CO₂ (23). The Synpcc7942_2078 protein shares structural features with an *E. coli* PGM control but is nonessential; thus, it was annotated as a PGM performing this regulatory function. Synpcc7942_0485 shares strong structural similarity to a phosphoserine phosphatase (PSP) in *Hydrogenobacter thermophilus* (24) and has sequence homology to the recently characterized PSP in PCC 6803 (25). Thus this gene was confidently annotated as a PSP in *S. elongatus*. Synpcc7942_1516, however, has structural features that could not be classified as a traditional PGM or PSP and is essential in vivo. Based on genomic neighborhood analysis and transcriptome mapping data (26), we hypothesized that Synpcc7942_1516 plays a regulatory role in an uncharacterized signaling network. As a regulatory enzyme, it fell outside the scope of the metabolic model. These results indicate that structural homology modeling is a promising annotation tool to increase the quality of genome-scale reconstructions and hypothesize enzyme function.

Improved reconstruction through incorporation of essential gene data.

The essential gene calls for *S. elongatus* determined by RB-TnSeq-enabled refinement of the gene reaction annotations during development of the reconstruction (16). This in vivo dataset provides a gauge of gene importance by identifying genes that cannot sustain insertion mutants, which are interpreted to be essential, and the growth rate of those that can, which are interpreted to be beneficial or nonessential. Alternatively, in silico essentiality calls are made by quantifying the impact on growth when the flux through each enzyme in the model is independently set to zero (*SI Materials and Methods*). Discrepancies between essential gene calls in the in silico draft model and the RB-TnSeq results were investigated, and with sufficient evidence, the gene assignment for the model reaction was updated accordingly. For example, the *S. elongatus* genome encodes two annotated uroporphyrinogen methyltransferases (Synpcc7942_0271 and Synpcc7942_2610, EC 2.1.1.107), catalyzing an early step of both vitamin B12 and siroheme biosynthesis. However, both genes are essential in vivo, suggesting that they are not compensatory. Genomic neighborhood analysis indicated that *synpcc7942_2610* is adjacent to an iron chelatase gene; thus, we proposed that it is dedicated to the biosynthesis of siroheme, not vitamin B12. However, discrepancies between the in silico and in vivo essential gene data were not forced into agreement without additional evidence, and these remaining discrepancies are contained in [Dataset S3](#). For example, *S. elongatus* has two genes annotated for type II NADH oxidoreductases (*synpcc7942_0101* and *synpcc7942_0198*, EC 1.6.5.9); however, the in vivo data indicated that one is essential (*synpcc7942_0101*) and thus, the enzymes are not redundant. Nevertheless, the manual curation process did not reveal any significant difference between the two genes, and they were annotated as isozymes, although the RB-TnSeq data suggest divergent functions.

Discrepancies also led to a more complete representation of cellular biomass. For example, alkanes were not initially included in the biomass, because their function is unknown. However, their synthesis is essential. Recently, it was discovered that alkanes can play a role in cyclic electron flow in PCC 6803 (27). Hypothesizing a homologous function in *S. elongatus*, they were added to the biomass equation. Thus, applying the essential gene data to the curation process increased the quality of the reconstruction.

Explicit modeling of light absorption. The dominant constraints on photoautotrophic growth are light and CO₂ availability. A fundamental barrier to predictive modeling of photoautotrophic growth has been the inability to translate light irradiance into a metabolite. Traditionally, light uptake has been inferred and not

explicitly determined. A two-step optimization method is commonly used, where the CO₂ uptake rate is fixed based on experimental values at the observed growth rate followed by minimization of the photon flux (19, 28). However, the resulting values reflect only the photons that perform metabolic work, not the totality of absorbed light. Excess photon absorption has a significant impact on growth rate and metabolism. Reactive oxygen species, such as singlet oxygen generated in the pigment antenna and photosystem II (29) as well as superoxide at photosystem I (30), damage the photosynthetic apparatus and consume metabolic resources. Thus, the two-step optimization method does not accurately account for total light absorption and is incapable of predictive modeling of phototrophic metabolism.

In a more extensive description of photon capture, a metabolic reconstruction of *Chlamydomonas reinhardtii* accounted for light source quality (31). This approach was more mechanistic than the two-step optimization but still did not consider cellular pigmentation as a factor in photon absorption, and therefore, light uptake could not be quantified into a typical modeling input flux. In another paper, a light distribution function was combined with flux-balance analysis to model cyanobacterial growth in photobioreactors; however, a mechanistic model of light harvesting was not included (32). Our approach goes further by combining light source irradiance with *in vivo* absorption to define photon use from measurements of incident light. We incorporated the chlorophyll-normalized optical absorption cross-section (33), derived from the *in vivo* absorption spectrum, to link photon uptake to cellular composition. Because the model's biomass explicitly defines the cellular composition, the biomass-normalized photon absorption rate was calculated from the combination of irradiance, optical absorption cross-section, and the chlorophyll component of the biomass. This approach enabled comparison between the photon absorption capacity of the cell and the photon delivery rate of the light source at a given irradiance (Fig. 1).

Accurate modeling of obligate phototrophic metabolism also required a new level of detail in the reconstruction of the photosystem. We used recent proteomics (34) and fluorescence microscopy data (35) to reconcile membrane localization of electron transport complexes that previous cyanobacterial models had included inaccurately in the cytoplasmic membrane. Additionally, the ferredoxin:plastoquinone oxidoreductase complex gene associations were updated to include additional subunits and the use of ferredoxin as the electron donor (36, 37). We also included the photoinactivation of the D1 subunit of photosystem II. Using the photodamage rates in PCC 6803 (38), we were able to calculate a D1 repair metabolic cost as a proportion of flux through photosystem II. Finally, we generated stoichiometric reactions accounting for the energy transfer efficiencies of each of the photosynthetic pigments. Targeted excitation of *S. elongatus* permaplasts provided relative efficiency metrics for the transfer of energy from a given pigment to the photosystems (39). Reactions including these efficiencies enabled the model to account for light spectrum-specific photosynthetic efficiency. This comprehensive reconstruction of light gathering set the framework for accurate constraint-based modeling of phototrophic metabolism.

The completed reconstruction, *iJB785*, consists of 785 genes, 850 metabolic and transport reactions, and 768 nonunique metabolites distributed over seven cellular compartments (Dataset S2). The reconstruction was completed in the BiGG Models format (40), enabling standardization and cross-referencing to external databases (bigg.ucsd.edu). Combining an updated annotation, whole-genome RB-TnSeq data, and an advanced representation of light harvesting resulted in a comprehensive phototrophic model.

Modeling Phototrophic Growth. Conversion of a reconstruction into a mathematical model and the subsequent application of biologically relevant constraints enable the simulation of cellular

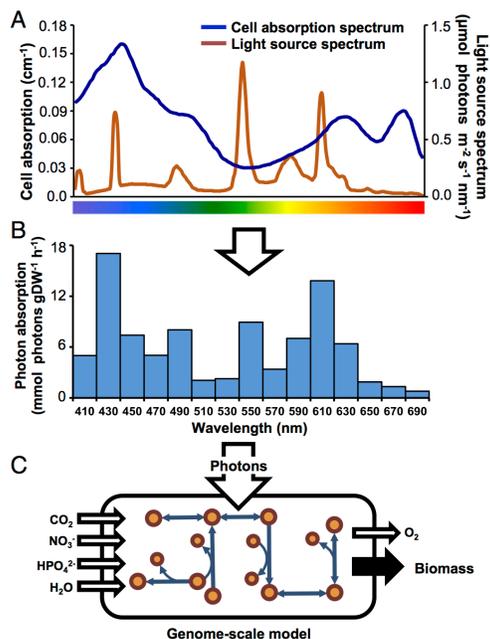


Fig. 1. Deriving light uptake rates for the GEM. (A) The chlorophyll *a*-normalized optical absorption cross-section was calculated from the cellular absorbance in the photosynthetically active range (400–700 nm) and compared with the spectral distribution of the incident light. (B) The photon absorption rate as a function of cell biomass [millimoles photons gram dry weight (DW)⁻¹ hour⁻¹] was determined by combining the optical absorption cross-section and incident light. (C) The GEM was constrained by setting the photon uptake flux to the calculated absorption rate, which was split into 15 20-nm bins across the photosynthetic range.

phenotypes. Modeling cellular growth is typically computed as either yield or specific growth rate, which assumes exponential growth (41). In both cases the inputs are normalized, such that the simulation reports a single value representative of cellular growth as long as there is a constant ratio between the biomass and the input flux. However, as the culture becomes denser, the photon absorption capacity can exceed the amount of light delivered by the light source and in the process, alter the ratio of uptake to biomass. When this phenomenon occurs, it results in a linear growth curve caused by self-shading–induced light limitation. Cells closest to the light source absorb excess photons, preventing cells in the inner culture from achieving their maximum growth rate. We accounted for shading with an unsteady-state growth modeling methodology (42), which resulted in an accurate model of linear photoautotrophic growth over the duration of a batch culture.

An additional constraint was required to capture the point when light absorption became excessive. As the other dominant constraint on growth, the maximum carbon uptake rate marks the transition between a light- and a carbon-limited culture. It is dependent on the availability of inorganic carbon in the media (43) and the acclimated state of the cell (44). However, an approximation of maximum photosynthetic output, a proxy for carbon uptake, can be captured in a single parameter: the oxygen evolution

rate (45). We used experimentally determined oxygen evolution rates to constrain the maximum photosynthetic output at a given irradiance (Fig. S2). Excess photon absorption was allowed to leave the system in a manner that did not incur a metabolic cost, simulating loss as heat or fluorescence. Reactive oxygen species production in the light-harvesting antenna caused by excess light is currently not modeled.

A growth curve for a typical *S. elongatus* culture was simulated using the oxygen evolution constraint and the calculated photon uptake rate. To account for self-shading, at 1-h intervals, the flask was sectioned into 50 concentric rings, with each ring modeling the biomass production of a 2% fraction of the culture. Light was modeled from the side of the flask, and the photon absorption of an outer ring was made unavailable to the remainder of the inner rings (Fig. S3). For the nonshading simulation, the growth rate matched the in vivo culture until light limitation, at which point the in silico growth remained exponential, whereas the in vivo curve became linear (Fig. 2). Growth simulations factoring in self-shading transitioned into linear phase on light limitation, characteristic of in vivo growth. This more accurate prediction of photoautotrophic growth was made possible by the combination of modeling photon uptake and shading as a function of culture density.

Model parameters are specific to the cellular phenotype: in particular, the chlorophyll-normalized optical absorption cross-section, which depends on the photoacclimation state of the cell. The primary photon-harvesting complex in *S. elongatus*, the phycobilisome, which can efficiently deliver light energy to both photosystems (39), is highly adaptive (46) but devoid of chlorophyll. Photosystem I contains 80–95% of the chlorophyll *a* in *S. elongatus* (39). Because photoacclimation causes fluctuation in the photosystem I to photosystem II ratio and changes to the phycobilisome (47), a given chlorophyll-normalized optical absorption cross-section may no longer be representative. The model can be easily reparameterized to account for this adaptation. Additionally, the oxygen evolution rate can be adjusted to account for phenotypic changes in different carbon environments. The mechanistic nature of this approach is generally applicable to any phototrophic reconstruction and expands genome-scale modeling into new phenotypes, such as photoacclimation and with additional constraints, photoinhibition.

Validation and Refinement of the GEM Through Essential Gene Comparison. The RB-TnSeq data served as not only a powerful guide during reconstruction but also a validation metric of the resulting model. The accuracy of the model was determined through a comparison of the in silico essential gene calls with the in vivo dataset. The gap between model predictions and in vivo realities highlights the

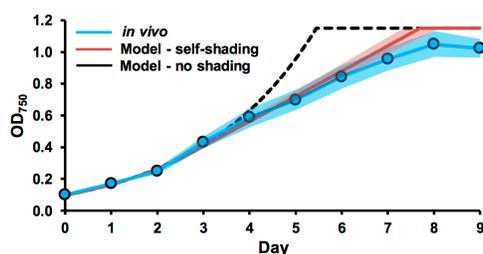


Fig. 2. Modeling the self-shading of cultures increases the accuracy of growth rate prediction. OD₇₅₀ values for the in vivo culture were the mean of three individual cultures. SE is represented by the shaded areas. SE for the in silico growth rate was determined from the SE of the in vivo inputs to the model.

limitations of the model as well as additional constraints on cellular metabolism. Evaluation of the disagreements enabled development of additional constraints beyond simple network connectivity (nonnetwork constraints), increasing the accuracy of the model predictions.

Essential gene-based model validation. We compared the in silico essential gene calls between our reconstructed model, iJB785, and a previous model of *S. elongatus*, iSyf715 (48) (Table 1). Minimal standardized constraints were applied to both models, allowing a direct comparison of the metabolic flexibility of the two networks. Of 752 genes in iJB785 with in vivo data, 587 (78%) were correctly assigned as either essential or nonessential (Fig. 3A and Dataset S3). The 165 disagreements were separated into explanatory categories (Fig. 3B). iSyf715 contained 683 genes with in vivo data, of which 377 (55%) were correctly assigned. Neither model was able to accurately predict the reduced growth rate phenotype indicative of genes categorized as beneficial (growth defect when mutated). The 319 genes essential in vivo that were not included in iJB785 participate in cellular processes, such as protein synthesis and transcription, that are out of scope for this GEM (Fig. S4). Incorporating the RB-TnSeq data during manual curation prevented the addition of excess metabolic flexibility (i.e., metabolic capabilities that are implied from the genome annotation but not observed in vivo) as evident by the increased accuracy of iJB785 compared with iSyf715.

Although both models incorporate only 25–30% of the ORFs identified in the *S. elongatus* genome, ~50% of the experimentally shown essential genes and 44% of functionally annotated ORFs are represented in iJB785. Additionally, of 157 genes labeled beneficial in vivo, 46% are present in iJB785. This enrichment of genes that impact cellular fitness underscores the value of GEMs for contextualizing meaningful in vivo genetic perturbations.

Increased model accuracy through nonnetwork constraints. GEMs offer a tool for visualizing the metabolic network use for a given KO genotype. The resulting flux map identifies alternate routes available to the network to respond to genetic perturbations. However, there are instances where the network connectivity indicates a metabolic pathway, but an additional constraint prevents its use in vivo. Such nonnetwork constraints resulted in a disagreement in the essentiality call for the pyruvate dehydrogenase (PDH) complex. Single-gene deletion of PDH in silico indicated that phosphoketolase (Synpcc7942_2080, EC 4.1.2.9) enables bypass of lower glycolysis by generating acetyl phosphate from the Calvin cycle intermediate fructose-6-phosphate or xylose-5-phosphate. Acetyl phosphate is converted to acetyl-CoA by the combined action of reversible acetate kinase and acetyl-CoA synthase, enabling bypass of PDH. This bypass has been investigated in PCC 6803 (49), and flux balance analysis in that organism also bypassed lower glycolysis with this pathway (50). The essential nature of PDH in vivo indicated an additional constraint that prevents this bypass from carrying sufficient flux to satisfy the acetyl-CoA needs of the cell. However, proteomics (51) and transcriptomics (26) datasets for *S. elongatus* indicated phosphoketolase abundance on the same order of magnitude as PDH subunits.

Because enzyme abundance could not explain the essentiality of PDH, we investigated metabolite channeling as a factor. Channeling is the result of spatial aggregation of pathway enzymes that prevents the intermediates from being acted on by enzymes outside of the pathway. MFA in PCC 6803 suggested metabolite channeling of Calvin cycle intermediates (11). This phenomenon can be modeled by coupling the flux between two reactions, forcing a ratio, and analyzing the result on the metabolic network. The analysis in silico of metabolite channeling indicated that, if more than 1% of Calvin cycle intermediates were allowed to enter the phosphoketolase bypass, PDH would be nonessential. These results indicate that either substantial metabolite channeling occurs in the Calvin cycle of

Table 1. Comparison of essentiality results between *iJB785* and *iSyf715*

Gene category	ORFs	<i>S. elongatus</i> PCC 7942					
		<i>iJB785</i> (this study)			<i>iSyf715</i> (previous model)		
		Included in GEM	Essentiality prediction	Correct* (%)	Included in GEM	Essentiality prediction	Correct* (%)
Essential	718	399	457	350 (88)	360	134	118 (33)
Beneficial	157	72	5	0 (0)	56	1	0 (0)
Nonessential	1,748	281	323	237 (85)	266	579	258 (97)
No in vivo data	100	33	0	N/A	32	0	N/A
Total genes	2,723	785	785	587 (78 [†])	714	714	376 (55 [†])

N/A, not applicable due to a lack of in vivo data.

*Equal to the number of genes correctly predicted to be essential in silico.

[†]Total correct genes/(total genes included in GEM – model genes with no in vivo data) × 100%.

S. elongatus or the phosphoketolase pathway is functioning in a yet uncharacterized way.

An additional nonnetwork constraint suggested by discrepancies between the model and the in vivo data is the phototrophic reaction catalyzed by ferredoxin-NADP oxidoreductase (FNOR; Synpcc7942_0978, EC 1.18.1.2), an essential reaction in vivo. The model indicated that FNOR was bypassed by NADPH:NAD⁺ transhydrogenase (Synpcc7942_1610, Synpcc7942_1611, and Synpcc7942_1612, EC 1.6.1.2), resulting in a discrepancy between the model and in vivo data. The canonical function of transhydrogenase is to provide metabolic flexibility by interconverting the two primary redox carriers. However, previous work in PCC 6803 had called into question the presence of an active transhydrogenase in cyanobacteria (52, 53). Previous modeling of PCC 6803 also observed dramatic changes in flux predictions depending on the activity of the transhydrogenase reaction (50), with the authors retaining the canonical transhydrogenase function. However, when we repeated the in silico essential gene assessment setting the trans-

hydrogenase reaction bounds to zero; along with additional non-network constraints, such as routing flux through PDH, 13 genes that had previously been discrepancies fell into alignment with in vivo data. These additional constraints, suggested for modeling in constant light, are provided in *SI Results*.

Central carbon metabolism flux predictions. The intracellular flux distribution maps the metabolic reaction use in a given condition. This visualization provides insight into highly used pathways that can be drawn on for product synthesis. The photoautotrophic flux distribution for central carbon metabolism in *S. elongatus* predicted by our GEM is shown in Fig. 4. Our flux values were in close alignment with PCC 6803 ¹³C MFA data (11), the in vivo equivalent of these in silico data. In PCC 6803, the flux ratio between carbon uptake and fixation was 1:1.27, whereas our model prediction for *S. elongatus* was 1:1.14. The flux ratio between the Calvin cycle and lower glycolysis, indicative of the biosynthetic carbon requirements of the cell, was 7.9:1 in our model prediction, in good agreement with the ratio of 9.8:1 observed in PCC 6803. The differences may be attributed to variations in the biomass composition for the two species or the fact that the model predicts optimality, whereas the in vivo data reflect inefficiencies naturally present in a living organism. Overall, the high accuracy of the essential gene assessments and consistency with published photoautotrophic flux data underscore the quality of the GEM.

Unusual Attributes of *S. elongatus* Metabolism. Taking into account the essential gene dataset (16) during the modeling process greatly improved the consistency of the model and the in vivo data. Still, reactions for which no evidence existed to bring the in silico prediction into agreement with the in vivo data deserved additional attention. The network reconstruction represents a repository of current knowledge; thus, discrepancies between in silico and in vivo results highlighted potential gaps in understanding of *S. elongatus* metabolism. Disagreements between *iJB785* and the RB-TnSeq essential gene calls were separated into categories reflecting the hypothesized source of the discrepancies (Fig. 3B and Dataset S3). Some categories, such as the Network Bypass and Annotated Isozymes, have already been discussed. Only 7% of discrepancies were categorized as Out of Scope, and these reactions were included in *iJB785* for completeness but operate in cellular processes not necessary for in silico growth. For example, the tRNA modification queuosine (54) is beneficial according to the RB-TnSeq data but was not explicitly modeled in silico, resulting in a discrepancy for all genes in the pathway.

The Nonessential Biomass discrepancies include enzymes that synthesize biomass components of the WT cellular composition but do not result in a significant in vivo growth defect when missing. Because every component of the defined biomass is required for in silico growth, this category represents in vivo flexibility not present in *iJB785*. Two such examples of known nonessential biomass components are sulfoquinovosyl diacylglycerol,

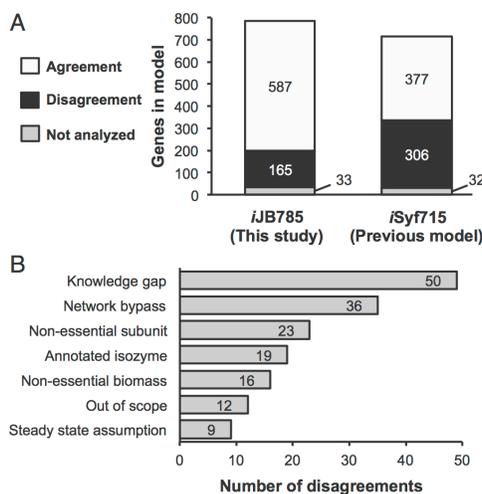


Fig. 3. Comparison of in vivo vs. in silico gene essentiality. (A) Comparison of in silico gene essentiality results for the model *iJB785* and the previous model of *S. elongatus* *iSyf715*. The agreements/disagreements are based on the comparison with in vivo RB-TnSeq results. (B) Sources of disagreements between in silico *iJB785* and in vivo gene essentiality.

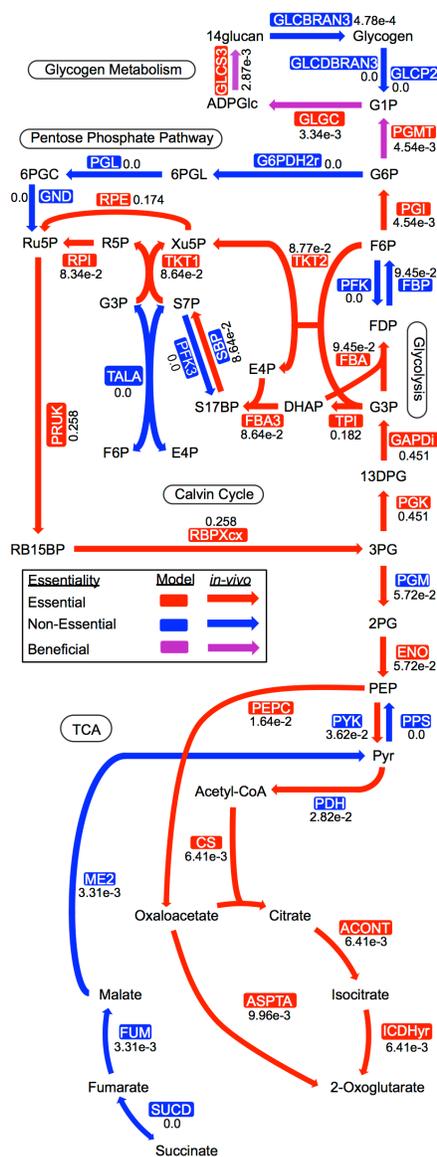


Fig. 4. Central carbon metabolism. In silico flux values are given numerically next to each reaction. The model essentiality calls are given by the color of the enzyme name, whereas arrow color indicates *in vivo* gene essentiality. Beneficial indicates a growth defect phenotype when the gene is mutated. Reaction and metabolite abbreviations are given in BIGG format (bigg.ucsd.edu) and found in [Dataset S2](#).

a component of the photosynthetic membranes (55), and the *O*-antigen polysaccharide, which even confers a fitness advantage against predators when mutated (56).

We also identified Nonessential Subunits of multiprotein complexes. If one gene in the complex is essential in silico, every subunit associated with that reaction is considered essential, even if the loss is tolerated *in vivo*. These disagreements included known nonessential subunits of the PCC 6803 photosynthetic electron transport chain (57) and ferredoxin: plastoquinone oxidoreductase complex (36).

Nucleotide salvage metabolism. Other discrepancies were placed into the categories Steady-State Assumption and Knowledge Gaps. Examples of both can be found in nucleotide salvage metabolism. Although *S. elongatus* encodes a complete set of enzymes for de novo biosynthesis of both purine and pyrimidine nucleotides, salvage reactions are mostly absent (Fig. S5). One exception is adenine phosphoribosyltransferase (Synpcc7942_2454, EC 2.4.2.7), which recycles adenine into AMP and is nonessential *in vivo* but essential in silico. Without this reaction, adenine produced during biosynthesis of polyamines would accumulate and violate the Steady-State Assumption in the model. However, the *in vivo* data suggest that adenine accumulation is not lethal to the organism, likely because of the small predicted flux through this pathway. Conversely, the catabolism of uracil into UMP by the enzyme uracil phosphoribosyltransferase (Synpcc7942_1715, EC 2.4.2.9) is essential *in vivo* but not in silico. The *in vivo* source of uracil and the metabolic requirement to salvage it to UMP represent a Knowledge Gap in our understanding of *S. elongatus*. Therefore, overlaying *in vivo* essentiality information over the model's predictions reveals multiple classes of unknowns in nucleotide salvage alone.

Photorespiration. The key carbon-fixing enzyme of the Calvin cycle in photosynthetic metabolism is ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO; Synpcc7942_1426 and Synpcc7942_1427, EC 4.1.1.39). This enzyme can fix not only CO₂ but O₂ as well, generating 2-phosphoglycolate. Buildup of this molecule is toxic and needs to be recycled through the process of photorespiration (58). To represent photorespiration in silico, a basal level of oxygenase activity needed to be added to the model's RuBisCO reaction. Based on extrapolation from ¹³C flux analysis in PCC 6803 (11) and metabolite concentrations in low- vs. high-carbon experiments in *S. elongatus* (59), we set the model RuBisCO oxygenase flux at 1% of total RuBisCO activity. Interestingly, the first step of the photorespiratory pathway is nonessential *in vivo*, although the model predicts it to be essential. In this step, phosphoglycolate phosphatase catalyzes the conversion of 2-phosphoglycolate to glycolate (Fig. S6). This enzyme, present upstream of the branching of photorespiration, has no high-confidence isozymes in *S. elongatus*. The nonessentiality of phosphoglycolate phosphatase suggests either an unknown enzyme for this function or dispensability of the pathway.

It is improbable that the photorespiratory pathway is non-essential. In PCC 6803, three pathways of photorespiration exist and have been included in previous GEMs of this species (50): the plant-like C2 cycle, full decarboxylation, and the glycerate pathway—which when disrupted in concert, cause a high-CO₂ dependency (60). *S. elongatus* may be missing the last of these pathways (Fig. S6). The glycerate pathway begins with glyoxylate carboxylase (GCL), an enzyme that combines two molecules of glyoxylate and ends with the central carbon metabolite 3-phosphoglycerate after the investment of ATP, NAD(P)H, and the release of CO₂. The gene found in PCC 6803 (*sll1981*) that is most similar to GCL in *E. coli* does not have a homolog in *S. elongatus*. Therefore, it is possible that only the plant-like C2 cycle and full decarboxylation via formate occur in *S. elongatus*.

The potential to generate glycine through photorespiration raised the possibility that the process could compensate for de novo glycine/serine biosynthesis. When RuBisCO oxygenase activity was set at 1%, the model predicted that sufficient glycine

would be created to support growth, similar to predictions in previous cyanobacterial GEMs (50, 61). Therefore, even when the de novo synthesis pathway of glycine from 3-phosphoglycerate via serine was broken in the model, it still predicted that cells grow at 72% of their normal rate. This finding runs counter to the in vivo data, which show that the entire biosynthesis pathway of serine from 3-phosphoglycerate is essential. The experimental data suggest low flux to amino acid biosynthesis through photorespiration, which could be explained by lower than expected photorespiration activity. Along these lines, in silico essential gene results became consistent with the in vivo data only when RuBisCO's oxygenase activity was lowered to 0.15% of its carbon-fixing activity. Another possibility is decreased flux specifically toward the amino acid biosynthesis pathway of photorespiration. Glycine hydroxymethyltransferase (Synpcc7942_0282, EC 2.1.2.1) in PCC 6803 is a choke point for the conversion of glycine from photorespiration into serine (62); it is possible that the same limitation exists in *S. elongatus*. There is also evidence for essentiality of the de novo serine biosynthetic pathway beyond simple metabolic requirements in PCC 6803 (25). Therefore, because of lower than expected photorespiration, limited flux toward glycine synthesis, or undiscovered requirements of de novo serine synthesis, photorespiration is not able to replace de novo amino acid biosynthesis from 3-phosphoglycerate.

A truncated TCA cycle. The completeness of the cyanobacterial TCA cycle has been an oft-debated subject (63). Since the discovery that 2-oxoglutarate dehydrogenase is missing in cyanobacteria, it was accepted for many years that the TCA cycle is incomplete (64). However, a complete TCA cycle is responsible for the majority of energy intermediates created by oxygenic metabolism and nearly ubiquitous throughout nature (65). For this reason, extensive effort has been applied to uncover routes that complete cyanobacterial TCA cycles. A number of bypasses of the missing 2-oxoglutarate dehydrogenase have been discovered, such as the 2-oxoglutarate decarboxylase pathway (66), the GABA shunt (67), and the glyoxylate shunt (68). These bypasses revealed that cyanobacteria harbor complete, albeit noncanonical TCA cycles (63). More recently, however, the necessity of the newly circularized TCA cycles of cyanobacteria has been called into question by experimental (16) and modeling studies (50). Therefore, the structure and biological relevance of a TCA cycle remains an open question in cyanobacteria.

For these reasons, we were particularly interested in discrepancies between the draft model simulations and the in vivo essentiality for two enzymes of the TCA cycle: fumarase (Synpcc7942_1007, EC 4.2.1.2) and malic enzyme (Synpcc7942_1297, EC 1.1.1.40). These enzymes are required by the model's steady-state assumption for the recycling of fumarate, a by-product of both purine and arginine synthesis. However, the overlay of the in vivo data on the model not only suggests that this recycling is not an essential function but more broadly, led us to evaluate the importance of a complete TCA cycle in *S. elongatus*.

We began by examining the bypasses that complete TCA cycles in other model cyanobacteria (63) for their potential presence in *S. elongatus*. Our reconstruction, however, revealed none of the known cyanobacterial bypasses (Fig. 5A). Additionally, we were unable to find evidence of the core TCA-cycle enzymes malate dehydrogenase (EC 1.1.1.37), malate:quinone oxidoreductase (EC 1.1.5.4), and succinyl-CoA synthetase (EC 6.2.1.5). Furthermore, the succinate dehydrogenase genes (synpcc7942_0314, synpcc7942_0641, and synpcc7942_1533, EC 1.3.5.1) are nonessential in vivo in addition to genes for fumarase and malic enzyme. Together, these data provide evidence that the metabolically important portion of the TCA cycle in *S. elongatus* is highly abridged. We call this oxidative, noncyclic portion of the TCA cycle that is essential in *S. elongatus* the TCA pathway (Fig. 5B). To explain the feasibility of this TCA pathway, we examined whether it would be sufficient to accomplish

the central functions of the TCA cycle: precursor metabolite production, by-product recycling, and energy generation.

The TCA pathway preserves the enzymes necessary for the synthesis of oxaloacetate and 2-oxoglutarate, which are precursors for many required biomass components; additionally, 2-oxoglutarate serves as the gateway to nitrogen assimilation. Therefore, functionality in producing precursor metabolites and nitrogen assimilation can be provided by the TCA pathway.

The TCA pathway does not include functionality for the recycling of fumarate. Fumarate is created as a by-product of nucleotide and arginine synthesis, the salvage of which is posited by previous models to be essential (69). However, loss of function mutants for fumarase and malic enzyme show that this recycling function is not required for viability on solid media or in liquid culture (Fig. 5C and Fig. S7A and B). The dispensability of fumarate salvage could be explained by the ability of *S. elongatus* to excrete fumarate into the media (70), and when this possibility was added into the model, the in silico predictions for fumarase and malic enzyme become nonessential in agreement with the in vivo data. In addition, *iJB785* shows a minor cost of excreting useable carbon backbones (Fig. S7C). This cost is commensurate with the in vivo data that show a significant decrease in colony size in the fumarase and malic enzyme mutants (Fig. 5C). Therefore, fumarate recycling is a dispensable function of the TCA cycle, despite a slight fitness cost.

The final core function of the TCA cycle is energy production. During photosynthetic metabolism in *S. elongatus*, however, full oxidation of pyruvate by the TCA cycle would amount to "metabolic suicide," in which the cell is fixing and degrading the same carbon compounds concurrently (71). The wastefulness of a complete TCA cycle for energy generation in *S. elongatus* is supported by viability of mutants defective for succinate dehydrogenase subunit B (Synpcc7942_1533, EC 1.3.5.1) (Fig. 5C), which in addition to its importance for cycle flux, is an electron donor to the electron transport chain in PCC 6803 (53). At nighttime, however, when the cell switches from photosynthesis to glycogen as its energy source (72), we might expect that a cyclic TCA process would become essential, because it would enable further energy generation from glycogen. In fact, expression of TCA cycle enzymes has been shown to occur in the dark period during light-dark cultivation (73), and modeling has shown cyclic flux through the cycle in PCC6803 under these conditions (50). Therefore, we repeated the viability assay under day-night conditions but found that fumarase, malic enzyme, and succinate dehydrogenase remain nonessential (Fig. 5D). Furthermore, *iJB785* simulations of dark metabolism indicated that full oxidation of glycogen through the oxidative pentose phosphate (OPP) pathway could generate equivalent ATP compared with a complete TCA cycle in *S. elongatus* (24.4 mol ATP/mol glucose in OPP vs. 24.7 mol ATP/mol glucose via TCA bypass). This prediction is supported by previous experimental evidence that the OPP pathway is important for diurnal survival in *S. elongatus* (74–76). Taken together, the model, the essential gene dataset, and our loss of function mutants support the hypothesis that an abridged TCA pathway, focused on generation of precursor metabolites, and not the traditional TCA cycle is the physiologically relevant TCA process for *S. elongatus*. This finding diverges from the current paradigm of complete TCA cycles in cyanobacteria (63).

TCA cycles focused on biosynthesis instead of energy generation have precedents. Green sulfur bacteria run their TCA cycle in reverse to fix CO₂ in a process called the reductive TCA cycle (77). It also is common for obligate autotrophs to lack the enzyme 2-oxoglutarate dehydrogenase (71). The hypothesized TCA mechanism in the absence of 2-oxoglutarate is a bifurcated process, in which a reductive branch leads to succinyl-CoA and an oxidative branch leads to 2-oxoglutarate (71). According to *iJB785*, however, *S. elongatus* does not require succinyl-CoA or any other metabolites of the reductive branch of this TCA process. Thus, it is likely that just the

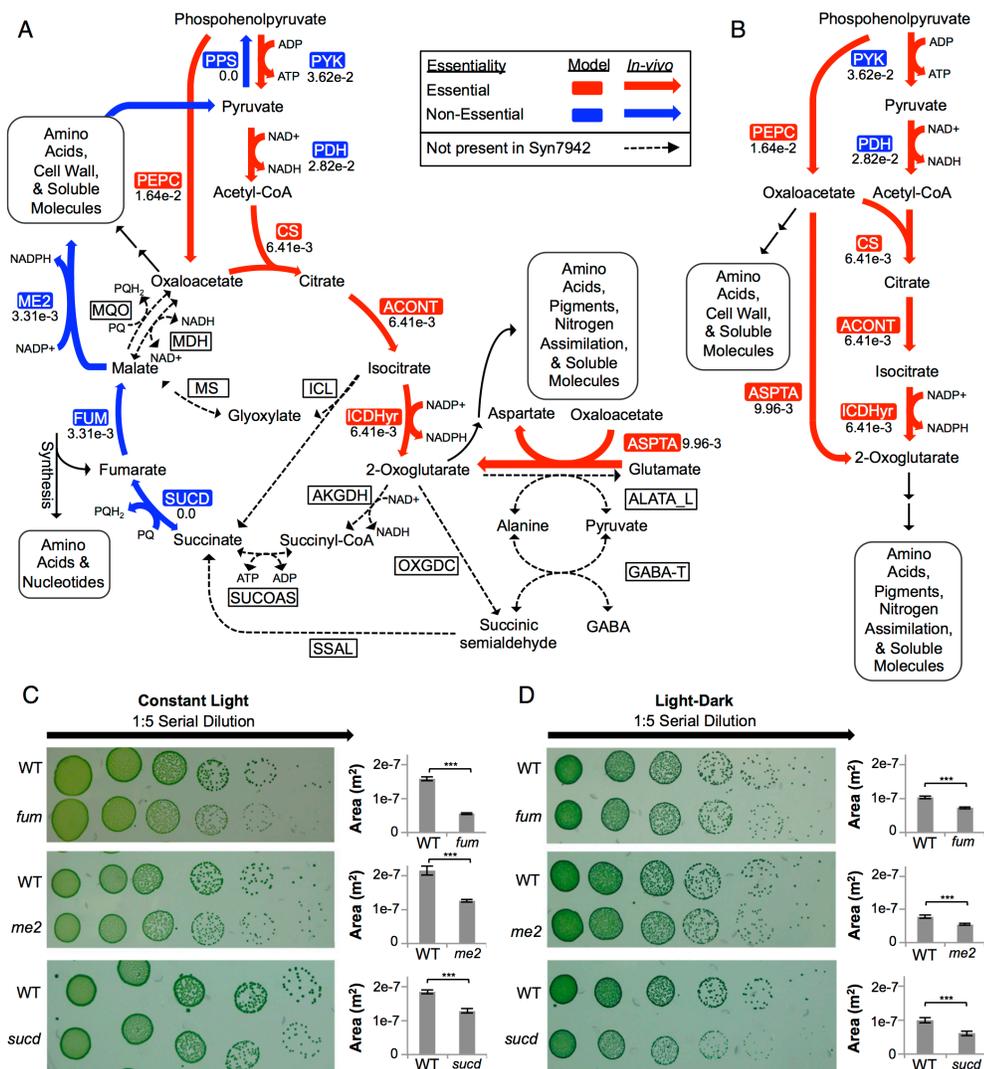


Fig. 5. The TCA pathway. (A) The canonical TCA cycle is shown, with bypasses and alternative reactions present in other cyanobacterial species indicated by dotted lines. (B) The TCA pathway, the proposed biologically relevant TCA process. Insertion loss of function mutants in *fum* (8542-06; *synpcc7942_1007*), *me2* (8529-16; *synpcc7942_1297*), and *sucd* (851-114; *synpcc7942_1533*) were made (SI Materials and Methods, Table S2), and growth was compared with the WT in both (C) continuous light and (D) cycles of alternating light-dark (12–12 h). Colony area was measured using ImageJ (80). Reaction and metabolite abbreviations are given in BiGG format (bigg.ucsd.edu) and found in Dataset S2. ***Significance level of 0.001 (t test).

oxidative wing of the TCA cycle, represented as the TCA pathway, is important largely as a biosynthetic pathway in *S. elongatus*.

The truncated TCA pathway of *S. elongatus* probably generalizes to other members of the phylum. Even if we artificially model the complete TCA cycle in *S. elongatus* by adding malate dehydrogenase and the 2-oxoglutarate dehydrogenase bypass present

in *Synechococcus* sp. PCC 7002 (66), the model still predicts no cyclic flux through the completed TCA cycle. In PCC 6803, which contains a complete TCA cycle, both flux balance analysis (50) and ¹³C MFA (78) show negligible flux from 2-oxoglutarate to the rest of the TCA cycle. Furthermore, when the complete TCA cycle of PCC 6803 is blocked, only minor decreases in growth rate are

observed (67). Finally, a bifurcated TCA cycle with a reductive branch for succinyl-CoA synthesis is likely to be nonessential in many cyanobacteria because of the presence of a heme biosynthesis pathway that begins with 2-oxoglutarate instead of succinyl-CoA (79). Based on these data, the abridged TCA pathway as opposed to a complete or bifurcated TCA process is likely relevant in other cyanobacteria, even those with genes making a complete TCA cycle.

Conclusions

The *iJB785* GEM of metabolism in *S. elongatus* presented here is a comprehensive representation of obligate phototrophic metabolism. Our mechanistic modeling of photon absorption and self-shading addresses the persistent challenge of accurately modeling light as a metabolite. This approach can be applied to any phototrophic GEM, enabling modeling of core aspects of light-driven metabolism. The predictive nature of the method also enables tailored light regimes for bioprocess optimization of photosynthetic platforms.

In addition to *iJB785*'s value for metabolic engineering and its technical improvement to phototrophic modeling, it serves as a platform for biological discovery. In synthesizing much of the physiological understanding available for *S. elongatus*, *iJB785* reveals the holes in this knowledge. Some of these holes include missing elements of the nucleotide salvage system, the reason that phosphoketolase is unable to bypass lower glycolysis, and the apparent noncanonical activity of the transhydrogenase. Furthermore, the incorporation of essential gene data both improved the model's accuracy and highlighted disagreements, which could not

be explained by published data for *S. elongatus*. Many of these inconsistencies represent new biology for *S. elongatus*, such as the importance of a linear, noncyclic TCA pathway. Finally, as a representation of our current best understanding of *S. elongatus*, *iJB785* is an ideal surface on which to overlay whole-genome datasets. The future addition of omic datasets to the model will both greatly improve in silico representation of *S. elongatus* metabolism and identify additional biological unknowns.

Materials and Methods

Methods used to generate the genome-scale reconstruction, derive constraints, and generate in silico results are presented in *SI Materials and Methods*. Briefly, the metabolic reconstruction was assembled using an established protocol (20), and biomass-normalized photon absorption rate for a given wavelength range was calculated from the combination of irradiance, optical absorption cross-section, and the chlorophyll a component of the biomass equation. Growth rates and reaction fluxes were simulated by maximizing the biomass objective function. All additional experimental protocols can be found in *SI Materials and Methods*.

ACKNOWLEDGMENTS. A draft reconstruction of *S. elongatus* metabolism was provided by Henning Knoop and Ralf Steuer. This research was supported by NIH Cell and Molecular Genetics Training Grant T32GM00724 (to B.E.R.), University of California, San Diego Frontiers of Innovation Scholars Program 2-P3071 (to D.G.W.), National Science Foundation Grant MCB1244108 (to S.S.G.), and US Department of Energy, Office of Science, Office of Biological and Environmental Research Award DE-SC0008593 (to B.O.P.).

- Angermayr SA, Gorchs Rovira A, Hellingwerf KJ (2015) Metabolic engineering of cyanobacteria for the synthesis of commodity products. *Trends Biotechnol* 33(6):352–361.
- Atsumi S, Higashide W, Liao JC (2009) Direct photosynthetic recycling of carbon dioxide to isobutyraldehyde. *Nat Biotechnol* 27(12):1177–1180.
- Oliver JW, Machado IM, Yoneda H, Atsumi S (2013) Cyanobacterial conversion of carbon dioxide to 2,3-butanediol. *Proc Natl Acad Sci USA* 110(4):1249–1254.
- Clerico EM, Ditty JL, Golden SS (2007) Specialized techniques for site-directed mutagenesis in cyanobacteria. *Methods Mol Biol* 362:155–171.
- Gudmundsson S, Nogales J (2015) Cyanobacteria as photosynthetic biocatalysts: A systems biology perspective. *Mol Biosyst* 11(1):60–70.
- Savakis P, Hellingwerf KJ (2015) Engineering cyanobacteria for direct biofuel production from CO₂. *Curr Opin Biotechnol* 33:8–14.
- Bordbar A, Monk JM, King ZA, Palsson BO (2014) Constraint-based models predict metabolic and associated cellular functions. *Nat Rev Genet* 15(2):107–120.
- Borodina I, et al. (2015) Establishing a synthetic pathway for high-level production of 3-hydroxypropionic acid in *Saccharomyces cerevisiae* via β -alanine. *Metab Eng* 27:57–64.
- Yim H, et al. (2011) Metabolic engineering of *Escherichia coli* for direct production of 1,4-butanediol. *Nat Chem Biol* 7(7):445–452.
- Shih PM, et al. (2013) Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing. *Proc Natl Acad Sci USA* 110(3):1053–1058.
- Young JD, Shastri AA, Stephanopoulos G, Morgan JA (2011) Mapping photoautotrophic metabolism with isotopically nonstationary (13)C flux analysis. *Metab Eng* 13(6):656–665.
- Kim TY, Sohn SB, Kim YB, Kim WJ, Lee SY (2012) Recent advances in reconstruction and applications of genome-scale metabolic models. *Curr Opin Biotechnol* 23(4):617–623.
- Lewis NE, Nagarajan H, Palsson BO (2012) Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods. *Nat Rev Microbiol* 10(4):291–305.
- Monk J, Nogales J, Palsson BO (2014) Optimizing genome-scale network reconstructions. *Nat Biotechnol* 32(5):447–452.
- Chowdhury R, Chowdhury A, Maranas CD (2015) Using gene essentiality and synthetic lethality information to correct yeast and CHO cell genome-scale models. *Metabolites* 5(4):536–570.
- Rubin BE, et al. (2015) The essential gene set of a photosynthetic organism. *Proc Natl Acad Sci USA* 112(48):E6634–E6643.
- Reed JL, et al. (2006) Systems approach to refining genome annotation. *Proc Natl Acad Sci USA* 103(46):17480–17484.
- Barouk C, Muñoz-Tamayo R, Steyer JP, Bernard O (2015) A state of the art of metabolic networks of unicellular microalgae and cyanobacteria for biofuel production. *Metab Eng* 30:49–60.
- Shastri AA, Morgan JA (2005) Flux balance analysis of photoautotrophic metabolism. *Biotechnol Prog* 21(6):1617–1626.
- Thiele I, Palsson BO (2010) A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat Protoc* 5(1):93–121.
- Reed JL, Vo TD, Schilling CH, Palsson BO (2003) An expanded genome-scale model of *Escherichia coli* K-12 (JUR904 GSM/GPR). *Genome Biol* 4(9):R54.
- Brunk E, et al. (2016) Systems biology of the structural proteome. *BMC Syst Biol* 10:26.
- Jablonsky J, Hagemann M, Schwarz D, Wolkenhauer O (2013) Phosphoglycerate mutases function as reverse regulated isoenzymes in *Synechococcus elongatus* PCC 7942. *PLoS One* 8(3):e58281.
- Chiba Y, et al. (2013) Structural units important for activity of a novel-type phosphoserine phosphatase from *Hydrogenobacter thermophilus* TK-6 revealed by crystal structure analysis. *J Biol Chem* 288(16):11448–11458.
- Klemke F, et al. (2015) Identification of the light-independent phosphoserine pathway as an additional source of serine in the cyanobacterium *Synechocystis* sp. PCC 6803. *Microbiology* 161(Pt 5):1050–1060.
- Vijayan V, Jain IH, O'Shea EK (2011) A high resolution map of a cyanobacterial transcriptome. *Genome Biol* 12(5):R47.
- Berla BM, Saha R, Maranas CD, Pakrasi HB (2015) Cyanobacterial alkanes modulate photosynthetic cyclic electron flow to assist growth under cold stress. *Sci Rep* 5:14894.
- Hendry JL, Prasanna CB, Joshi A, Dasgupta S, Wangikar PP (2016) Metabolic model of *Synechococcus* sp. PCC 7002: Prediction of flux distribution and network modification for enhanced biofuel production. *Bioresour Technol* 213:190–197.
- Krieger-Liszak A (2005) Singlet oxygen production in photosynthesis. *J Exp Bot* 56(411):337–346.
- Patterson CO, Myers J (1973) Photosynthetic production of hydrogen peroxide by *Anacystis nidulans*. *Plant Physiol* 51(1):104–109.
- Chang RL, et al. (2011) Metabolic network reconstruction of *Chlamydomonas* offers insight into light-driven algal metabolism. *Mol Syst Biol* 7:518.
- He L, Wu SG, Wan N, Reding AC, Tang YJ (2015) Simulating cyanobacterial phenotypes by integrating flux balance analysis, kinetics, and a light distribution function. *Microb Cell Fact* 14:206.
- Dubinsky Z, Berman T, Schanz F (1984) Field experiments for in situ measurement of photosynthetic efficiency and quantum yield. *J Plankton Res* 6(2):339–349.
- Liberton M, et al. (2016) Global proteomic analysis reveals an exclusive role of thylakoid membranes in bioenergetics of a model cyanobacterium. *Mol Cell Proteomics* 15(6):2021–2032.
- Liu LN, et al. (2012) Control of electron transport routes through redox-regulated redistribution of respiratory complexes. *Proc Natl Acad Sci USA* 109(28):11431–11436.
- Gao F, et al. (2016) NdhV is a subunit of NADPH dehydrogenase essential for cyclic electron transport in *Synechocystis* sp. strain PCC 6803. *Plant Physiol* 170(2):752–760.
- Battchikova N, et al. (2011) Identification of novel Ssl0352 protein (NdhS), essential for efficient operation of cyclic electron transport around photosystem I, in NADPH: plastoquinone oxidoreductase (NDH-1) complexes of *Synechocystis* sp. PCC 6803. *J Biol Chem* 286(42):36992–37001.
- Allakhverdiev SI, Murata N (2004) Environmental stress inhibits the synthesis de novo of proteins involved in the photodamage-repair cycle of Photosystem II in *Synechocystis* sp. PCC 6803. *Biochim Biophys Acta* 1657(1):23–32.
- Stamatakis K, Tsimilli-Michael M, Papageorgiou GC (2014) On the question of the light-harvesting role of β -carotene in photosystem II and photosystem I core complexes. *Plant Physiol Biochem* 81:121–127.
- King ZA, et al. (2016) BIGG Models: A platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Res* 44(D1):D515–D522.
- Feist AM, Palsson BO (2010) The biomass objective function. *Curr Opin Microbiol* 13(3):344–349.

42. Varma A, Palsson BO (1994) Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110. *Appl Environ Microbiol* 60(10):3724–3731.
43. Yu J, et al. (2015) *Synechococcus elongatus* UTEX 2973, a fast growing cyanobacterial chassis for biosynthesis using light and CO₂. *Sci Rep* 5:8132.
44. Shibata M, et al. (2001) Distinct constitutive and low-CO₂-induced CO₂ uptake systems in cyanobacteria: Genes involved and their phylogenetic relationship with homologous genes in other organisms. *Proc Natl Acad Sci USA* 98(20):11789–11794.
45. Woodger FJ, Badger MR, Price GD (2005) Sensing of inorganic carbon limitation in *Synechococcus* PCC7942 is correlated with the size of the internal inorganic carbon pool and involves oxygen. *Plant Physiol* 139(4):1959–1969.
46. McConnell MD, Koop R, Vasil'ev S, Bruce D (2002) Regulation of the distribution of chlorophyll and phycobilin-absorbed excitation energy in cyanobacteria. A structure-based model for the light state transition. *Plant Physiol* 130(3):1201–1212.
47. Muramatsu M, Hihara Y (2012) Acclimation to high-light conditions in cyanobacteria: From gene expression to physiological responses. *J Plant Res* 125(1):11–39.
48. Triana J, et al. (2014) Generation and evaluation of a genome-scale metabolic network model of *Synechococcus elongatus* PCC7942. *Metabolites* 4(3):680–698.
49. Xiong W, et al. (2015) Phosphoketolase pathway contributes to carbon metabolism in cyanobacteria. *Nat Plants* 2:15187.
50. Knoop H, et al. (2013) Flux balance analysis of cyanobacterial metabolism: The metabolic network of *Synechocystis* sp. PCC 6803. *PLoS Comput Biol* 9(6):e1003081.
51. Guerreiro AC, et al. (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(8):2042–2055.
52. Angermayr SA, Paszota M, Hellingwerf KJ (2012) Engineering a cyanobacterial cell factory for production of lactic acid. *Appl Environ Microbiol* 78(19):7098–7106.
53. Cooley JW, Vermaas WF (2001) Succinate dehydrogenase and other respiratory pathways in thylakoid membranes of *Synechocystis* sp. strain PCC 6803: Capacity comparisons and physiological function. *J Bacteriol* 183(14):4251–4258.
54. Reader JS, Metzgar D, Schimmel P, de Crécy-Lagard V (2004) Identification of four genes necessary for biosynthesis of the modified nucleoside queuosine. *J Biol Chem* 279(8):6280–6285.
55. Aoki M, Sato N, Meguro A, Tsumuki M (2004) Differing involvement of sulfoquinovosyl diacylglycerol in photosystem II in two species of unicellular cyanobacteria. *Eur J Biochem* 271(4):685–693.
56. Simkovsky R, et al. (2012) Impairment of O-antigen production confers resistance to grazing in a model amoeba-cyanobacterium predator-prey system. *Proc Natl Acad Sci USA* 109(41):16678–16683.
57. Schneider D, Volkmer T, Rögnér M (2007) PetG and PetN, but not PetL, are essential subunits of the cytochrome b6f complex from *Synechocystis* PCC 6803. *Res Microbiol* 158(1):45–50.
58. Husic DW, Tolbert NE (1987) Inhibition of glycolate and D-lactate metabolism in a *Chlamydomonas reinhardtii* mutant deficient in mitochondrial respiration. *Proc Natl Acad Sci USA* 84(6):1555–1559.
59. Schwarz D, et al. (2011) Metabolic and transcriptomic phenotyping of inorganic carbon acclimation in the Cyanobacterium *Synechococcus elongatus* PCC 7942. *Plant Physiol* 155(4):1640–1655.
60. Eisenhut M, et al. (2008) Metabolome phenotyping of inorganic carbon limitation in cells of the wild type and photorespiratory mutants of the cyanobacterium *Synechocystis* sp. strain PCC 6803. *Plant Physiol* 148(4):2109–2120.
61. Nogales J, Gudmundsson S, Knight EM, Palsson BO, Thiele I (2012) Detailing the optimalities of photosynthesis in cyanobacteria through systems biology analysis. *Proc Natl Acad Sci USA* 109(7):2678–2683.
62. Eisenhut M, et al. (2006) The plant-like C₂ glycolate cycle and the bacterial-like glycerate pathway cooperate in phosphoglycolate metabolism in cyanobacteria. *Plant Physiol* 142(1):333–342.
63. Steinhäuser D, Fernie AR, Araújo WL (2012) Unusual cyanobacterial TCA cycles: Not broken just different. *Trends Plant Sci* 17(9):503–509.
64. Pearce J, Leach CK, Carr NG (1969) The incomplete tricarboxylic acid cycle in the blue-green alga *Anabaena variabilis*. *J Gen Microbiol* 55(3):371–378.
65. Fernie AR, Carrari F, Sweetlove LJ (2004) Respiratory metabolism: Glycolysis, the TCA cycle and mitochondrial electron transport. *Curr Opin Plant Biol* 7(3):254–261.
66. Zhang S, Bryant DA (2011) The tricarboxylic acid cycle in cyanobacteria. *Science* 334(6062):1551–1553.
67. Xiong W, Brune D, Vermaas WF (2014) The γ -aminobutyric acid shunt contributes to closing the tricarboxylic acid cycle in *Synechocystis* sp. PCC 6803. *Mol Microbiol* 93(4):786–796.
68. Zhang S, Bryant DA (2015) Biochemical validation of the glyoxylate cycle in the cyanobacterium *Chlorogobiosis fritschii* strain PCC 9212. *J Biol Chem* 290(22):14019–14030.
69. Beck C, Knoop H, Axmann IM, Steuer R (2012) The diversity of cyanobacterial metabolism: Genome analysis of multiple phototrophic microorganisms. *BMC Genomics* 13:56.
70. Hickman JW, et al. (2013) Glycogen synthesis is a required component of the nitrogen stress response in *Synechococcus elongatus* PCC 7942. *Algal Res* 2(2):98–106.
71. Wood AP, Aurikko JP, Kelly DP (2004) A challenge for 21st century molecular biology and biochemistry: What are the causes of obligate autotrophy and methanotrophy? *FEMS Microbiol Rev* 28(3):335–352.
72. Diamond S, Jun D, Rubin BE, Golden SS (2015) The circadian oscillator in *Synechococcus elongatus* controls metabolite partitioning during diurnal growth. *Proc Natl Acad Sci USA* 112(15):E1916–E1925.
73. Welkie D, et al. (2014) Transcriptomic and proteomic dynamics in the metabolism of a diazotrophic cyanobacterium, *Cyanothece* sp. PCC 7942 during a diurnal light-dark cycle. *BMC Genomics* 15:1185.
74. Doolittle WF, Singer RA (1974) Mutational analysis of dark endogenous metabolism in the blue-green bacterium *Anacystis nidulans*. *J Bacteriol* 119(3):677–683.
75. Broedel SE, Jr, Wolf RE, Jr (1990) Genetic tagging, cloning, and DNA sequence of the *Synechococcus* sp. strain PCC 7942 gene (gnd) encoding 6-phosphogluconate dehydrogenase. *J Bacteriol* 172(7):4023–4031.
76. Scanlan DJ, Sundaram S, Newman J, Mann NH, Carr NG (1995) Characterization of a *zwf* mutant of *Synechococcus* sp. strain PCC 7942. *J Bacteriol* 177(9):2550–2553.
77. Atomi H (2002) Microbial enzymes involved in carbon dioxide fixation. *J Biosci Bioeng* 94(6):497–505.
78. You L, Berla B, He L, Pakrasi HB, Tang YJ (2014) 13C-MFA delineates the photo-mixotrophic metabolism of *Synechocystis* sp. PCC 6803 under light- and carbon-sufficient conditions. *Biotechnol J* 9(5):684–692.
79. Obornik M, Green BR (2005) Mosaic origin of the heme biosynthesis pathway in photosynthetic eukaryotes. *Mol Biol Evol* 22(12):2343–2353.
80. Schindelin J, Rueden CT, Hiner MC, Eliceiri KW (2015) The ImageJ ecosystem: An open platform for biomedical image analysis. *Mol Reprod Dev* 82(7–8):518–529.
81. Camacho C, et al. (2009) BLAST+: Architecture and applications. *BMC Bioinformatics* 10:421.
82. UniProt Consortium (2015) UniProt: A hub for protein information. *Nucleic Acids Res* 43(Database issue):D204–D212.
83. Marchler-Bauer A, et al. (2015) CDD: NCBI's conserved domain database. *Nucleic Acids Res* 43(Database issue):D222–D226.
84. Ebrahim A, Lerman JA, Palsson BO, Hyduke DR (2013) COBRAPy: Constraints-Based Reconstruction and Analysis for Python. *BMC Syst Biol* 7:74.
85. Perez F, Granger BE (2007) IPython: A system for interactive scientific computing. *Comput Sci Eng* 9(3):21–29.
86. Bond CS, White MF, Hunter WN (2001) High resolution structure of the phosphohistidine-activated form of *Escherichia coli* cofactor-dependent phosphoglycerate mutase. *J Biol Chem* 276(5):3247–3253.
87. Rigden DJ, Littlejohn JE, Henderson K, Jedrzejak MJ (2003) Structures of phosphate and triphosphate complexes of *Bacillus stearothermophilus* phosphatase PhoE: Structural and functional analysis in the cofactor-dependent phosphoglycerate mutase superfamily. *J Mol Biol* 325(3):411–420.
88. Roy A, Kucukural A, Zhang Y (2010) I-TASSER: A unified platform for automated protein structure and function prediction. *Nat Protoc* 5(4):725–738.
89. Yang J, Roy A, Zhang Y (2013) Protein-ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment. *Bioinformatics* 29(20):2588–2595.
90. Humphrey W, Dalke A, Schulten K (1996) VMD: Visual molecular dynamics. *J Mol Graph* 14(1):27–38.
91. Roberts E, Eargle J, Wright D, Luthey-Schulten Z (2006) MultiSeq: Unifying sequence and structure data for evolutionary analysis. *BMC Bioinformatics* 7:382.
92. Russell RB, Barton GJ (1992) Multiple protein sequence alignment from tertiary structure comparison: Assignment of global and residue confidence levels. *Proteins* 14(2):309–323.
93. Ritchie RJ (2006) Consistent sets of spectrophotometric chlorophyll equations for acetone, methanol and ethanol solvents. *Photosynth Res* 89(1):27–41.
94. Falkowski PG, Raven JA (2007) *Aquatic Photosynthesis* (Princeton Univ Press, Princeton), 2nd Ed.
95. Platt T, Gallegos CL, Harrison WG (1980) Photoinhibition of photosynthesis in natural assemblages of marine-phytoplankton. *J Mar Res* 38(4):687–701.
96. Schellenberger J, et al. (2011) Quantitative prediction of cellular metabolism with constraint-based models: The COBRA Toolbox v2.0. *Nat Protoc* 6(9):1290–1307.
97. Lewis NE, et al. (2010) Omic data from evolved *E. coli* are consistent with computed optimal growth from genome-scale models. *Mol Syst Biol* 6:390.
98. Orth JD, et al. (2011) A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism—2011. *Mol Syst Biol* 7:535.
99. 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*, eds Burnap R, Vermaas W (Springer, Dordrecht, The Netherlands), pp 119–137.
100. Holtman CK, et al. (2005) High-throughput functional analysis of the *Synechococcus elongatus* PCC 7942 genome. *DNA Res* 12(2):103–115.
101. Ashby MK, Houmar J (2006) Cyanobacterial two-component proteins: Structure, diversity, distribution, and evolution. *Microbiol Mol Biol Rev* 70(2):472–509.

Supporting Information

Broddrick et al. 10.1073/pnas.1613446113

SI Materials and Methods

Genome-Scale Reconstruction for *Synechococcus elongatus*. The genome annotation of *Synechococcus elongatus* PCC 7942 was obtained from the Cyanobase database (genome.microbedb.jp/cyanobase/SYNPCC7942; date accessed: 9/2015). Functional annotation of the predicted ORFs was performed using the BLAST command line tool (81). Initially, the amino acid sequences of the *S. elongatus* proteins were compared with all reviewed cyanobacterial proteins in the Uniprot/Swissprot database (82) (BLASTp; e-value cutoff $1e^{-20}$). The *S. elongatus* genes without a best hit in the cyanobacterial subset of the database were then searched against all reviewed Uniprot/Swissprot sequences that had evidence at the protein or transcript level (BLASTp; e-value cutoff $1e^{-20}$). Additionally, the *S. elongatus* protein sequences were queried for conserved domains using the National Center for Biotechnology Information Batch CD-Search Tool (83) with the default settings. A draft reconstruction for *S. elongatus* based on a previous model for sp. PCC 6803 (50), provided by Henning Knoop and Ralf Steuer, Humboldt-Universität zu Berlin Institute for Theoretical Biology, Berlin, served as the starting point for the reconstruction. The genome functional annotation was curated based on the BLASTp and CD-Search outputs, and the network was reconstructed as described previously (20) using the COBRAPy Python package (84) in iPython Notebook (85).

Structural Homology Modeling of Annotated PGMs. All protein structures were downloaded from the Protein Data Bank (PDB). PDB ID codes 4IJ5 [*Hydrogenobacter thermophilus* TK-6 metal-independent phosphoserine phosphatase 1 (iPSP1)] (24), 1E59 [*Escherichia coli* cofactor-dependent phosphoglycerate mutase (dPGM)] (86), and 1H2F [*Bacillus stearothermophilus* PhoE] (87) were selected for this analysis. Full-length homology modeling of amino acid sequences was conducted using a locally downloaded version of the I-TASSER v4.4 (iterative threading assembly refinement) package (88). The COACH package (89), which is contained within I-TASSER, was used for binding site and substrate binding predictions, which were cross-referenced with known dPGM and iPSP binding residues annotated within UniProt (82) and various literature sources. Additionally, the full-length homology model of the *E. coli* dPGM (UniProt ID code P62707) was obtained from a database of *E. coli* homology models previously generated with I-TASSER (zhanglab.ccmb.med.umich.edu/Ecoli/). To compare the positions of binding and active residues, structure files were simultaneously loaded into VMD (90), and sequences were aligned using the MultiSeq tool (91). MultiSeq also contains the structural alignment tool STAMP (92), which was used to align the regions of interest. The positions of known histidine phosphatase residues and residues that are known to contribute to either iPSP or dPGM activity were then compared between all homology models and the selected experimental PDB files.

Strains and Culture Conditions. All WT assays were done in *S. elongatus* PCC 7942, which is stored in our laboratory's culture collection as AMC06. Mutants were also constructed in this WT background. All culturing occurred at 30 °C. Liquid cultures were grown in 100 mL BG-11 medium in 250-mL flasks (PYREX) and shaken at 150 rpm (Thermo Fischer MaxQ 2000 Orbital Shaker).

Chlorophyll Determination. Working under reduced irradiance to prevent degradation of extracted pigments, 1–2 mL cultures were harvested in triplicate in 2-mL tubes by centrifugation at

15,000 × g at laboratory temperature for 7 min. The supernatant fraction was removed, and cells were resuspended in 1 mL cold 100% (vol/vol) methanol. Samples were placed in a light-free container incubated at 4.0 °C for 1 h to extract the pigments from the cells. After incubation, cellular material was centrifuged at 15,000 × g for 10 min at 4.0 °C, and the supernatant was used for spectrophotometry. A Beckman Coulter DU 640B Spectrophotometer was calibrated using methanol as a blank, and absorbance at 665 and 720 nm was measured. Chlorophyll *a* concentrations were determined using the following equation: chlorophyll *a* (micrograms/milliliter) = 12.9447 (A665 – A720) (93).

Whole-Spectrum Light Absorption. Whole-cell absorption spectra were measured every 2 nm using an Infinite 200 PRO Multiplate Reader (Tecan) from 400 to 800 nm. Before measurements, culture densities were adjusted to an OD₇₅₀ of 0.05. Measurements using BG-11 medium as a blank were subtracted from the sample results, and then, results were normalized to OD₇₅₀ for comparison.

Modeling of Photon Uptake. Absorption spectra were collected as described above, corrected for light scattering by subtracting the OD₇₅₀ value from each absorbance value, and then, normalized to a 1-cm path length. The chlorophyll *a*-normalized optical absorption cross-section was calculated from the absorbance by the following equation (modified from ref. 33):

$$a_{\lambda}^* = 2.303 \frac{\text{absorbance}}{[\text{chlorophyll } a]}$$

The photosynthetically active radiation (PAR) range (400–700 nm) was divided into 15 bins of 20 nm each, and the a_{λ}^* values across each bin were averaged to give a spectrally averaged, chlorophyll *a*-normalized absorbance cross-section for each bin. The SE for each bin was determined from the SD of the in vivo absorption spectrum. The spectral distribution of the Osram Sylvania Octron (R) 741 Fluorescent Lamp was obtained from the manufacturer (www.sylvania.com/en-us/Pages/default.aspx). The spectrum was divided into 15 bins of 20 nm each, and the area of each bin was calculated using the trapz function in the NumPy Python package (www.numpy.org/).

After dilution to minimize light scattering, chlorophyll *a* measurements were used to normalize the absorption spectrum. The ratio of chlorophyll *a* to OD remained constant during the time course, indicating a lack of significant photoacclimation (Fig. S1). Because the model's biomass equation explicitly defines the cellular composition, the biomass-normalized photon absorption rate, $E_{a(\lambda)}$ [micromoles photons gram dry weight (DW)⁻¹ hour⁻¹], was calculated from the combination of irradiance $E_{0(\lambda)}$ (micromoles photons meter⁻² second⁻¹), optical absorption cross-section a_{λ}^* [centimeters² milligram chlorophyll *a* (Chl *a*)⁻¹], and the chlorophyll *a* component of the biomass equation [modified from P259 (94)]:

$$E_{a(\lambda)} = \frac{\text{mg Chl } a}{\text{g DW}} \int_{\lambda_1}^{\lambda_2} E_{0(\lambda)} a_{\lambda}^* d\lambda.$$

The resulting photon absorption flux was modeled as 15 metabolites, each representing a 20-nm segment of the 400- to 700-nm photosynthetic range.

Physiological Assays. *S. elongatus* was grown under 100 $\mu\text{mol photons m}^{-2} \text{s}^{-1}$ from cool white fluorescent bulbs. Growth was monitored by taking OD_{750} measurements and dry cell weight measurements at six time points over the course of 10 d. Cells were diluted appropriately to achieve OD_{750} measurements between 0.05 and 1.0. Dry cell weight was determined by vacuum filtration of 50 mL culture material using preweighed 0.45- μm hydrophilic polypropylene filters (47-mm GH Polypro; p/n 66548; Life Sciences). Filter disks containing cellular material were then placed in a large glass petri plate, dried at 90 °C for 1 h, allowed to cool to room temperature, and weighed. The initial weight of the filter disk was subtracted from the final to get the weight of the dry cells. Filters were placed back in for an additional 30 min and weighed again to ensure that drying was complete.

Oxygen Evolution. Activities of photosynthetic oxygen evolution were determined using a Clark-type oxygen electrode (ALGinstruments). The cells grown under 100 $\mu\text{mol photons m}^{-2} \text{s}^{-1}$ light at 30 °C were placed in 15-mL sterile conical tubes and quickly transported to another room for oxygen evolution analysis. Cells with chlorophyll *a* concentrations of 1–2 $\mu\text{g/mL}$ were placed in a water-heated apparatus and kept at 30 °C. Light was provided at various intensities following a regimen of initial dark for 5 min followed by 2 min of constant illumination and 2 min of complete darkness. Light intensities included 0, 50, 100, 500, 1,000, and 2,000 $\mu\text{mol photons m}^{-2} \text{s}^{-1}$. BG-11 medium was used as a blank, and rates were calculated by dividing by chlorophyll *a* concentration.

Simulations of in Silico Growth Rates. Growth rates were simulated for a 100-mL culture of *S. elongatus* in a 250-mL shaken flask. Irradiance (micromoles photons meter⁻² second⁻¹) was measured outside the flask using a QSL-100 PAR Irradiance Sensor (Biospherical Instruments Inc.). A correction factor for light loss caused by the apparatus was calculated by taking light measurements inside an empty flask at different irradiances and plotting the outside irradiance vs. the inside irradiance. A correction factor was derived from a linear regression of the data. The photon delivery rate was determined from the light source irradiance and the surface area of the flask. The flask was modeled as a frustum of a cone, with the lower radius equal to the bottom of a 250-mL flask and the upper radius equal to the flask radius at the culture height (culture volume = 100 mL). The lateral surface area was calculated and used to derive the photon delivery rate (micromoles photons second⁻¹) by multiplying the photon flux density (micromoles photons meter⁻² second⁻¹) by the calculated surface area (square meters). Self-shading was determined by dividing the 100 mL culture volume into 50 sections totaling 2% of the culture biomass in each section. For each section, the photon absorption rate and the photon delivery rate of each 20-nm bin were compared, and the lesser of two values was set as the upper and lower bounds of the photon exchange reaction. For each section (*n*), the photon delivery rate [$\nu_{(\lambda_{bin})_n}$] was equal to the initial photon delivery rate [$\nu_{(\lambda_{bin})_0}$] minus the photon absorption of all previous slices:

$$\nu_{(\lambda_{bin})_n} = \nu_{(\lambda_{bin})_0} - \sum_{i=1}^{n-1} \nu_{EX_photon(\lambda_{bin})_i}$$

Growth curves were simulated by dividing the culture duration into 240 1-h segments. To avoid numerical precision issues, the flux units were converted to micromoles milligram DW⁻¹ hour⁻¹. At each time point, the biomass-normalized constraints were converted to the total metabolite flow across the reaction for the 1-h time period. For example, at a time *t* where the biomass equals 100 mg dry cell weight, the nongrowth-associated maintenance constraint, which is 0.071 $\mu\text{mol mg DW}^{-1} \text{h}^{-1}$, becomes 7.1 μmol . After accounting for light attenuation caused by

self-shading, biomass was maximized for each of 50 culture sections as follows:

Maximize $\nu_{biomass}$.

Subject to

$$S \cdot \nu = 0$$

$$lb_i \leq \nu_i \leq ub_i$$

The following constraints were set before simulation:

$$\nu_{NADTRHD} = 0 \text{ (transhydrogenase set to 0),}$$

$$\nu_{TALA} \geq 0 \text{ (transaldolase set as irreversible),}$$

$$\nu_{ORNNTA} = 0 \text{ (ornithine transaminase set to 0),}$$

$$\nu_{CYOOxm} = 0 \text{ (cytochrome oxidase set to 0),}$$

$$\nu_{LDH_D} \leq 0 \text{ (lactate dehydrogenase set as irreversible),}$$

$$\nu_{EX_o2} \leq 153(1 - e^{(-1.4x/153)})e^{(-9.3 \times 10^{-14}x/153)}, \text{ where } I = \text{irradiance [oxygen evolution set to the experimental values with Platt fitting (ref. 95, pp. 687–701)],}$$

$$\nu_{PSII} \leq 7 \times S_{mgDW} \times 1 \text{ h (maximum flux through photosystem II), and}$$

$$0.01 \times \nu_{TKT2} \geq \nu_{PKETF} + \nu_{PKETX} \text{ (metabolite channeling constraint of phosphoketolase pathway),}$$

where S_{mgDW} is equal to 2% of the biomass of a given time point. Nongrowth-associated maintenance was set to 0.071 $\mu\text{mol mg DW}^{-1} \text{h}^{-1}$ for all simulations. The biomass yield for each time point was the summation of the biomass output of each of the 50 sections. Inorganic phosphate use was tracked to determine when the culture entered phosphate limitation-induced stationary phase. Biomass yield was converted to OD using a standard curve (Fig. S2). Simulations were performed using the COBRAToolbox (96) optimizeCbModel function in Matlab (Mathworks) with Gurobi 6.5.1 (Gurobi Optimization).

Intracellular Flux Distribution During Linear Growth. Reaction fluxes were taken from a time point in the linear portion of the growth curve. Simulations were constrained as above, with the exception that a secondary objective of minimizing the taxicab norm of the flux vector was applied (97). The metabolite flow in micromoles was normalized to the biomass of the given time point to convert the units to millimoles gram DW⁻¹ hour⁻¹.

In Silico Essential Gene Comparison. The previous *S. elongatus* model, *iSyf715*, was downloaded in SBML format, and the gene reaction rules were changed to a Boolean format and added to the model based on the supplemental files in the publication (48). The growth rate was fixed to 0.02 h⁻¹, and the minimum photon uptake to achieve the set growth rate was determined using the optimize_minimal_flux function in the COBRAPy Python package (84) in iPython Notebook (85) using the Gurobi 6.5.1 solver (Gurobi Optimization). Photon flux to the photosystem II reaction (`_lightII_r`) was set to 7.382 mmol g DW⁻¹ h⁻¹, and photon flux was set to 7.18 mmol g DW⁻¹ h⁻¹ for the photosystem I reaction (`_lightI_r`). All other constraints were left at their default values; the CO₂ and bicarbonate default uptake rates were both 1.99 mmol g DW⁻¹ h⁻¹. The photon uptake rate for *iJB785* was set by varying the light irradiance until a growth rate of 0.02 h⁻¹ was achieved. All other constraints were set to the same value as in *iSyf715*, with the exception of the nongrowth-associated maintenance, which was set to the calculated value of 0.071 mmol g DW⁻¹ h⁻¹. The in silico essential gene assessment for both models was made using the single_gene_deletion function in COBRAPy (84). Gene deletions where greater than 80% of the target growth rate was still achieved were considered to reflect nonessential genes, 80–10% were considered as beneficial genes,

and below 10% were labeled essential genes. Comparison between RB-TnSeq in vivo essentiality calls and in silico assessments were performed using in-house Python scripts in iPython Notebook (76).

TCA in Silico Simulations. The impact of TCA gene KO's on growth rate was simulated by setting the reaction flux to zero for the TCA enzymes fumarase and malic enzyme. Growth curves were simulated as outlined above at a model irradiance of $68 \mu\text{mol m}^{-2} \text{s}^{-1}$ ($\sim 90\text{--}100 \mu\text{mol m}^{-2} \text{s}^{-1}$ in experimental conditions). The TCA cycle was completed in silico by adding the alpha-ketoglutarate dehydrogenase and succinyl-CoA synthase genes from the *E. coli* model *iJO1366* (98). Growth curves and reaction fluxes were determined as outlined above. The TCA bypass reaction alpha-ketoglutarate decarboxylase was manually added to the model, and the succinyl-CoA dehydrogenase reaction from the *E. coli* model *iJO1366* was added to complete the bypass. Growth curves and reaction fluxes were determined as outlined above. The malate dehydrogenase (BiGG ID code MDH) reaction from the *E. coli* model *iJO1366* was added to the model. Growth curves and reaction fluxes were determined as outlined above. For simulations in the dark, the light uptake was set to zero, and the glucose released by glycogen degradation was set to an arbitrary rate of $6 \text{ mmol g DW}^{-1} \text{ h}^{-1}$. The objective function was set to maximize ATP production.

Generating and Assaying TCA Cycle Mutants. Mutants were generated from plasmids taken from the unigene set, an arrayed mutant library for *S. elongatus* (99, 100). Standard transformation protocols were used for mutant generation (4), and genotyping was done using colony PCR (Table S2) with Taq DNA Polymerase (NEB). Liquid culture assays were conducted in BG-11 medium containing kanamycin ($5 \mu\text{g/mL}$) under light levels of $\sim 100 \mu\text{mol photons m}^{-2} \text{s}^{-1}$, with OD_{750} taken every 24 h.

For spot plates, $4 \mu\text{L}$ culture was plated onto solid BG-11 kanamycin medium in a 1:5 dilution series. Constant light-incubated spot plates were put under light levels of $\sim 100 \mu\text{mol photons m}^{-2} \text{s}^{-1}$ for 5 d. Light-dark-incubated plates were grown in a 12:12 -h cycle with square transitions at the same light intensity for 8–9 d. Colony area was measured using ImageJ analysis software (80).

SI Results

Structural Homology Modeling. PGMs fall into two categories that are structurally distinct: dPGM (Synpcc7942_2078, Synpcc7942_1516, and Synpcc7942_0485, EC 5.4.2.11) and cofactor-independent phosphoglycerate mutase (iPGM; Synpcc7942_0469, EC 5.4.2.12). The iPGM family performs the mutase reaction exclusively, whereas the dPGM family has been assigned various catalytic functions (86). Because the enzymatic activity of a given PGM is structure-dependent, we attempted to categorize the *S. elongatus* PGM reaction specificity through structural analysis.

Because the iPGM family has only been shown to perform the mutase reaction, structural homology modeling focused on the dPGM family. We generated structural homology models for the *S. elongatus* dPGMs using an automated in silico platform for protein structural prediction (88). The resulting homologous enzyme scaffolds included a dPGM from *E. coli* (86) and a PSP from *H. thermophilus*. Interestingly, the *S. elongatus* network reconstruction indicated a gap in the gene assignment for PSP (EC 3.1.3.3). The *H. thermophilus* PSP is a member of the dPGM

family, and the crystal structure, including features necessary for catalytic function, has been elucidated (24). Structural features of the resulting homology models were compared with the *E. coli* dPGM and the *H. thermophilus* PSP controls to refine the functional annotation of the enzymes (SI Materials and Methods).

S. elongatus synpcc7942_0469 is the only gene encoding an iPGM, and it is essential in vivo. Thus, synpcc7942_0469 was annotated as the primary glycolytic PGM in *S. elongatus*. The protein encoded by synpcc7942_2078 shares structural features with the *E. coli* dPGM control and lacks features that are important for PSP activity in *H. thermophilus*. Because it is non-essential in vivo, synpcc7942_2078 was annotated as a dPGM: possibly performing the “reverse regulatory” function with Synpcc7942_0469 as proposed previously (23). These researchers also suggested that Synpcc7942_0485 functions as a PSP, and the recently characterized PSP in PCC 6803 has amino acid homology to Synpcc7942_0485 (25) (Table S1).

Synpcc7942_0485, its homolog in PCC 6803 (*str1124*), and the *H. thermophilus* PSP shared strong structural similarity; synpcc7942_1516 was essential in vivo, and the protein carried structural features that could not be classified into a dPGM or PSP. Its genomic neighbor, synpcc7942_1517, encodes an essential cyanobacterial-conserved histidine kinase, and transcriptome mapping data indicated that synpcc7942_1516 and synpcc7942_1517 are coexpressed on the same transcript (26). Cyanobacteria have a variety of two-component systems comprising a histidine kinase and at times, a phosphatase to regulate signal transduction activity (101). We hypothesized that synpcc7942_1516 encodes a histidine phosphatase regulator of an uncharacterized cyanobacterial two-component system. As a regulatory enzyme, Synpcc7942_1516 fell outside the scope of the metabolic model and was not included in the model gene list.

Suggested Constraints for Modeling in Constant Light.

Upper and lower flux bounds through NADP:NADH transhydrogenase reaction set to zero ($\nu_{\text{NADTRHD}} = 0$). Justification is given in the text.

The lower bound of the transaldolase reaction set to zero ($\nu_{\text{TALA}} \geq 0$). Corrects a central carbon essentiality discrepancy for sedoheptulose-1,7-bisphosphatase (BiGG ID code SBP), resulting in a more accurate intracellular flux map.

Upper and lower flux bounds through ornithine transaminase set to zero ($\nu_{\text{ORNTA}} = 0$). Ornithine transaminase connects proline and arginine biosynthesis. If active, neither of those pathways would be essential. However, both proline and arginine biosynthesis pathways are essential in vivo.

Upper and lower flux bounds through cytochrome oxidase set to zero ($\nu_{\text{CYOoxm}} = 0$). Presence allows for a biologically infeasible pseudocyclic electron flow. Fluorescence microscopy experiments in *S. elongatus* indicated that the concentration of this complex is very low based on the inability to visualize GFP constructs (35). Transcriptomics data also indicated very little expression of this complex (26).

The upper bound of the transaldolase reaction set to zero, making it irreversible away from pyruvate ($\nu_{\text{LDH}_D} \leq 0$). Avoids biologically irrelevant bypass of lower glycolysis through cell wall degradation into lactate and conversion back into pyruvate.

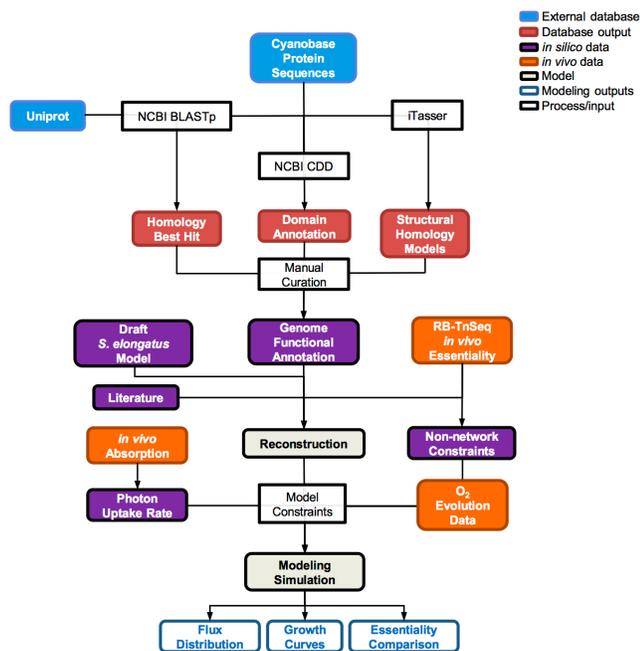


Fig. S1. Model development flowchart. This diagram outlines the steps that went into genome reconstruction and model simulation. CDD, Conserved Domains Database; NCBI, National Center for Biotechnology Information.

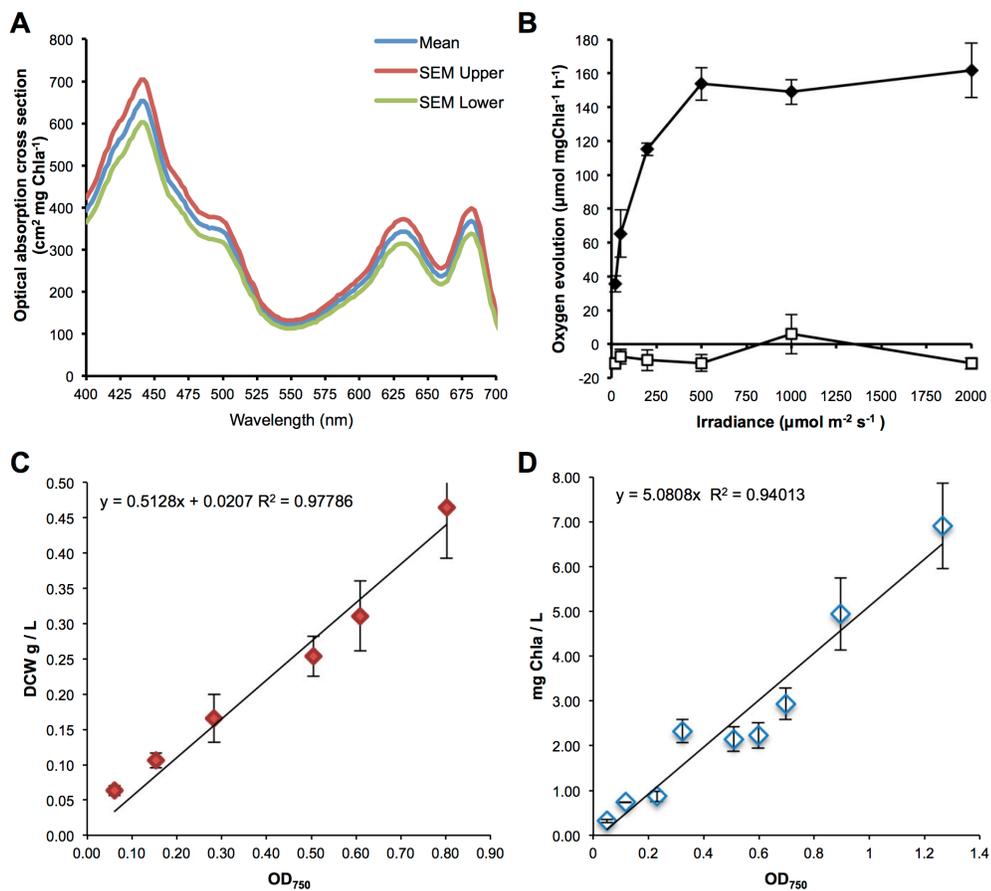


Fig. 52. In vivo physiological data incorporated as constraints in *in silico* flux balance analysis. (A) Whole-cell absorption spectra. (B) Oxygen evolution under illumination at various light intensities shown as ♦. Dark condition oxygen consumption after each light period shown as ◻. (C) Dry cell weight (DCW) vs. OD₇₅₀. (D) Chlorophyll a concentration vs. OD₇₅₀. The data shown for each graph are averages of three independent measurements, and error bars indicate SDs.

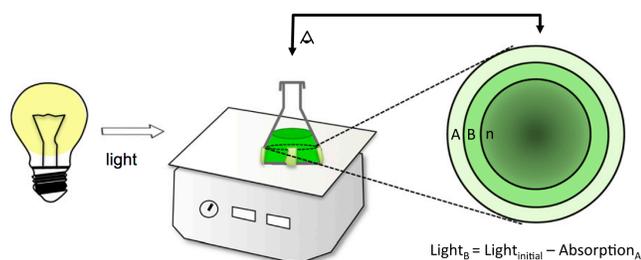


Fig. 53. Diagram of cell-shading calculation. Increased light limitation caused by cell-cell shading was factored into the model simulations for biomass prediction in a typical 250-mL Erlenmeyer flask with 100 mL culture medium.

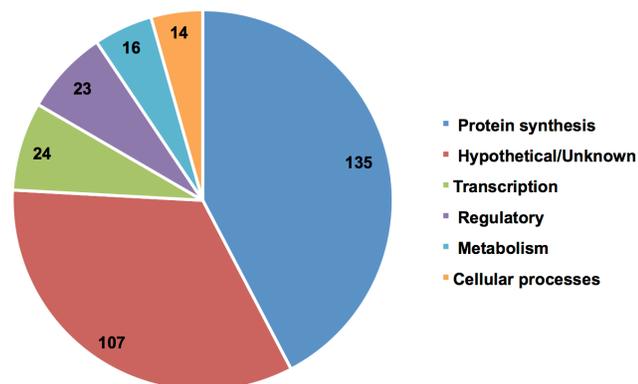


Fig. 54. Essential genes in *S. elongatus* not present in *iJB785*. The number of essential genes in each category is indicated.

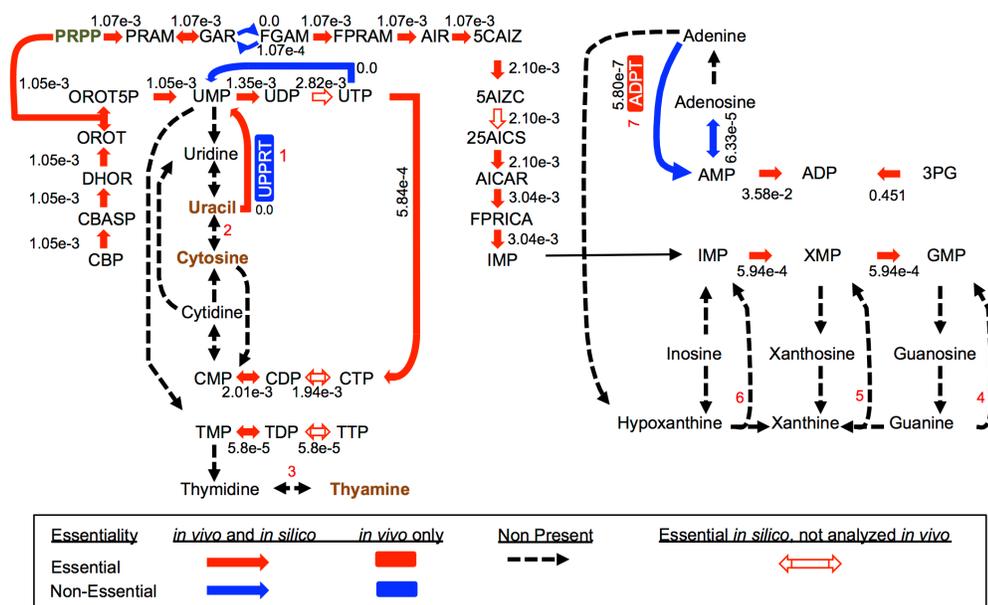


Fig. 55. Nucleotide salvage pathway. Reactions not present in *S. elongatus* are indicated by dashed black arrows, essential reactions are indicated by red arrows, nonessential reactions are indicated by blue arrows, and those essential in the model but not analyzed *in vivo* are indicated by red-outlined arrows. Reactions important for nucleotide salvage are numbered as follows: 1, uracil phosphoribosyltransferase (UPPRT); 2, cytosine deaminase; 3, thymidine phosphorylase; 4, guanine phosphoribosyltransferase; 5, xanthine phosphoribosyltransferase; 6, hypoxanthine phosphoribosyltransferase; 7, and adenine phosphoribosyltransferase (ADPT). Reaction and metabolite abbreviations are given in BiGG format (bigg.ucsd.edu) and found in Dataset S2.

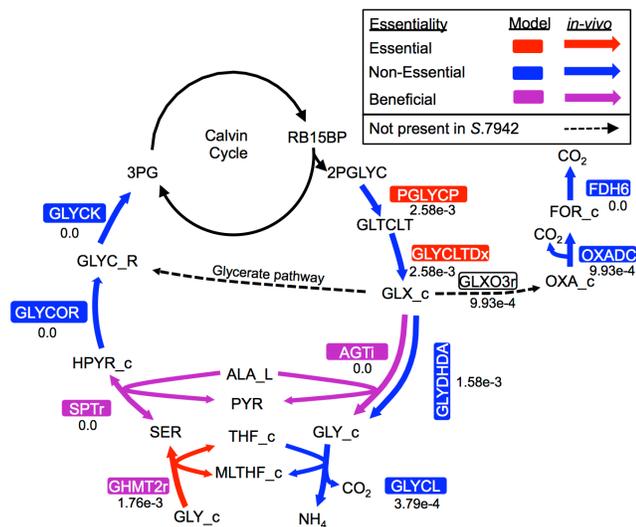


Fig. S6. Photorespiration reactions. Beneficial arrows (purple) indicate a growth defect phenotype when a gene is mutated. Reaction and metabolite abbreviations are given in BiGG format (bigg.ucsd.edu) and found in Dataset S2.

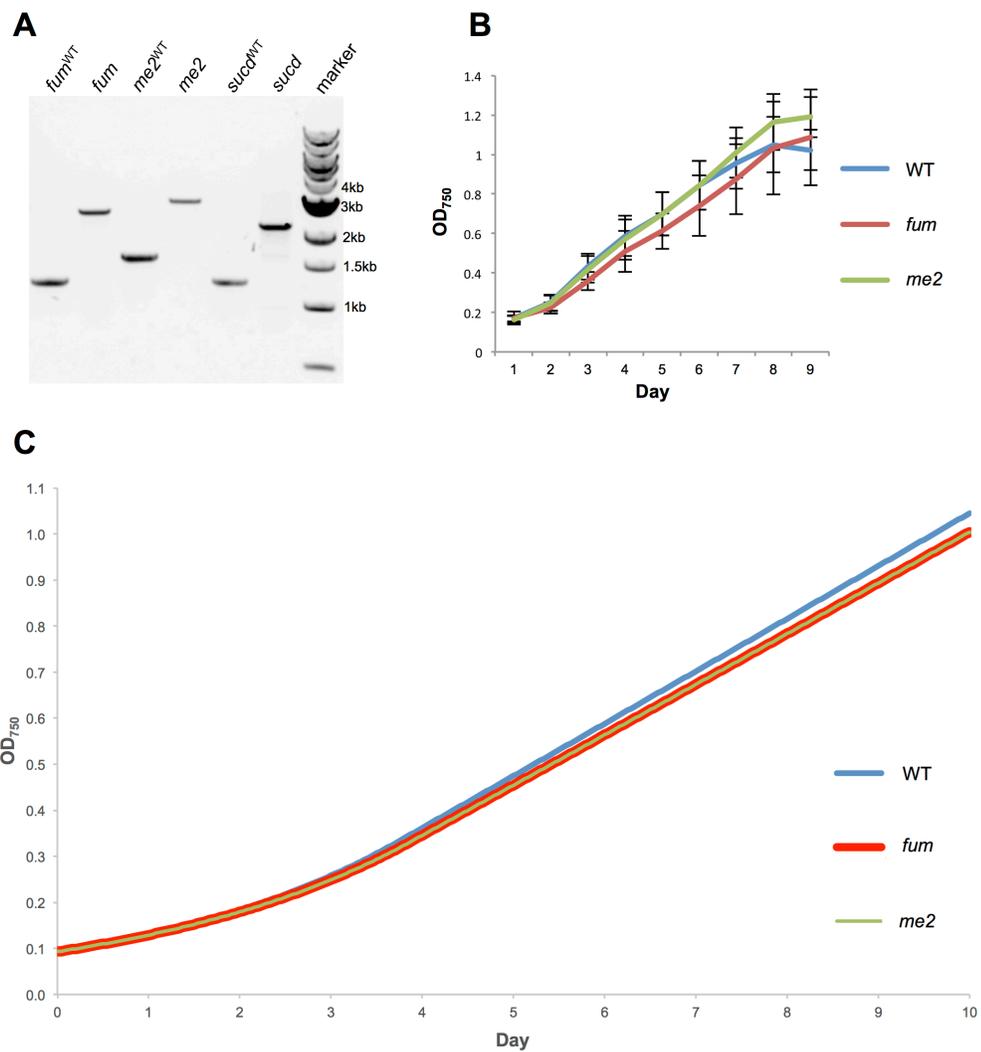


Fig. 57. Genotypic characterization and growth curve of the TCA cycle mutants. (A) Lane 1 (*fum*^{WT}), amplification of WT DNA with primers surrounding the *fum* gene (*synpcc7942_1007*); lane 2 (*fum*), amplification with the same primers from *fum* transposon insertion KO mutant (8542-O6), in which a 1.3-kb insertion is present; lane 3 (*me2*^{WT}), amplification of WT DNA with primers surrounding the gene encoding malic enzyme (*synpcc7942_1297*); lane 4 (*me2*), amplification with the same primers from the transposon insertion KO mutant for the gene encoding malic enzyme (8529-J6); lane 5 (*succ*^{WT}), amplification of WT DNA with primers surrounding the *succ* subunit B gene (*synpcc7942_1533*); lane 6 (*succ*), amplification with the same primers from the *succ* transposon insertion KO mutant (851-JJ4); and lane 7 (marker), standard 1-kb ladder (New England BioLabs). Each band is representative of three colonies tested. (B) Growth of WT, fumarase mutant, and malic enzyme mutant strains in liquid culture. Error bars show SDs of three independent replicates for each mutant. (C) Single-gene deletions of each of the TCA cycle genes *fum* (*synpcc7942_1007*) and *me2* (*synpcc7942_1297*) were performed in silico, and the resulting impact to growth rate was analyzed.

Table S1. Structural homology analysis of *S. elongatus* PGMs

Gene	Organism	Annotation	C-terminal chain	His85	Gln22	Conclusion	Ref.
HTH_0103	<i>H. thermophilus</i>	PSP (control)	+	+	+	Control	24
b0755	<i>E. coli</i>	dPGM (control)	-	Tyr	Thr	Control	86
Synpcc7942_0485	<i>S. elongatus</i>	PGM	+	+	+	PSP	
Synpcc7942_2078	<i>S. elongatus</i>	PGM	-	Phe	Ser	dPGM	
Synpcc7942_1516	<i>S. elongatus</i>	PGM	+	Leu	Leu	Histidine phosphatase	
Synpcc7942_0469	<i>S. elongatus</i>	iPGM				iPGM	

Table S2. Primers for validation of the TCA pathway mutants

Purpose	Primer name	Forward primer (5' to 3')	Reverse primer (5' to 3')
Segregation check of fumarate hydratase mutant	Synpcc 1007 F/R	AACATCAAATCCAGTCGGCG	TCATTTGCCCATTTACTCGCG
Segregation check of malic enzyme	Synpcc 1297 F/R	CATTAAGACTCTTCGCAC	GAAACAATATCTGCCCTAC
Segregation check of succinate dehydrogenase subunit B	Synpcc 1533 F/R	TTTGGGTACGGCCTATT	CTCTAGAACAACCTGAATCC

Other Supporting Information Files

[Dataset S1 \(DOCX\)](#)

[Dataset S2 \(XLSX\)](#)

[Dataset S3 \(XLSX\)](#)

3.4 RB-TnSeq Screens: Biofilm Formation

Introduction. Although our previous work has identified gene products that are involved in repressing or enabling biofilm formation in *S. elongatus*, there are still many unknowns. These poorly understood areas include the repression of biofilm formation that occurs in the laboratory WT strain as well as the mechanisms that allow biofilm formation when this repression is blocked. To gain a more global and thorough list of genes involved in cyanobacterial regulation of biofilm formation, we performed two next-generation sequencing-based experiments: RNA-Seq and RB-TnSeq. These two data sets not only illuminated novel genes of interest, but also demonstrate the respective benefits of these two experimental techniques for revealing genotype-phenotype associations in the context of complex behaviors.

RB-TnSeq Biofilm Screening Method. To take advantage of Rb-TnSeq for understanding genotypic contributors to biofilm formation, we grew replicate biological samples of the RB-TnSeq library under the biofilm-forming conditions performed for the RNA-Seq experiments, but for a two-week time course rather than four days (Fig. 3.4-1). Under these conditions the library samples generated biofilms. These cultures were then split into three fractions for sequencing analysis: (1) planktonic cells that were removed by decanting the media, (2) settled cells that were removed from the emptied test tubes with gentle water washes, and (3) biofilmed cells that required scraping to be removed from the test tubes. Bar codes from cells in each fraction were then amplified and

sequenced with the aim of identifying mutants that are over- and under-represented in the biofilm-forming portion of the population.

RB-TnSeq Results. Two experiments were conducted using the method described above in order to reveal biofilm-involved mutants. In the first experiment, 34 hits were enriched in the biofilming fraction while no over- or under-represented mutants were found in the settler fraction. Included in the 34 hits were all six previously published biofilm-forming mutants. In the second experiment, 6 hits were significant in only the biofilming fraction with an additional 18 hits being significant in both the biofilming and settler fractions. Four of the six known biofilm-forming mutants were significantly over-abundant in the biofilming fraction and significantly absent from the planktonic fraction. This result demonstrates the precision of this technique for uncovering genotypic-phenotypic relationships while providing a manageable set of ranked candidate genes to investigate through experimental validation.

Rb-TnSeq vs. RNA-Seq Validation. To compare the efficacy of these two techniques in providing genetic-phenotypic correlations concerning biofilm formation, a number of putative targets were validated based on either the RNA-Seq or RB-TnSeq data. In total, 29 genes were selected based on RNA-Seq data, while 39 genes were selected based on RB-TnSeq data. Of the 29 RNA-Seq-derived candidates, only one showed film formation when mutated. Of the 39 genes investigated based on RB-TnSeq data, 13 mutants produced films in

test tubes, 96-well plates, or flasks. It is likely that the effectiveness of Rb-TnSeq in identifying biofilm-involved genes can be attributed to its direct identification of mutants present in biofilms, as opposed to RNA-Seq where gene involvement in biofilm formation is inferred indirectly from patterns of gene expression.

Figures.

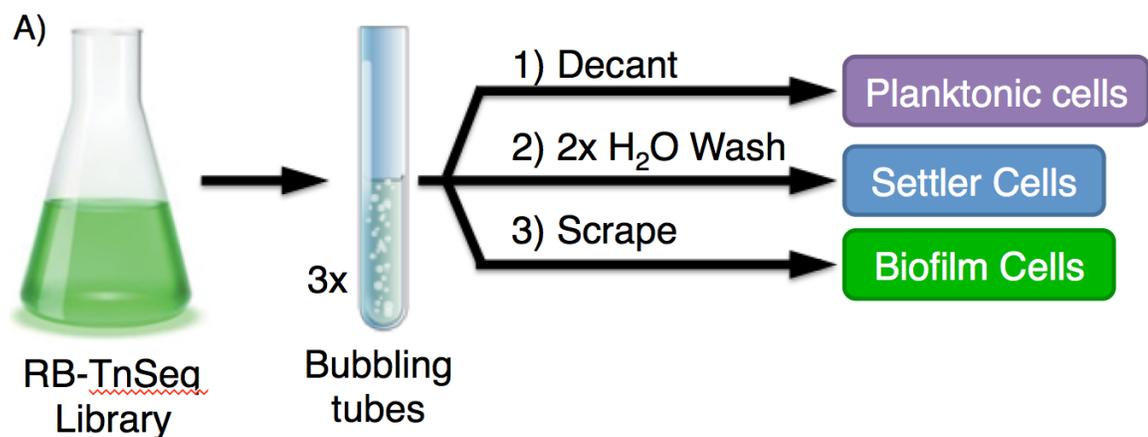


Figure 3.4-1: RB-TnSeq for biofilm formation. (A) Experimental setup for RB-TnSeq of biofilming cultures. The RB-TnSeq library was grown as a starter culture and used to inoculate bubbling test tube cultures. After 2 weeks, each test tube was fractionated to produce planktonic, settler, and biofilming fractions. Triplicate experiments were pooled to acquire enough mass for sequencing each fraction.

3.5 RB-TnSeq Screens: Amoeba Grazing

Introduction. In their natural environment, cyanobacteria are exposed to a wide variety of stressors, both biotic and abiotic. Grazing by protozoan predators such as amoebae is a major source of cyanobacterial mortality, and is a complex multistep activity (Jousset, 2011). A broad, unbiased search for genes in cyanobacteria that influence fitness under selection by grazing will implicate the cellular processes involved and will contextualize the stress of grazing within the landscape of environmental stimuli. Here, a whole-genome search for the genetic basis of susceptibility to cyanobacterial grazing was conducted using RB-TnSeq.

Amoeba Grazing Screen Method. In fitness experiments using the amoebae HGG1 and LPG1 on solid agar media, the RB-TnSeq library was grown as lawns on plates at high light. Mature lawns were transferred to lower light to slow cyanobacterial growth, and were inoculated with a liquid suspension of amoebae spotted at the center of each plate. Amoebae initially graze the bacteria at the site of inoculation and then spread outwards, resulting in an expanding yellow plaque that indicates where cyanobacteria have been grazed. After the amoebae completely cleared the cyanobacterial lawns, plates were moved back to high light to encourage the rapid recovery and growth of surviving cyanobacterial cells. As a negative control, lawns of the library were grown in the same manner without the amoebal inoculation. We then harvested all lawns, extracted genomic DNA, and conducted RB-TnSeq sequencing and analysis.

Results RB-TnSeq on solid media enabled clear differentiation of genes with amoeba-resistant phenotypes (Fig. 3.5-1). Many of these genes have common functions in the biosynthesis of lipopolysaccharide, particularly in the attachment of O-antigen, indicating that the disruption of this process at various steps in the pathway allows resistance to grazing on solid media. These findings expand on earlier experiments that showed impairment of O-antigen synthesis provides protection against grazing (Simkovsky et al., 2012). It is clear from the RB-TnSeq experiment that the protective effect of an incomplete LPS synthesis pathway is valid against LPG1 as well as the distant HGG1 species, suggesting that the phenotype is not highly specific to particular grazers, and may apply to other grazers of *S. elongatus*. In addition, several genes had markedly different fitness values between grazing by LPG1 and by HGG1, indicating resistance mechanisms that are grazer-specific. This work both illuminates important pathways of amoeba grazing and more broadly identifies RB-TnSeq as a useful tool for illuminating predator prey interactions.

Figures.

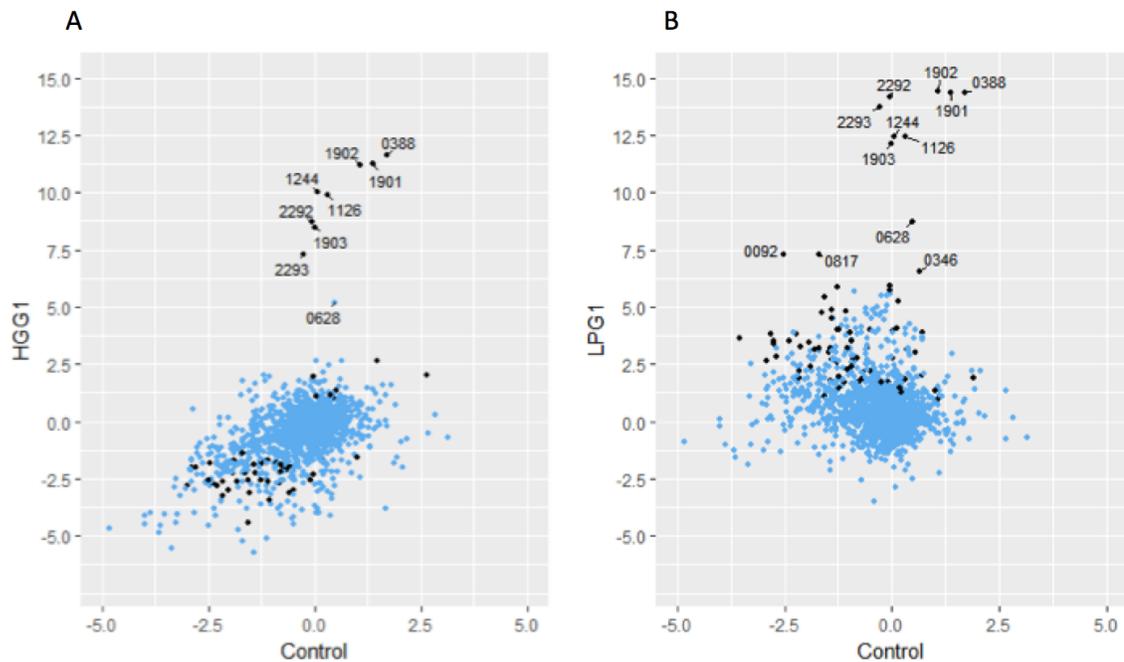


Figure 3.5-1: HGG1 and LPG1 grazing. Comparison of mutant fitness values in solid grazing conditions. **(A)** Mutant fitness in ungrazed control plates (x axis) and plates grazed by amoeba HGG1 (y axis). **(B)** Mutant fitness in ungrazed control plates (x axis) and plates grazed by amoeba LPG1 (y axis). Values shown are averages of two replicates. Black dots indicate significance for both replicates in the grazed condition for a gene; blue dots indicate that at least one of the grazed conditions was insignificant. Gene numbers for high-fitness mutants are labeled.

References.

- Jousset, A., 2011. Ecological and evolutive implications of bacterial defences against predators. *Environ. Microbiol.* 14, 1830–1843.
- Simkovsky, R., Daniels, E.F., Tang, K., Huynh, S.C., Golden, S.S., Brahamsha, B., 2012. Impairment of O-antigen production confers resistance to grazing in a model amoeba–cyanobacterium predator–prey system. *Proc Natl Acad Sci USA* 109, 16678–16683.

3.6 RB-TnSeq Screens: Analysis Across Conditions

The library is a powerful tool for identifying genes that are beneficial or detrimental under different growth conditions. To identify phenotypes for previously hypothetical genes, we screened the constituents of the *S. elongatus* library under 30 stress conditions (Price et al., 2016). These data were grouped with data from similar screens on 24 proteobacteria. The conditional fitness phenotypes for these 25 bacteria under hundreds of screening conditions were analyzed to assign cellular roles to 8,456 genes which previously had no known function. This annotation amounts to functional predictions for 14% of all sequenced bacterial genes that previously had no known role.

References.

Price, M.N., Wetmore, K.M., Waters, R.J., Callaghan, M., Ray, J., Kuehl, J.V., Melnyk, R.A., Lamson, J.S., Suh, Y., Esquivel, Z., Sadeeshkumar, H., Chakraborty, R., Rubin, B.E., Bristow, J., Blow, M.J., Arkin, A.P., Deutschbauer, A.M., 2016. Deep Annotation of Protein Function across Diverse Bacteria from Mutant Phenotypes. bioRxiv.

3.7 Acknowledgments

Chapter 3, in part, is made up of reprints of two published manuscripts and two manuscripts in preparation. The first published manuscript in section **3.2 RB-TnSeq Development and Essential Genes in *S. elongatus*** is: Rubin BE et al. (2015) The essential gene set of a photosynthetic organism. *Proc Natl Acad Sci*. 112(48). The dissertation author was the primary author of this paper. The second published manuscript in section **3.3 RB-TnSeq Guided Metabolic Modeling** is: Broddrick J*, Rubin BE*, Welkie DG*, Nui D, Nathan M, Diamond S, Jenny JL, 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* 113(51). The dissertation author was the co-primary author of this paper. The first manuscript in preparation in section **3.4 RB-TnSeq Screens: Biofilm Formation** is: Simkovsky R, Rubin BE, Wang J, Nagar E, Parnasa R, Yevgeni J, Veltman B, Sendersky E, Schwarz R, Golden SS. RNA-Seq vs RB-TnSeq: Each reveal critical elements in biofilm repression and formation in the cyanobacterium *Synechococcus elongatus* PCC 7942. (In Preparation). The dissertation author will be the secondary author of this paper. The Second manuscript in preparation in section **3.5 RB-TnSeq Screens: Amoeba Grazing** is: Ota M, Simkovsky R, Rubin BE, Golden SS, Brahamsha B, Golden JW. A genome-wide fitness assay identifies determinants of cyanobacterial resistance against protozoan grazing. (*In preparation*). The dissertation author will be the tertiary author of this paper.

CHAPTER 4: Applying RB-TnSeq to Light-Dark

Physiology

4.1 Chapter Summary

Although cyanobacteria live in LDCs in nature, and are being developed for bio-production outdoors, they are rarely studied under these conditions. This chapter addresses this deficiency largely by the application of RB-TnSeq. In section **4.2 RB-TnSeq Under Light-Dark Cycles**, which consists of a manuscript in preparation on which I'm second author, the RB-TnSeq library was screened under LDCs. This screen allowed the identification of mutants that grow poorly under LDCs, and hence genes that are important for surviving these conditions. Many of the candidates that came out of this screen fit our current paradigm for the circadian clock's role in the survival of LDCs. However, the top hit from this screen was one of three members of the *S. elongatus* core circadian clock oscillator, *kaiA*, whose mutants had never been recognized to be sensitive to LDCs. We have gone on to identify the mechanisms through which this mutant locks the clock in an LDC-sensitive state. These findings have led to a better understanding of the circadian clock and its role in surviving LDCs.

Section **4.3 The Role of c-di-AMP in Light-Dark Cycles**, which includes a manuscript in preparation on which I'm first author, presents the use of RB-TnSeq to understand the role of a newly discovered signaling nucleotide in LDCs. The molecule, cyclic-di-AMP (c-di-AMP), has been speculated to function in cyanobacteria, but never previously studied *in vivo*. Our work identified its

presence, and the responsible cyclase in *S. elongatus*. In addition, it showed that c-di-AMP quantity fluctuates over LDCs and that the organism is impaired in LDCs in the absence of the signaling molecule. Finally, we developed and implemented an RB-TnSeq based approach for quantitative whole-genome interaction screens (IRB-Seq). We applied IRB-Seq along with a traditional protein pull-down method to discover the interaction network of the signaling nucleotide. This section identifies c-di-AMP as an LDC-involved signaling molecule in cyanobacteria and describes the development of an approach for high-throughput quantitative interaction screens.

4.2 RB-TnSeq Under Light-Dark Cycles

Abstract. Recently, the creation of a dense transposon mutant library in *S. elongatus* quantified the fitness contribution of each individual gene in the genome under continuous-light conditions and described the first essential gene set for any photosynthetic organism. Here we describe results from screening this library to assess the genetic fitness contributions of each gene under cycling light-dark conditions. Intriguingly, loss of the core circadian clock protein KaiA is specifically detrimental under light-dark conditions, whereas loss of the entire *kaiABC* gene cluster is not. This work identifies the genes essential for growth under light-dark conditions in *S. elongatus* and explains how a proper functioning circadian program, particularly the presence of KaiA, confers a fitness advantage in naturally cycling environmental conditions.

Introduction. Circadian rhythms can be found throughout Nature, suggesting a clear fitness advantage for such timing mechanisms in a cyclic environment (Dunlap et al. 2003). For photosynthetic cyanobacteria that rely on conversion of light energy from the sun, many of the core metabolic processes including those involved with CO₂ fixation are clock controlled (Diamond et al. 2015; Ito et al. 2009). The circadian clock's contribution towards environmental fitness was first shown conclusively in experiments that mixed *Synechococcus elongatus* sp. PCC 7942 strains that had different intrinsic circadian periods under different natural or non-24-h day lengths; in each case, the strain whose intrinsic circadian period most closely matched the external light cycle outcompeted strains that have different periods (Ouyang et al. 1998; Woelfle et al.

2004). This fitness advantage of appropriate circadian signaling is not exclusive to cyanobacteria and is also observed in the plant model species *Arabidopsis* (Yerushalmi et al. 2011; Dodd et al. 2005), and in humans, where circadian disruption can result in a myriad of health issues including cardiovascular disease (Morris et al. 2012), cancer (Kelleher et al. 2014), mental illness (McCarthy and Welsh 2012), and sleep disorders (LeGates et al. 2014).

The central oscillators in all of the eukaryotic species studied share similar underlying feedback mechanisms (Bell-Pedersen et al. 2005). *S. elongatus* is the lone prokaryotic model organism used to study circadian biology, and its mechanism is fundamentally different than that described for eukaryotes. While the molecular mechanisms of the clock is well established in constant environments for *S. elongatus*, the physiological programs that confer a fitness advantage are not understood, and the switch from daytime (class 2) to nighttime (class 1) metabolic programs remain vague. To address the question of how circadian rhythms translate to physiological fitness in cycling environments, we employed the method of random bar code transposon-site sequencing (RB-TnSeq) to quantitatively identify the abundance changes of a pooled mutant population of *S. elongatus*. This approach provided a fitness contribution measure for each individual gene in the genome.

In this study we describe the metabolic processes that are essential for growth in light-dark cycling (LDC) conditions and report that disruptions to the circadian clock protein KaiA results in significant growth attenuation specifically in

LDCs. We further investigated why loss of KaiA results in more severe light-dark sensitivity than deletion of the complete gene cassette *kaiABC* (Iwasaki et al. 2000) or solely *kaiC* (Diamond et al. 2015), and gained new insights into protein interactions of the Kai oscillator that confer signal information to induce night-time essential processes.

Results and Discussion.

Genome-wide fitness contributions during photoautotrophic growth under light-dark cycles. The same dense bar-coded transposon mutant library used previously to identify essential genes in *S. elongatus* was grown under conditions of continuous light as well as alternating 12-h light, 12-h dark cycles (12:12 LDCs). Of the 2,723 genes comprising the genome of *S. elongatus*, 718 are essential (Rubin et al. 2015). Of the remaining 2,005 genes, data from insertions in 1,872 genes were analyzed and scored to determine fitness effects of gene knockouts in LDCs relative to continuous light. Mutants were scored based on two criteria: a fitness score representative of a > 1 generation difference and a false discovery rate of $\leq 1\%$. An initial 452 genes had significant fitness scores but 362 of these did not meet the fitness score cutoff that would indicate a strong phenotypic affect, *i.e.* little changes in abundance over the course of the experiment in LDC vs. constant light. The remaining 90 genes were divided approximately evenly between those significantly increasing in abundance in the population when mutated in LDCs (49 gene disruptions were beneficial in LDCs) and those

decreasing in abundance under LDCs (41 gene disruptions that caused sensitivity in LDC) (Figure 4.2-1A).

Genes essential for growth in LDCs included those involved in the oxidative pentose phosphate pathway such as *zwf*, *opcA*, *pgl*, *tal*, and *gnd* (Figure 4.2-1C). This pathway facilitates carbon flow from the breakdown of storage carbohydrates and is the major source of reducing equivalents in the absence of photosynthetic electron transfer. Both reactions for catabolism of glycogen, *glgP* and *glgX*, are also required for LDC growth. In addition, genes encoding enzymes needed for DNA repair (*mdf*), chaperone systems (*clpB1*), and circadian regulation scored highly. LDC growth defects in mutants with disruptions in some genes of the circadian regulatory network have been described previously. Early studies showed that strains of *S. elongatus* with mutations in *kaiC* are out-competed in direct competition with WT and established that clock timing provides a fitness contribution for the cell (Ouyang et al. 1998; Woelfle et al. 2004). Moreover, the clock output transcription factor RpaA is essential for growth in light-dark conditions, with rapid death in the dark mediated by redox unbalance (Diamond et al. 2017). These results are validated in this study, as *kaiC* and *sasA* mutants have a fitness score indicating strong sensitivity to LDCs (-1 and -1.5, respectively). Strains with disruptions in *rpaA* also showed LDC sensitivity, although with a more modest score of -0.7, presumably because the *rpaA* mutant is less fit under continuous light conditions as well; thus, its comparative fitness magnitude is reduced even though the

phenotype is severe. Notably, the highest scoring LDC-sensitive mutants were insertions in the circadian clock protein gene *kaiA*, with a score of -3.0.

Disruptions in a number of other genes were found to be beneficial to growth in LDCs. Among the mutant strains that thrived were those with that disrupt pilus assembly and protein export (8 genes), glycine cleavage and purine metabolism (6 genes), sigma factor RpoD2 involved in circadian regulation, and the circadian-associated phosphatase, CikA. Additionally, disruptions to genes involved with metal ion homeostasis such as *smtA* and *lipA*, cobalamin biosynthesis, *cobL*, and NADH dehydrogenase, *ndhD2*, allow cells to thrive under LDCs, possibly by lessening the level of stress acquired during the high-light conditions during the day period.

Locking the clock in a repressive state is fatal in LDCs. The mechanism of circadian regulation in cyanobacteria is fairly well understood (Cohen and Golden 2015). Three core clock proteins, KaiA, KaiB, and KaiC orchestrate circadian oscillation in *S. elongatus* (Ishiura et al. 1998), whereby KaiC undergoes a daily rhythm of phosphorylation and dephosphorylation at two amino acid residues (Nishiwaki et al. 2004; Xu et al. 2004). During the day period, KaiA stimulates KaiC autophosphorylation and during the night, KaiB opposes KaiA's stimulatory activity by binding an inactive form of KaiA and sequestering it (Tseng et al., 2017). Temporal information from the clock is relayed to the genome to generate rhythmic gene expression via two histidine kinase proteins, SasA and CikA, which regulate the rhythmic phosphorylation and dephosphorylation of the

transcription factor RpaA. RpaA is necessary for circadian rhythms of gene expression and directly binds to downstream genetic targets (Markson et al. 2013; Gutu and O'Shea 2013; Takai et al. 2006). As KaiC becomes phosphorylated during the daytime, it stimulates the autophosphorylation of the histidine kinase SasA (Gutu and O'Shea 2013) which then transfers a phosphate group to RpaA. Thus, phosphorylated RpaA accumulates during the day, peaking at the day-night transition, and activates the nighttime circadian program. During the night, SasA kinase activity decreases (Chang et al. 2015) and phosphatase activity of CikA increases. These influences together cause dephosphorylation of RpaA, repressing the nighttime circadian-regulated program in preparation for the day.

In a *kaiA* insertion mutant, KaiC and KaiB levels are low and KaiC is unphosphorylated (Ditty et al., 2005). For this reason, a *kaiA* mutant was previously expected to be similar to a *kaiC* null, which does not have a notable defect in LDC when grown in pure culture without competition. We hypothesized that without KaiA, the cell will generate less phosphorylated KaiC and cause an imbalance of SasA kinase and CikA phosphatase, leading to the failure to initiate processes needed for darkness (Figure 4.2-1B). Further, we predicted that the defect would be more severe in a deletion mutant that lacks all of the *kaiA* open reading frame, including a *cis* element that represses *kaiBC* expression, in which more unphosphorylated KaiC and KaiB are present, potentially activating CikA phosphatase. We assessed a number of *kaiA* mutant strains under LDC and continuous-light conditions. While no significant difference in growth is observed

between WT cells and those that lack *kaiA* ($KaiA^{\text{insertion}}$ and $KaiA^{\text{deletion}}$) under continuous light, growth of the *kaiA* mutants is severely attenuated under LDCs, particularly in the case of the complete deletion strain (Figure 4.2-2A). Moreover, in cells that lack KaiA, when the status of gene expression is visualized using a class 1 reporter (*PkaiBC::luc*) that serves as a proxy for *rpaA*-stimulated genes, bioluminescence is locked at or below the circadian trough in the *kaiA*-deficient strains (Figure 4.2-2B). This reduced luminescence indicates a clock signal that is locked in a repressive class 2 state (Paddock et al., 2013).

We further hypothesized that inactivation of *cikA* in a deletion background would increase the magnitude of expression from the RpaA-dependent reporter gene and suppress the phenotype of the *kaiA* mutant. These results are evident in Figure 4.2-2A and B. The increase in *kaiBC::luc* reporter expression remains robust when *cikA* is replaced by a full-length variant that cannot engage with the oscillator to turn on RpaA phosphatase activity.

All about RpaA. Because RpaA-dependent reporter gene expression is reduced in the *kaiA* deletion mutant and a high percentage of LDC-sensitive mutants are RpaA-regulated genes, we measured the levels of RpaA phosphorylation in cultures during LDCs. RpaA phosphorylation in WT cycles throughout the 24-h period (Figure 4.2-3A); higher resolution sampling shows that phosphorylation levels peak at the LD transition (Espinosa et al., 2015). In the *kaiA* mutant phosphorylated RpaA levels are constitutively low. This constant low level of phosphorylated RpaA is similar to an *rpaA* null mutant in that both

negatively affect gene expression that is vital to surviving the dark (Boyd et al. 2013). The inability to accumulate phosphorylated RpaA when KaiA is absent is relieved when *cikA* is also knocked out. These data support the hypothesis that the respective activities of KaiA and CikA balance clock output.

However, KaiC phosphorylation, which is stimulated by KaiA, never approaches levels comparable to the WT peak when KaiA is absent, regardless of the CikA status (Figure 4.2-3 B). This result can be explained by two distinct functions of KaiA in the clock: KaiA acts alone to stimulate KaiC phosphorylation at the KaiC C-terminal domain, and competes with CikA for binding to KaiB, which docks at the KaiC N-terminal domain. We proposed that the low RpaA-phosphate level, and LCD sensitivity, in the *kaiA* mutant strains resulted from lack of competition for CikA for binding to KaiB. Thus, we proposed that the LDC-sensitive phenotype results from out of control dephosphorylation of RpaA by CikA.

Severity of the *kaiA* and *rpaA* mutants in LDC. Other physiological similarities between the *kaiA* and the *rpaA* mutants exist. The *rpaA* mutant accumulates excessive reactive oxygen species during the day that it is unable to alleviate during the night. Metabolomic measurements show that knocking out RpaA causes a reductant deficit in the dark, and it is hypothesized that the lack of primary nighttime NADPH-producing reactions that are targets of RpaA prevent the cell from detoxifying this ROS. These physiological effects are also

seen with the *kaiA* mutant (Figure 4.2-3C), which has increased levels of ROS compared to that of the WT and the *kaiA cikA* double mutant.

Concluding Remarks. This is the first study to quantify the genome-wide genetic fitness contributions of genes in cyanobacteria under particular growth conditions. Good indication of the success of our screen was finding that genes whose mutants were previously known to be sensitive to LDCs such as those of the oxidative pentose phosphate pathway and glycogen breakdown show up in our analysis. We also identified previously unknown genes important for LDCs and identified an unexpected role for the highly studied *kaiA*.

Redox homeostasis is especially important during LDC conditions. ROS can be managed by the cell during the day via reductant (NADPH) generated from electron transport through the photosystems, but when cells enter the dark, this photosynthetically derived reductant ceases to be made and the primary source of NADPH is acquired via the oxidative pentose phosphate pathway (Guo et al. 2014). In the case of the *kaiA* mutant strains described in this study, we also observed increases in ROS and a failure to generate the signal to turn on the nighttime genetic program of the cell.

Methods.

Bacterial strains and culture conditions. All cultures were constructed using wild type *S. elongatus* PCC 7942 stored in our laboratory as AMC06. Cultures were grown at 30 C and when in BG-11 liquid or solid medium with

antibiotics as needed for selection at standard concentrations. Liquid cultures were either cultivated in 100ml volumes in 250ml flasks shaken at ~150rpm on an orbital shaker or in 400ml volume in top lit bioreactors (Photometrics) under 500 μ E square wave LD cycles. The bioreactors were mixed via filtered air bubble agitation from the bottom with a rate of 0.1ml per minute.

Bioluminescence Monitoring. Bioluminescence of *PkaiBC-luc* firefly luciferase fusion reporter strains was monitored at 30 C under LL conditions as described previously (Mackey et al. 2011). Data were analyzed with the Biological Rhythms Analysis Software System (<http://millar.bio.ed.ac.uk/pebrown/brass/brasspage.htm>) import and analysis program using Microsoft Excel. Results shown are representative of at least four independent experiments each averaging the reads from replicate wells.

Immunoblotting. SDS/PAGE was performed according to standard methods with the following exceptions. Phosphorylation of RpaA and KaiC was detected using 10% SDS-polyacrylamide gels supplemented with Phos-tag ligand (Wako Chemicals USA) at a final concentration of 25 μ M and manganese chloride at a final concentration of 50 μ M. Gels were incubated once for 10 min in transfer buffer supplemented with 100 mM EDTA, followed by a 10-min incubation in transfer buffer without EDTA before standard wet transfer. For phospho-RpaA detection, protein extracts and electrophoretic apparatus were kept chilled to minimize hydrolysis of heat-labile phospho-Asp. Protein extracts for use in Phos-tag gels were prepared in Tris-buffered saline, and extracts for

standard SDS/PAGE were prepared in PBS. RpaA anti-serum (a gift of E. O'Shea, Harvard University, Cambridge, MA) was used at a dilution of 1:1,000. KaiC immunoblotting was performed as described previously (Ivleva and Golden 2007) with modifications described elsewhere (Dong et al. 2016). Densitometric analyses were performed using National Institutes of Health ImageJ software (Schneider et al. 2012).

Gene fitness calculations. To estimate fitness the fitness effects of gene knockouts in light-dark conditions relative to continuous light, we developed an analysis pipeline that consisted of curating the data, normalizing it, and then analyzing it using linear models. We first counted the number of reads for each sample to use as a normalizing factor among samples. Barcodes are dispersed across the genome, and we removed any barcode falling outside a protein coding region (24868/154949) or within a gene but not within the middle 80% (27763/154949). Based on the barcodes remaining, we removed any gene not represented by at least three barcodes in different positions (114/2075). This cutoff yielded 102136 barcodes distributed across 1961 genes.

For each barcode in each sample we added a pseudocount of one to the number of reads, divided by the total number of reads for the sample as calculated before, and took the log-2 transformation of this sample-normalized number of reads. The experiment involved two different starting pools of strains (called T0), each of which was divided into LL and LD samples. To account for different starting percentages of each barcode within the T0 pools, we averaged

the log-2 transformed values for a barcode across the four replicate T0 samples for each pool then subtracted these average starting barcode values from the LL and LD values in the respective pools. We also removed any gene without at least 15 T0 reads (across the four replicates and before adding the pseudocount in each pool (89/1961), leaving 1872 genes and 101258 barcodes.

For each gene, we used maximum likelihood to fit a pair of nested linear mixed effects models to the sample- and read-normalized log-2 transformed counts:

$$(1) \quad y_{i,j,k} = \mu_g + C_j + B_i + \varepsilon_{i,j,k} ; B_i \sim iid N(0, \zeta_g^2) ; \varepsilon_{i,j,k} \sim iid N(0, \sigma_g^2)$$

$$(2) \quad y_{i,j,k} = \mu_g + B_i + \varepsilon_{i,j,k} ; B_i \sim iid N(0, \zeta_g^2) ; \varepsilon_{i,j,k} \sim iid N(0, \sigma_g^2)$$

where $y_{i,j,k}$ is the normalized log-2 value for barcode i in gene g in condition j for sample k , μ_g is the average value for the gene, C_j is the fixed effect of condition j , B_i is a random effect for barcode i , and $\varepsilon_{i,j,k}$ is the residual. We identified genes with significant fitness differences between conditions by comparing the difference in the $-2 \cdot \log$ likelihoods of the models to a chi-square distribution with one degree of freedom, estimating a p-value, accounting for multiple testing by the method of Benjamini and Hochberg (Benjamini and Hochberg 1995) and selecting those gene with adjusted p-values less than 0.01. We took the contrast $C_{LD} - C_{LL}$ to be the estimated fitness effect of knocking out the gene.

Quantification of ROS. Total cellular ROS load in WT and the mutant was quantified using a fluorescent compound H₂DCFDA (Life Technologies catalog no. D399)(Rastogi et al. 2010). Briefly, 2 mL of photobioreactor-grown culture

was collected and split into 1-mL aliquots. H₂DCFDA was added to one sample at a final concentration of 5 μM. Tubes were protected from light and shaken at 30 C for 30 min. After incubation, 200 μL from each tube was added to a separate well in a clear-bottom black 96-well plate. Fluorescence was quantified at an excitation of 480 nm and an emission of 520 nm on a Tecan Infinite M200 plate reader with the gain manually set to 120. Fluorescence data were normalized to OD₇₅₀ of each sample, and untreated-sample background fluorescence was then subtracted from treated sample fluorescence values.

Tables.

Table 4.2-1. Strains.

Strain	Genetic background	Reporter plasmid	Antibiotic resistance	Source or reference
AMC541	Wild type	<i>PkaiBC</i> :: <i>luc</i> ;	Cm	(Chen et al. 2009)
AMC1161	<i>kaiA</i> insertion; pAM2969 in AMC541	<i>PkaiBC</i> :: <i>luc</i> ;	Cm Km	
AMC702	<i>kaiA</i> deletion; in-frame markerless constructed in AMC541	<i>PkaiBC</i> :: <i>luc</i> ;	Cm	(Ditty et al. 2003)
AMC1936	<i>kaiABC</i> knockout; pAM4252 in AMC541	<i>PkaiBC</i> :: <i>luc</i> ;	Cm Km	(Paddock et al. 2013)
AMC1679	<i>kaiA cikA</i> double knockout; pAM2152 in AMC702	<i>PkaiBC</i> :: <i>luc</i> ;	Cm Gm	(Dong et al. 2010)
KaiACikA+	AMC1679 with <i>P_{trc}:cikA(C644R mutation)</i> in NS1	<i>PkaiBC</i> :: <i>luc</i> ;	Cm Gm	This study

Figures.

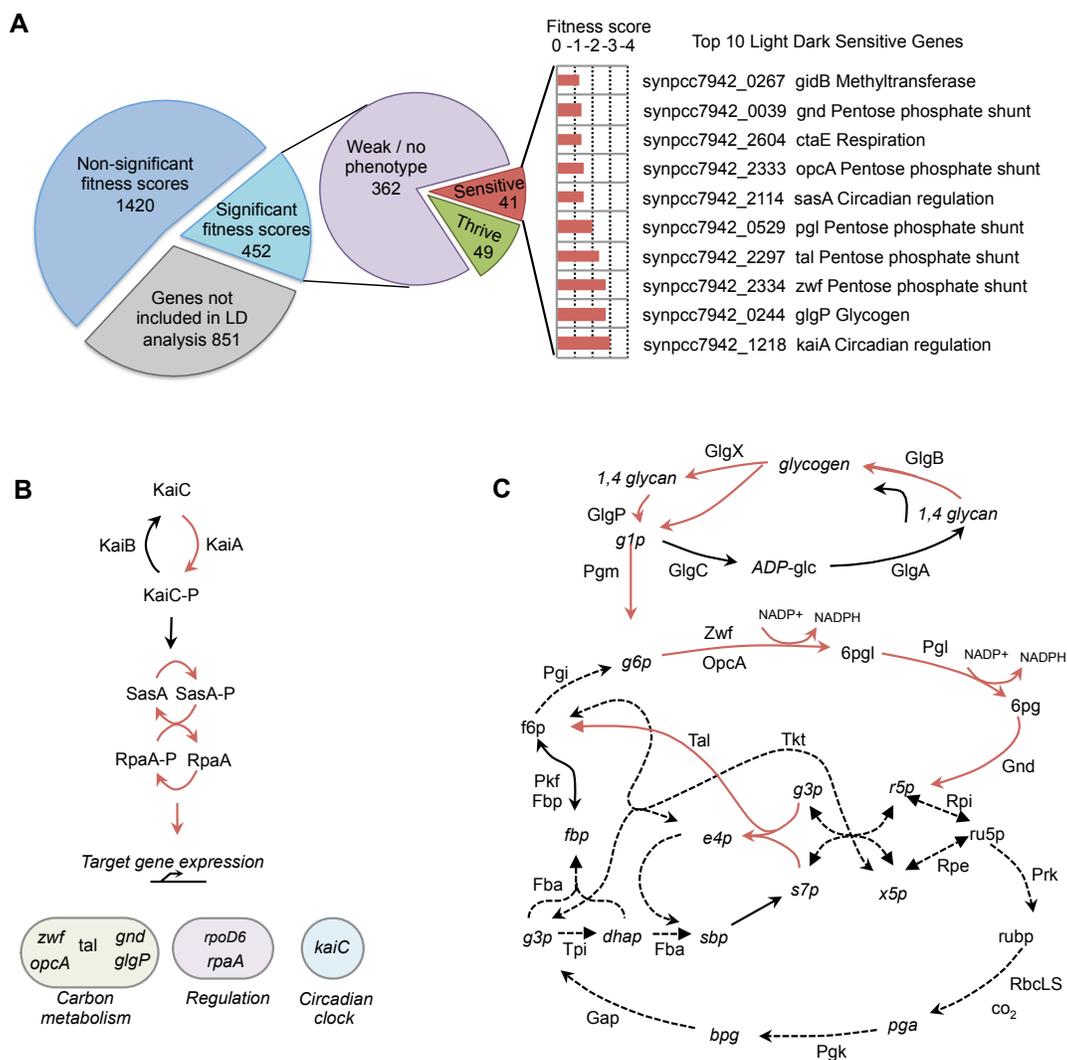


Figure 4.2-1: LDC sensitive genes (A) Overview of the number of genes and their respective LDC sensitivity screening results. The top LDC-sensitive gene scores, locus id, gene symbol, and functional category are shown in the bar graph. **(B)** KaiA disruption leads to a cascade of regulatory dysfunction resulting in adverse physiological consequences for carbon metabolism, transcriptional regulation, and of circadian clock component expression. The genes in the highlighted circles are known to be direct targets of clock output RpaA signaling. Not shown is the RpaA phosphatase CikA, whose activity requires interaction with KaiB-KaiC complex, where it binds in competition with KaiA. **(C)** Highlighted in the metabolic map are the main reactions that are essential for surviving LDCs. Red arrows represent reactions that are essential for growth under cycling light-dark conditions, solid black arrows represent reactions with observed phenotypic consequence when disrupted, and dashed arrows represent reactions that are essential under continuous light conditions (Rubin et. al 2016).

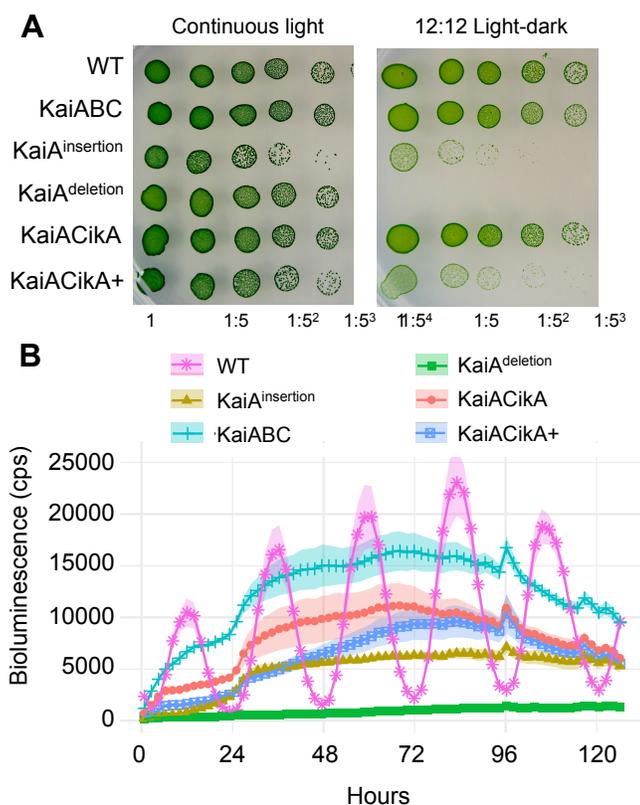


Figure 4.2-2: Growth and circadian rhythms of KaiA mutants. (A) Growth in LDC and continuous light on spot plates. (B) After entrainment to a 12-h light:12-h dark cycle, bioluminescence traces were captured over 5 d under continuous light. Inactivation of *kaiA* by deletion of the entire coding region (green) results in low arrhythmic expression from the *PkaiBC::luc* reporter, whereas inactivation by insertion of an antibiotic-resistance cassette results in arrhythmic rhythms of slightly higher levels. Inactivation of *cikA* in a deletion background increases the magnitude of the values, closer to levels observed in the *kaiABC* mutant (orange). This increase in *kaiBC::luc* reporter expression is evident even when *cikA* is intact but replaced by a variant that cannot engage the circadian oscillator to turn on RpaA-phosphatase activity (blue). Representative traces are shown. cps, counts per second.

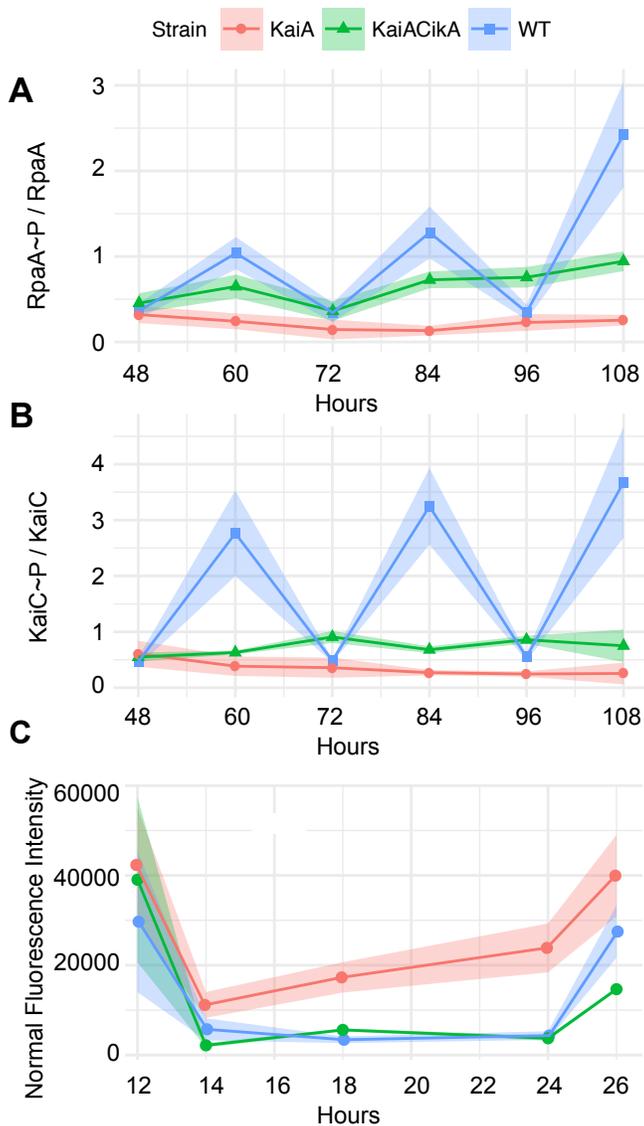


Figure 4.2-3: Phosphorylation time-course. Densitometric analysis showing average \pm SE values from multiple immunoblots showing the *in vivo* RpaA (A) and KaiC (B) phosphorylation cycles, demonstrating overall reduced levels of RpaA and KaiC phosphorylation in the KaiA and *kaiAcikA* mutants compared to WT entrained to a 12:12 LDC and sampled every 12 h. The RpaA phosphorylation level in the *kaiA* mutants is arrhythmic and lower than in WT, with the *kaiAcikA* double mutant accumulating significantly higher phosphorylated RpaA throughout the sampling period. A plot of H₂DCFDA fluorescence over a 12 h dark period of an LDC indicating total cellular ROS in WT, *kaiA*, and *kaiAcikA*. The shaded area in all graphs indicates the standard error.

References.

- Bell-Pedersen D, Cassone VM, Earnest DJ, Golden SS, Hardin PE, Thomas TL, Zoran MJ (2005) Circadian rhythms from multiple oscillators: lessons from diverse organisms. *Nature Reviews Genetics* 6 (7):544-556. doi:10.1038/nrg1633
- Benjamini Y, Hochberg Y (1995) Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *J Roy Stat Soc B Met* 57 (1):289-300
- Boyd JS, Bordowitz JR, Bree AC, Golden SS (2013) An allele of the *crm* gene blocks cyanobacterial circadian rhythms. *Proceedings of the National Academy of Sciences of the United States of America* 110 (34):13950-13955. doi:10.1073/pnas.1312793110
- Chang YG, Cohen SE, Phong C, Myers WK, Kim YI, Tseng R, Lin J, Zhang L, Boyd JS, Lee Y, Kang S, Lee D, Li S, Britt RD, Rust MJ, Golden SS, LiWang A (2015) Circadian rhythms. A protein fold switch joins the circadian oscillator to clock output in cyanobacteria. *Science* 349 (6245):324-328. doi:10.1126/science.1260031
- Chen Y, Kim YI, Mackey SR, Holtman CK, Liwang A, Golden SS (2009) A novel allele of *kaiA* shortens the circadian period and strengthens interaction of oscillator components in the cyanobacterium *Synechococcus elongatus* PCC 7942. *Journal of Bacteriology* 191 (13):4392-4400. doi:10.1128/JB.00334-09
- Cohen SE, Golden SS (2015) Circadian Rhythms in Cyanobacteria. *Microbiology and Molecular Biology Reviews* : MMBR 79 (4):373-385. doi:10.1128/MMBR.00036-15
- Diamond S, Jun D, Rubin BE, Golden SS (2015) The circadian oscillator in *Synechococcus elongatus* controls metabolite partitioning during diurnal growth. *Proceedings of the National Academy of Sciences of the United States of America* 112 (15):E1916-1925. doi:10.1073/pnas.1504576112
- 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. *Proceedings of the National Academy of Sciences of the United States of America* 114 (4):E580-E589. doi:10.1073/pnas.1613078114

- Ditty, J.L., Canales, S.R., Anderson, B.E., Williams, S.B., Golden, S.S., 2005. Stability of the *Synechococcus elongatus* PCC 7942 circadian clock under directed anti-phase expression of the kai genes. *Microbiology* 151, 2605–2613.
- Ditty JL, Williams SB, Golden SS (2003) A cyanobacterial circadian timing mechanism. *Annu Rev Genet* 37:513-543. doi:10.1146/annurev.genet.37.110801.142716
- Dodd AN, Salathia N, Hall A, Kevei E, Toth R, Nagy F, Hibberd JM, Millar AJ, Webb AA (2005) Plant circadian clocks increase photosynthesis, growth, survival, and competitive advantage. *Science* 309 (5734):630-633. doi:10.1126/science.1115581
- Dong G, Yang Q, Wang Q, Kim YI, Wood TL, Osteryoung KW, van Oudenaarden A, Golden SS (2010) Elevated ATPase activity of KaiC applies a circadian checkpoint on cell division in *Synechococcus elongatus*. *Cell* 140 (4):529-539. doi:10.1016/j.cell.2009.12.042
- Dong P, Fan Y, Sun J, Lv M, Yi M, Tan X, Liu S (2016) A dynamic interaction process between KaiA and KaiC is critical to the cyanobacterial circadian oscillator. *Scientific Reports* 6:25129. doi:10.1038/srep25129
- Dunlap JC, DeCoursey PJ, Loros JJ (2003) *Chronobiology: Biological Timekeeping*. Sinauer Associates, Incorporated,
- Espinosa, J., Boyd, J.S., Cantos, R., Salinas, P., Golden, S.S., Contreras, A., 2015. Cross-talk and regulatory interactions between the essential response regulator RpaB and cyanobacterial circadian clock output. *Proc Natl Acad Sci USA* 112, 2198–2203.
- Guo J, Nguyen AY, Dai Z, Su D, Gaffrey MJ, Moore RJ, Jacobs JM, Monroe ME, Smith RD, Koppelaar DW, Pakrasi HB, Qian WJ (2014) Proteome-wide light/dark modulation of thiol oxidation in cyanobacteria revealed by quantitative site-specific redox proteomics. *Molecular & Cellular Proteomics* : MCP 13 (12):3270-3285. doi:10.1074/mcp.M114.041160
- Gutu A, O'Shea EK (2013) Two antagonistic clock-regulated histidine kinases time the activation of circadian gene expression. *Mol Cell* 50 (2):288-294. doi:10.1016/j.molcel.2013.02.022
- Ishiura M, Kutsuna S, Aoki S, Iwasaki H, Andersson CR, Tanabe A, Golden SS, Johnson CH, Kondo T (1998) Expression of a gene cluster kaiABC as a circadian feedback process in cyanobacteria. *Science* 281 (5382):1519-1523. doi:Doi 10.1126/Science.281.5382.1519

- Ito H, Mutsuda M, Murayama Y, Tomita J, Hosokawa N, Terauchi K, Sugita C, Sugita M, Kondo T, Iwasaki H (2009) Cyanobacterial daily life with Kai-based circadian and diurnal genome-wide transcriptional control in *Synechococcus elongatus*. *Proceedings of the National Academy of Sciences of the United States of America* 106 (33):14168-14173. doi:10.1073/pnas.0902587106
- Ivleva NB, Golden SS (2007) Protein extraction, fractionation, and purification from cyanobacteria. *Methods in molecular biology* 362:365-373. doi:10.1007/978-1-59745-257-1_26
- Iwasaki H, Williams SB, Kitayama Y, Ishiura M, Golden SS, Kondo T (2000) A kaiC-interacting sensory histidine kinase, SasA, necessary to sustain robust circadian oscillation in cyanobacteria. *Cell* 101 (2):223-233. doi:10.1016/S0092-8674(00)80832-6
- Kelleher FC, Rao A, Maguire A (2014) Circadian molecular clocks and cancer. *Cancer Lett* 342 (1):9-18. doi:10.1016/j.canlet.2013.09.040
- LeGates TA, Fernandez DC, Hattar S (2014) Light as a central modulator of circadian rhythms, sleep and affect. *Nat Rev Neurosci* 15 (7):443-454. doi:10.1038/nrn3743
- Mackey SR, Golden SS, Ditty JL (2011) The itty-bitty time machine genetics of the cyanobacterial circadian clock. *Adv Genet* 74:13-53. doi:10.1016/B978-0-12-387690-4.00002-7
- Markson JS, Piechura JR, Puszynska AM, O'Shea EK (2013) Circadian control of global gene expression by the cyanobacterial master regulator RpaA. *Cell* 155 (6):1396-1408. doi:10.1016/j.cell.2013.11.005
- McCarthy MJ, Welsh DK (2012) Cellular Circadian Clocks in Mood Disorders. *Journal of biological rhythms* 27 (5):339-352. doi:10.1177/0748730412456367
- Morris CJ, Yang JN, Scheer FA (2012) The impact of the circadian timing system on cardiovascular and metabolic function. *Prog Brain Res* 199:337-358. doi:10.1016/B978-0-444-59427-3.00019-8
- Nishiwaki T, Satomi Y, Nakajima M, Lee C, Kiyohara R, Kageyama H, Kitayama Y, Temamoto M, Yamaguchi A, Hijikata A, Go M, Iwasaki H, Takao T, Kondo T (2004) Role of KaiC phosphorylation in the circadian clock system of *Synechococcus elongatus* PCC 7942. *Proceedings of the National Academy of Sciences of the United States of America* 101 (38):13927-13932. doi:10.1073/pnas.0403906101

- Ouyang Y, Andersson CR, Kondo T, Golden SS, Johnson CH (1998) Resonating circadian clocks enhance fitness in cyanobacteria. *Proceedings of the National Academy of Sciences of the United States of America* 95 (15):8660-8664
- Paddock ML, Boyd JS, Adin DM, Golden SS (2013) Active output state of the *Synechococcus* Kai circadian oscillator. *Proceedings of the National Academy of Sciences of the United States of America* 110 (40):E3849-3857. doi:10.1073/pnas.1315170110
- Pattanayak GK, Phong C, Rust MJ (2014) Rhythms in energy storage control the ability of the cyanobacterial circadian clock to reset. *Current biology : CB* 24 (16):1934-1938. doi:10.1016/j.cub.2014.07.022
- Rastogi RP, Singh SP, Häder D-P, Sinha RP (2010) Detection of reactive oxygen species (ROS) by the oxidant-sensing probe 2',7'-dichlorodihydrofluorescein diacetate in the cyanobacterium *Anabaena variabilis* PCC 7937. *Biochemical and Biophysical Research Communications* 397 (3):603-607. doi:10.1016/j.bbrc.2010.06.006
- 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. *Proceedings of the National Academy of Sciences of the United States of America*. doi:10.1073/pnas.1519220112
- Rust MJ, Golden SS, O'Shea EK (2011) Light-driven changes in energy metabolism directly entrain the cyanobacterial circadian oscillator. *Science* 331 (6014):220-223. doi:10.1126/science.1197243
- Schneider CA, Rasband WS, Eliceiri KW (2012) NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* 9 (7):671-675
- Takai N, Nakajima M, Oyama T, Kito R, Sugita C, Sugita M, Kondo T, Iwasaki H (2006) A KaiC-associating SasA-RpaA two-component regulatory system as a major circadian timing mediator in cyanobacteria. *Proceedings of the National Academy of Sciences of the United States of America* 103 (32):12109-12114. doi:10.1073/pnas.0602955103
- Tseng R, Chang YG, Bravo I, Latham R, Chaudhary A, Kuo NW, LiWang A (2014) Cooperative KaiA-KaiB-KaiC interactions affect KaiB/SasA competition in the circadian clock of cyanobacteria. *Journal of molecular biology* 426 (2):389-402. doi:10.1016/j.jmb.2013.09.040
- Tseng R, Goularte NF, Chavan A, Luu J, Cohen SE, Chang Y-G, Heisler J, Li S, Michael AK, Tripathi S, Golden SS, LiWang A, Partch CL (2017) Structural

basis of the day-night transition in a bacterial circadian clock. *Science* 355 (6330):1174-1180. doi:10.1126/science.aag2516

Williams SB, Vakonakis I, Golden SS, LiWang AC (2002) Structure and function from the circadian clock protein KaiA of *Synechococcus elongatus*: a potential clock input mechanism. *Proceedings of the National Academy of Sciences of the United States of America* 99 (24):15357-15362. doi:10.1073/pnas.232517099

Woelfle MA, Ouyang Y, Phanvijhitsiri K, Johnson CH (2004) The adaptive value of circadian clocks: an experimental assessment in cyanobacteria. *Current Biology* : CB 14 (16):1481-1486. doi:10.1016/j.cub.2004.08.023

Xu Y, Mori T, Pattanayek R, Pattanayek S, Egli M, Johnson CH (2004) Identification of key phosphorylation sites in the circadian clock protein KaiC by crystallographic and mutagenetic analyses. *Proceedings of the National Academy of Sciences of the United States of America* 101 (38):13933-13938. doi:10.1073/pnas.0404768101

Yerushalmi S, Yakir E, Green RM (2011) Circadian clocks and adaptation in *Arabidopsis*. *Mol Ecol* 20 (6):1155-1165. doi:10.1111/j.1365-294X.2010.04962.x

4.3 The Role of c-di-AMP in Light-Dark Cycles

Abstract. The broadly conserved signaling nucleotide c-di-AMP is essential for viability in most bacteria where it has been studied. However, characterization of c-di-AMP has largely been confined to a few genera of pathogenic organisms, limiting our functional understanding of the molecule among diverse phyla. Here we identify the cyclase responsible for c-di-AMP synthesis and characterize the molecule's role in survival of darkness in the model photosynthetic cyanobacterium, *Synechococcus elongatus* PCC 7942. In addition to using traditional genetic, biochemical, and proteomic approaches, we developed a high-throughput genetic interaction screen (IRB-Seq) to illuminate the pathways where the signaling nucleotide is active in *S. elongatus*. We found that the c-di-AMP cyclase is encoded by a previously unannotated gene, named here as *dacA*. In the loss-of-function mutant of *dacA* the cell experiences increased oxidative stress and death in the night portion of day-night cycles, in which potassium transport is implicated. These findings suggest that c-di-AMP is biologically active in Cyanobacteria, where the molecule's non-canonical role in oxidative stress management and day-night survival is potentially managed by traditional functions such as potassium homeostasis. The pipeline and analysis tools for IRB-Seq developed for these studies constitute a quantitative high-throughput approach for studying genetic interactions in microorganisms.

Introduction. The signaling nucleotide cyclic di-adenosine monophosphate (c-di-AMP) has been implicated in a wide range of biological

processes since its discovery less than a decade ago (Witte et al., 2008). The molecule is active in multiple pathways including potassium transport, regulation of central metabolism, cell wall homeostasis, gene expression, and DNA damage responses (Corrigan and Gründling, 2013). In addition, c-di-AMP is the only second messenger that is essential to many of the organisms in which it is studied (Commichau et al., 2015). Despite the biological importance of c-di-AMP, *in-vivo* studies of this signaling nucleotide have been focused on a narrow array of pathogenic Firmicutes and to a lesser extent Actinobacteria. Expanding the investigation of c-di-AMP to a broader diversity of microbial phyla is likely to reveal new functions for this molecule.

Organisms in the photosynthetic phylum, Cyanobacteria, are key players in global carbon, oxygen, and nitrogen cycling, and are promising platforms for renewable production of industrial chemicals (Angermayr et al., 2015; Bryant, 2003; Flombaum and Gallegos, 2013). Signaling nucleotides aside from c-di-AMP (cGMP, c-di-GMP, (p)ppGpp, and cAMP) serve important roles in their environmental responses. Studying these messengers in photosynthetic organisms elucidated new roles for the nucleotides, such as regulating phototaxis, photosystem repair, and dark survival (Agostoni and Montgomery, 2014; Hood et al., 2016). While c-di-AMP has not been reported in the phylum, the existence of enzymes and riboswitches that interact with the molecule have been suggested based on homology (Agostoni and Montgomery, 2014; Corrigan and Gründling, 2013; Huynh et al., 2015; Nelson et al., 2013). Discovery of c-di-AMP and its role

in Cyanobacteria would lead to better understanding of these important organisms and the function of c-di-AMP.

Genetic interaction screens in other organisms have offered important insights into c-di-AMP function. Screens focused on secondary mutations that relieve the phenotypes of c-di-AMP-related mutants (alleviating interactions) have successfully identified c-di-AMP interacting proteins (Corrigan et al., 2011; Whiteley et al., 2015). However, these screens have been limited to assaying positive interactions, are non-quantitative, and have been performed only under a single condition.

Next-generation sequencing paired with transposon mutagenesis (Tn-Seq) is a recently described method that can be used for quantitative genetic interactions screens to identify alleviating as well as aggravating interactions (Dejesus et al., 2017; Meeske et al., 2015; van Opijnen et al., 2009). This approach relies upon the generation of a new mutant library in the background of a knockout of interest for each screen (Brochado and Typas, 2013). While informative, the need to generate and characterize a new transposon mutant library for each interaction screen is both costly and labor intensive (Gray et al., 2015). These features have limited the applicability of Tn-Seq for interaction screens. A similar method, RB-TnSeq, provides a less costly and laborious approach for screening transposon mutant libraries, in which mutants are tracked by the sequencing of 20 base pair “barcodes” on each transposon (Wetmore et al., 2015). This approach has not yet been applied to interaction screens, but

offers great potential for mitigating their logistical weaknesses and enlarging their scope.

Here, we used genetic, biochemical, and proteomic approaches to elucidate the presence and roles of c-di-AMP in the cyanobacterium *Synechococcus elongatus* PCC 7942. To further probe the molecule's functions we developed and implemented an approach we named interaction-based RB-TnSeq, or IRB-Seq, to quantitatively measure genetic interactions with the c-di-AMP cyclase under multiple conditions.

Results.

Presence of c-di-AMP and its Cyclase. The putative c-di-AMP producing enzyme in *S. elongatus* was identified computationally. The software package HMMER (Johnson et al., 2010) identified a single protein in the organism that contains a DisA_N domain (PF02457), currently the only domain known to be responsible for c-di-AMP production (Xayarath and Freitag, 2015). This sequence, annotated at time of publication as “protein of unknown function DUF147”, predicts three transmembrane segments and a cytoplasmic DisA_N domain (Fig. 4.3-1A). The arrangement of domains identifies the protein as a member of the most abundant family of DAC domain-containing proteins, DacA (Corrigan et al., 2013). Thus we refer to this previously unannotated gene, which encodes the putative cyclase of c-di-AMP, as *dacA* (Synpcc7942_0263).

To validate the bioinformatics predication, we quantified c-di-AMP levels in wild type (WT) and a *dacA* mutant using LC-MS. The WT value of 18.8 μM is

several fold higher than in other organisms, including *Staphylococcus aureus* (Corrigan et al., 2015) and *Bacillus subtilis* (Oppenheimer-Shaanan et al., 2011), where the nucleotide plays biologically important roles. In contrast to many other c-di-AMP-producing organisms, in *S. elongatus* a fully segregated insertion mutant for *dacA* is viable under standard lab conditions of continuous light. In this mutant, the level of c-di-AMP was 3.3 μM (Fig. 4.3-1B) ($P < 10^{-7}$; Mann-Whitney-Wilcoxon Test). Difficulties in c-di-AMP extraction from *S. elongatus* samples caused substantial background noise in mass spectrometry quantification, which likely accounted for the detection of low-level c-di-AMP in the *dacA* mutant. Regardless, these data suggest that *S. elongatus* produces c-di-AMP, and that DacA catalyzes its synthesis.

Concentrations of some other signaling nucleotides studied in Cyanobacteria change in response to light (Agostoni and Montgomery, 2014). To address whether this may be the case with c-di-AMP in *S. elongatus*, we sampled over one 12 h:12 h light-dark cycle (LDC). The c-di-AMP concentration was variable in WT both within time points and between them, with an apparent, although non-significant, trend upwards at nighttime. Given that signaling nucleotide responses to light have been shown to occur within minutes in Cyanobacteria (Terauchi and Ohmori, 2004), we did higher resolution sampling around the light-to-dark and dark-to-light transitions. While light did not have a large or immediate effect on c-di-AMP, the onset of darkness caused a spike in the nucleotide's levels observable after 15 minutes and reaching a three-fold increase (Fig. 4.3-1C).

LDC Sensitivity in *dacA* Mutant. Based on the observed increase in c-di-AMP levels upon onset of darkness as well as previous research showing the importance of ppGpp, a closely linked signaling nucleotide, to dark survival in *S. elongatus* (Hood et al., 2016), we examined whether c-di-AMP is necessary to survive LDCs. On solid media the *dacA* loss of function mutant showed decreased growth in LDCs, but not in constant light (Fig. 4.3-2A), and this sensitivity to LDCs was reproducible in liquid culture (Fig. 4.3-2B). To confirm that the *dacA* mutation is responsible for this phenotype we added an ectopic copy of the *dacA* gene to the mutant, which complemented the LDC sensitivity of the mutant. Therefore, the *dacA* mutation and its decreased c-di-AMP levels are likely responsible for the LDC-specific defect.

To explore the nature and cause of the LDC sensitivity we examined cells over one LDC. While the *dacA* mutant grew similarly to WT over the light portion of the LDC, viability of the mutant, as measured by colony forming units in outgrowths, rapidly decreased upon the onset of darkness (Fig. 4.3-2C). This difference was significant after two hours of exposure to darkness ($P < .05$; *t* test), and became more pronounced over the course of the night (Fig 4.3-2C). Recent work showed that reactive oxygen species (ROS) correlate with death in darkness for this species (Diamond et al., 2017). Indeed, reactive oxygen species peaked in the mutant upon the onset of darkness at more than two fold the level in WT and remained higher through the course of the night ($P < 10^{-5}$; Mann-Whitney-Wilcoxon Test) (Fig. 4.3-2D). These data suggest an active death

mechanism in the *dacA* mutant, which occurs specifically in the dark stage of the LDC as a consequence of high oxidative stress.

Biochemical Interactions of c-di-AMP. Because no *in vivo* evidence exists for members of the c-di-AMP signaling pathway in Cyanobacteria we used an unbiased proteomic approach to identify binding partners. C-di-AMP-bound beads were used for affinity purification of interacting *S. elongatus* proteins (Sureka et al., 2014). These proteins were in turn identified by gel-free quantitative shotgun proteomics (Fig. 4.3-3A). There were eleven proteins enriched at least four fold over those from control beads, of which five had domains previously shown to bind c-di-AMP (Table 4.3-1). In addition to known c-di-AMP binding domains, the majority of the proteins have predicted functions previously associated with the signaling nucleotide, such as ion and particularly potassium transport (Synpcc7942_1729, Synpcc7942_1588, and Synpcc7942_0545), and DNA replication, repair, and homologous recombination (Synpcc7942_1416, Synpcc7942_1886, and Synpcc7942_0301), (Commichau et al., 2015; Corrigan and Gründling, 2013). The identification of known c-di-AMP binding domains and functional pathways among the pull-down candidates validate the approach.

To follow up on these results we conducted both targeted mutagenesis and individual binding assays for the top candidates. Of the eleven proteins enriched two fold or more, two had previously been shown to be essential (Synpcc7942_1416 and Synpcc7942_0301) (Rubin et al., 2015). We successfully made insertion mutants for all of the remaining genes, with the exception of

Synpcc7942_0420, and characterized their phenotypes. None of the mutants, however, phenocopied the *dacA* mutant's LDC sensitivity, or showed visible phenotypes under control conditions. In parallel we expressed nine of the eleven candidate proteins in *Escherichia coli* and tested them individually for binding to c-di-AMP using DRaCALA, a method based upon differential diffusion of bound and unbound radiolabelled c-di-AMP across a nitrocellulose membrane (Roelofs et al., 2011). Three of the expressed proteins showed binding to the nucleotide (Fig. 4.3-3B). Of those candidates whose binding was confirmed, KdpD encoded by *synpcc7942_1729*, has been associated with c-di-AMP previously and suggests a role for the molecule in potassium transport in *S. elongatus*. The PII-like protein encoded by *synpcc7942_1476* consists of a domain previously associated with the poorly understood PstA protein, which has been studied for its c-di-AMP binding in *Bacillus subtilis*, *Staphylococcus aureus*, and *Listeria monocytogenes* (Campeotto et al., 2015; Choi et al., 2015; Gundlach et al., 2015; Müller et al., 2015). Surprisingly, while it is highly conserved among those organisms, the protein sequence is divergent in *S. elongatus* (E-Value=.2 compared to the *L. monocytogenes* gene). This finding suggests that although c-di-AMP may interact with a functionally similar protein in *S. elongatus*, the binding pocket is disparate. The protein TrmH (Synpcc 7942_1874), which stabilizes tRNAs by addition of a methyl group to a conserved guanosine (Nakanishi and Nureki, 2005), has neither a function nor domain previously associated with c-di-AMP. Together, the pulldown results and confirmation by

DRaCALA suggest conserved function and binding of c-di-AMP, as well as domains and genes not previously connected with c-di-AMP, such as TrmH.

IRB-Seq design. Although the pull-down was able to inform us of binding partners of c-di-AMP, it did not reveal which of these interactions represent biologically meaningful roles. To address this issue we developed a genome-wide method for identifying genetic interactions of c-di-AMP's cyclase, DacA. The approach relied on a previously developed dense transposon insertion mutant library in *S. elongatus* (Rubin et al., 2015). Briefly, this library was built with the RB-TnSeq method (Wetmore et al., 2015), in which every loss-of-function mutant in the library contains a unique identifier sequence, or barcode (Fig. 4.3-4A). By using next-generation sequencing of barcodes the survival of the approximately 150,000 barcoded mutants in the library can be tracked under control as well as experimental conditions. In this way the *S. elongatus* RB-TnSeq library can be used for pooled, quantitative, whole genome mutant screens.

In interaction RB-TnSeq (IRB-Seq) a second mutation is added to all the library members in order to determine the fitness impact of the second mutation in combination with every other mutation in the library. In this case the second mutation inactivated *dacA*, and conferred a different antibiotic resistance than that used to construct the mutant library (see Materials and Methods). To a separate aliquot of the initial library, a non-deleterious mutation was added to serve as a control for potential changes to the library caused by the addition of a second mutation, or selection for the second antibiotic, that were not specific to the *dacA* mutation. The two double-mutant libraries were then grown under dual

selection so that every member of the library also contained the secondary mutation of interest (*dacA* or the control). At this point the barcodes were sequenced and compared to starting composition of the library and the control (Fig. 4.3-4B). These data were used to determine genetic interactions, or instances in which the fitness values of the library mutation and the *dacA* mutation were not simply additive. These interactions could in turn be used to identify relationships between genes and their pathways. In addition, because of our knowledge of the *dacA* mutant's decreased viability in LDCs we were able to sensitize our screen by exposing the library of double mutants to LDCs (Fig 4.3-4C). Frequencies of the barcodes after this stress were compared to a time zero sample and the control in order to determine genetic interactions apparent under stress conditions. In this way IRB-Seq enabled a quantitative assay of genetic interactions with *dacA*, under both control conditions as well as sensitizing conditions.

Before attempting IRB-Seq we ensured that the double-mutant library retained sufficient diversity for meaningful functional screens. Of the original 154,949 barcoded mutants in the library after addition of the second mutation we were able to recover, on average, 100,180 (65%). This number still represents approximately 30 insertion mutants for the average gene. To further test efficacy of the double-mutant library for screens we used the version containing the control mutation to reproduce the results of a previous screen for LDC-sensitive genes (Welkie et al., n.d.). Even though the new screen was conducted under different light conditions than the previous screen, and in flasks instead of

bioreactors, the two screens strongly correlate (Fig. 4.3-5A). It is of note that the slope of this correlation is approximately .4, suggesting that, although the relative phenotypes were similar, the mutants did not diverge as much over the course of the double-mutant experiment. Based on these results we carried out IRB-Seq using a *dacA* mutation layered on to the RB-TnSeq library for *S. elongatus*.

IRB-Seq interaction screen. The first genetic interactions examined were those that are detectable under standard laboratory conditions upon the addition of the *dacA* mutation to the library (Fig. 4.3-4B). As expected, given that the *dacA* mutation shows no fitness defects under control conditions (Fig. 4.3-2A), all of the strong genetic interactions (FDR < .01, genetic interaction absolute value > 1) were negative (aggravating or synthetic interactions) (Fig 4.3-5A, Dataset S1). These synthetic interactions can result from redundancy in function between the two mutated genes. Two of the top annotated synthetic interactions with *dacA* were with genes that encode flavoproteins Flv1 (synpcc7942_1810) and Flv3 (Synpcc7942_1809), which form a heterodimer that allows the release of high-energy electrons from photosystem II to oxygen without producing oxidative stress. Although these proteins are dispensable under standard laboratory conditions, under variable conditions such as fluctuating light, they become important for decreasing oxidative stress and allowing growth (Allahverdiyeva et al., 2013; Shaku et al., 2015). Also in the top ten annotated synthetic interactions were genes that encode a 6-pyruvoyl-tetrahydropterin synthase-like protein potentially involved in a pathway important for resisting UV-A stress (Synpcc7942_1184) (Matsunaga et al., 1993; Moon et al., 2012; Wachi et al.,

1995), a glutaredoxin-related protein (Synpcc7942_1145), and Psb28 (Synpcc7942_1679), a protein involved in repair of photosystem II in related cyanobacteria (Sakata et al., 2013). These interactions suggest that, in the absence of the functional *dacA*, a cell under increased oxidative stress (Fig 4.3-2D) becomes more dependent on proteins that offer electron sinks or protection against reactive oxygen species.

Another synthetic interaction apparent from the screen under standard laboratory conditions is with potassium transport. The fifth ranked candidate is the K⁺ transport protein, TrkA, encoded by *synpcc7942_1080*. Its genomic neighbor *synpcc7942_1081* and second ranked for interaction, is unannotated, but based on its TrkH domain (PF02386) and sequence homology (Finn et al., 2016), is likely TrkH, a complex partner of TrkA (Cao et al., 2013). Therefore, the *dacA* mutant is also sensitized to defects in potassium import.

IRB-Seq sensitized interaction screen. When the *dacA* double mutant library was sampled under sensitizing LDCs (Fig. 4.3-4C), we observed a number of positive (alleviating) interactions (Fig. 4.3-5C, Dataset S2). Alleviating interactions can occur when the affected genes are in the same pathway and their effects mask each other, or when the mutations counteract each other. Synthetic interactions were also present in the sensitized screen, identifying genes whose loss exacerbates the moderate survival impact of the *dacA* mutation in LDCs. For the sensitized interaction screen, we decreased the interaction score for consideration to greater than absolute value of .5. This change was based on our observation that the LDC screening conditions used enabled the differentiation of

genuine, but subtle, mutant phenotypes (Fig. 4.3-5A). Among the six top alleviating candidates for *dacA* in LDCs (FDR < .01, genetic interaction > .5) four are proteins involved in peptide and macromolecule degradation. The cell wall recycling protein MurQ (Synpcc7942_2577) as the top hit among alleviating interactions may be explained by decreased cell wall turnover being permissive in the *dacA* mutant, which has been associated with disruption of cell wall homeostasis in multiple species (Corrigan and Gründling, 2013). The reason for the more general enrichment of genes that encode degrading enzymes among alleviating mutations is unclear.

The fifth ranked hit among suppressors and the strongest hit among the synthetic interactions were circadian clock component CikA (Synpcc7942_0644) and circadian-involved sigma factor RpoD2 (Synpcc7942_1746), respectively. Mutants of *cikA*, the phosphatase for the master clock transcription factor RpaA, leave RpaA in a highly phosphorylated state (Gutu and O'Shea, 2013). This RpaA state locks the clock into activating nighttime processes and the detoxification of oxidative stress (Diamond et al., 2017; 2015; Markson et al., 2013). The single *cikA* mutant has a slightly positive effect on light-dark survival and has been shown capable of alleviating another LDC sensitive mutant, KaiA (Welkie et al., n.d.). Therefore, it might be expected that the *dacA* mutant, which is unable to fully clear oxidative stress and survive the night, would be suppressed by *cikA*. The *rpoD2* gene, which has by far the strongest synthetic interaction with *dacA* in the sensitized screen, encodes a sigma factor that causes changes in circadian rhythms when mutated (Tsinoremas et al., 1996).

However, the broader transcriptional and physiological effects of this mutation are unknown. Based on these candidates, it is likely that adjustments of the circadian clock can play a role in decreasing or improving the ability of the *dacA* mutant to survive LDCs, potentially by modulating the oxidative stress to which the mutant is exposed.

Discussion.

The Non-Essential Nature of c-di-AMP in S. elongatus. These investigations establish the presence of c-di-AMP at physiologically relevant levels in a cyanobacterium, and implicate the signaling nucleotide in response to light-dark transitions (Fig. 4.3-2). C-di-AMP controls many aspects of bacterial physiology and is essential in the bacterial phylum Firmicutes, where most research on the molecule has taken place. The essentiality of c-di-AMP may be related to its role in central metabolism. For instance, in *L. monocytogenes* the central metabolic enzyme pyruvate carboxylase is regulated by c-di-AMP (Sureka et al., 2014). More, recently it has been proposed that the essential nature of c-di-AMP can be explained by high levels of another signaling nucleotide, (p)ppGpp, that accumulates in the absence of c-di-AMP, which in turn leads to a starvation response (Whiteley et al., 2015). Importantly, this lethality in the absence of c-di-AMP occurs in rich media, while minimal media are permissive for the mutant. A possible explanation for *S. elongatus* viability with the *dacA* mutation is that, as a photoautotroph, it grows independent of carbon sources in the media. Regardless of its cause, the essential nature of c-di-AMP

in many organisms has limited the toolset available for research on the nucleotide (Xayarath and Freitag, 2015). Future genetic exploration of *dacA* in *S. elongatus* will be facilitated by the viability of the *dacA* mutant.

The IRB-Seq Approach to Genetic Interaction Screens. The IRB-Seq approach developed for this study enables high-throughput quantitative genetic interaction screens with minimal sequencing prep. A second mutation is added directly to an existing RB-TnSeq library, removing the need of previous approaches to recreate a library in a new background and determine all of the insertion loci for each screen (Fig. 4.3-4A)(Dejesus et al., 2017; Meeske et al., 2015; van Opijnen et al., 2009). Traditional sequencing preparation used in classical Tn-Seq studies is also avoided because survival of mutants is quantified by PCR and sequencing of 20bp “barcodes” present in each transposon which serve as identifiers for each clone in the mutant library (BarSeq) (Wetmore et al., 2015). IRB-Seq requires only one PCR and $\sim 1/100^{\text{th}}$ of an Illumina HiSeq 4000 lane per sample to provide a genome-wide quantitative measure of genetic interactions ranging from strong alleviation to full synthetic lethal. The advantage of a quantitative alternative to previous suppressor screens used in c-di-AMP research (Corrigan et al., 2011; Whiteley et al., 2015) is apparent in this study because the lack of fitness phenotype of the *dacA* mutant under constant light, and the moderate phenotype under the sensitizing LDCs, would have made traditional suppressor screens difficult, if not impossible. In addition, the high-throughput nature of this approach to genome wide genetic interaction screens

makes it feasible to conduct IRB-Seq screens in replicate and under many different sensitizing and permissive conditions.

Illuminating the Role of c-di-AMP in Nighttime Survival. Elucidating the survival of Cyanobacteria in LDCs is important for improved understanding of both a phylum of tremendous ecological impact and an environmental challenge relevant to all photosynthetic organisms. For ease of research, however, almost all experiments on cyanobacteria have been conducted in simplifying constant-light conditions. One of the findings of the LDC work that exists is that oxidative stress management is likely a key component for surviving LDCs (Diamond et al., 2017; Welkie et al., n.d.). Similar to the LDC-sensitive circadian clock mutants *rpaA* and *kaiA*, lethality in the *dacA* mutant occurred specifically upon the onset of night following high oxidative stress at dusk (Fig. 4.3-2). The death in the mutant begins concurrent with a spike in c-di-AMP level in the WT. This correlation suggests a role for the molecule in the day-night transition, a seemingly crucial period for surviving LDCs in the mutants where it has been studied (Fig. 4.3-1C). Notably, a number of the top synthetically interacting genes determined by IRB-Seq are involved in oxidative stress mitigation. Nevertheless, the pull-down of interactors of c-di-AMP did not identify circadian clock proteins, and the survival phenotype of the *dacA* mutant in LDCs is less severe than that of clock mutants *rpaA* and *kaiA* (Diamond et al., 2017; Welkie et al., n.d.). Thus, although the *dacA* mutant is similar to LDC-sensitive clock mutants in its oxidative stress and LDC sensitivity, these phenotypes may be caused through a different pathway.

A possible basis for the *dacA* mutant's sensitivity to LDCs and oxidative stress is through potassium transport. C-di-AMP has previously been associated with multiple potassium transporters (Corrigan and Gründling, 2013). In this study the Kdp potassium pump regulator, KdpD (Synpcc7942_1476), was one of three proteins that binds c-di-AMP through both pull-down and DRaCALA. Two proteins from the Trk family of potassium transporters (Synpcc7942_1080 and Synpcc7942_1081) were found among the top synthetic interactions, suggesting that the *dacA* mutant is sensitized to potassium import mutations. Altered potassium transport previously has been shown in cyanobacteria to sensitize the cells to the oxidative stress-producing conditions of high-light and heavy metal exposure (Checchetto et al., 2012; 2016). Therefore, the canonical role of c-di-AMP in potassium homeostasis may, in *S. elongatus*, be involved in the non-canonical function of LDC survival through oxidative stress regulation.

Future Uses of IRB-Seq. IRB-Seq as an inexpensive and straightforward approach to high throughput quantitative interaction screens and is suitable for addressing an array of questions in different organisms. With 25 published RB-TnSeq libraries (Price et al., 2016; Rubin et al., 2015) and many more under development, there exists ample starting material for IRB-Seq screens assuming the ability to deliver a second mutation into the host with high efficiency. The experimental pipeline and analysis tools developed here should make this assay feasible for genetic interactions of any gene of interest. A further use of IRB-Seq arises from our finding that mutants of DNA uptake and homologous recombination are underrepresented in the double mutant libraries. We are

currently leveraging this finding to reveal the full set of non-transforming strains to better understand the cellular machinery involved in competency and homologous recombination. Furthermore, the addition of multiple mutations into the library, or reporters paired with cell sorting, should enable screening for more complex genetic interactions as well as identification of effects on gene expression.

Tables.

Table 4.3-1. C-di-AMP binding candidates.

7942 ID	Name	Annotation	Domains with known c-di-AMP binding*
Synpcc7942_2278		conserved hypothetical protein	
Synpcc7942_1476		conserved hypothetical protein	Nitrogen regulatory PII-like, alpha/beta (IPR011322) {Choi:2015ey, Muller:2015kp, Gundlach:2015iw, Campeotto:2015ch}
Synpcc7942_1729		potassium-transporting ATPase D chain	Universal stress protein (PF00582) {Moscoso:2015cz}
Synpcc7942_1588		CBS	Universal stress protein (PF00582) {Moscoso:2015cz}; Two CBS domains (PF00571) {Huynh:2016iw}
Synpcc7942_1416	topA	DNA topoisomerase I	
Synpcc7942_0420		conserved hypothetical protein	
Synpcc7942_0546	nha4	Na ⁺ /H ⁺ antiporter	Universal stress protein (PF00582) {Moscoso:2015cz}
Synpcc7942_1322	psaE	Photosystem I subunit IV	
Synpcc7942_1886		Exonuclease RecJ	DHH (PF01368) and DHHA1 (PF02272) {Bai:2013gf} {Manikandan:2014iu}
Synpcc7942_1874	TrmH	RNA methyltransferase TrmH, group 2	
Synpcc7942_0301		Single-stranded DNA-binding protein	

*Domains in the proteins pulled-down with c-di-AMP probes, for which literature exists showing interaction with the molecule in other organisms.

Figures.

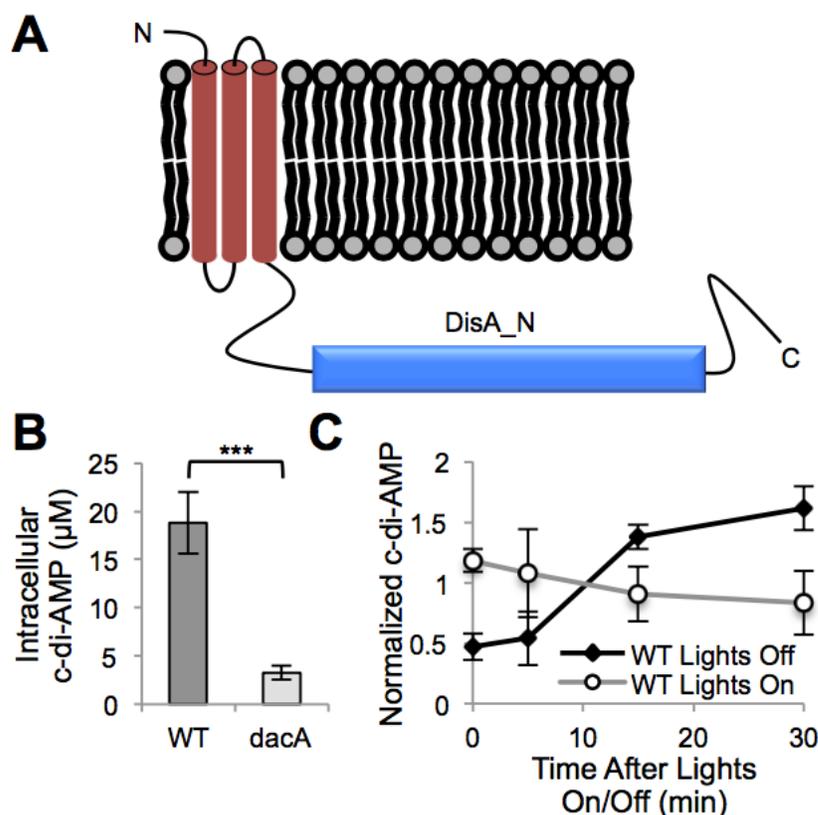


Fig. 4.3-1. Presence, synthase, and light dependence of c-di-AMP in *S. elongatus*. (A) DacA protein, with membrane association determined by phobius (Käll et al., 2004) and the DAC domain identified by Pfam (PF02457)(Finn et al., 2016). (B) Intracellular c-di-AMP measured by LC-MS for WT and *dacA* transposon mutant (8S16-L9). The error bars represent standard error (SE) of five time points taken throughout a 24 hour light-dark cycle in quadruplicate. *** $P < 10^{-7}$ (Mann-Whitney-Wilcoxon Test). (C) C-di-AMP quantities upon the onset of darkness in WT, normalized to average value of replicate. Error bars represent SE of four replicates.

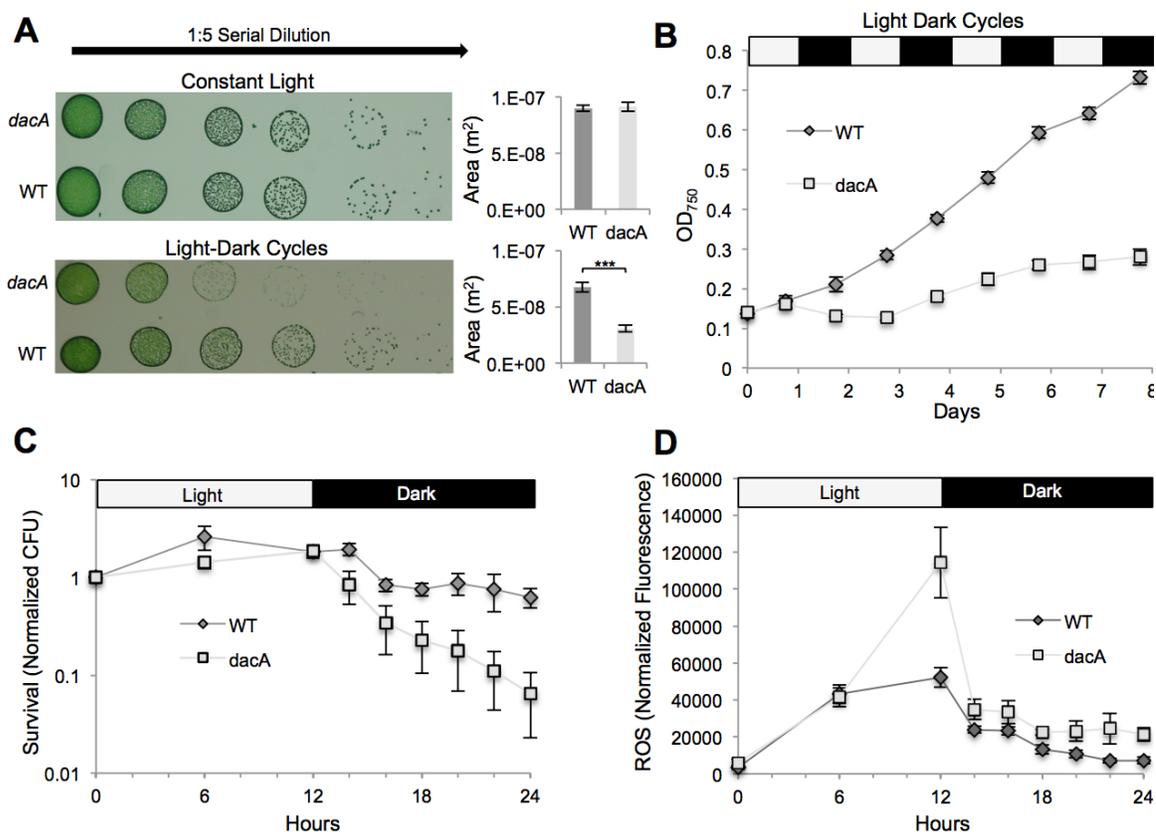


Fig. 4.3-2. Sensitivity of *dacA* mutant to LDCs. (A) Growth of WT and *dacA* transposon mutant (8S16-L9) measured by spot plate under constant light and LDCs. Colony area was measured with ImageJ (Schindelin et al., 2015). *** $P < 10^{-5}$ (Mann-Whitney-Wilcoxon Test). (B) Growth curve of WT and *dacA* mutant in liquid culture in bioreactors under LDCs. (C) High resolution measurement of survival of WT and *dacA* mutant throughout one LDC. Survival is quantified by CFU present at each time point normalized to CFU present at the first time point for each replicate. (D) ROS measured by H2DCFDA fluorescence Normalized by OD₇₅₀. Error bars in all figure parts indicate SE of four replicates.

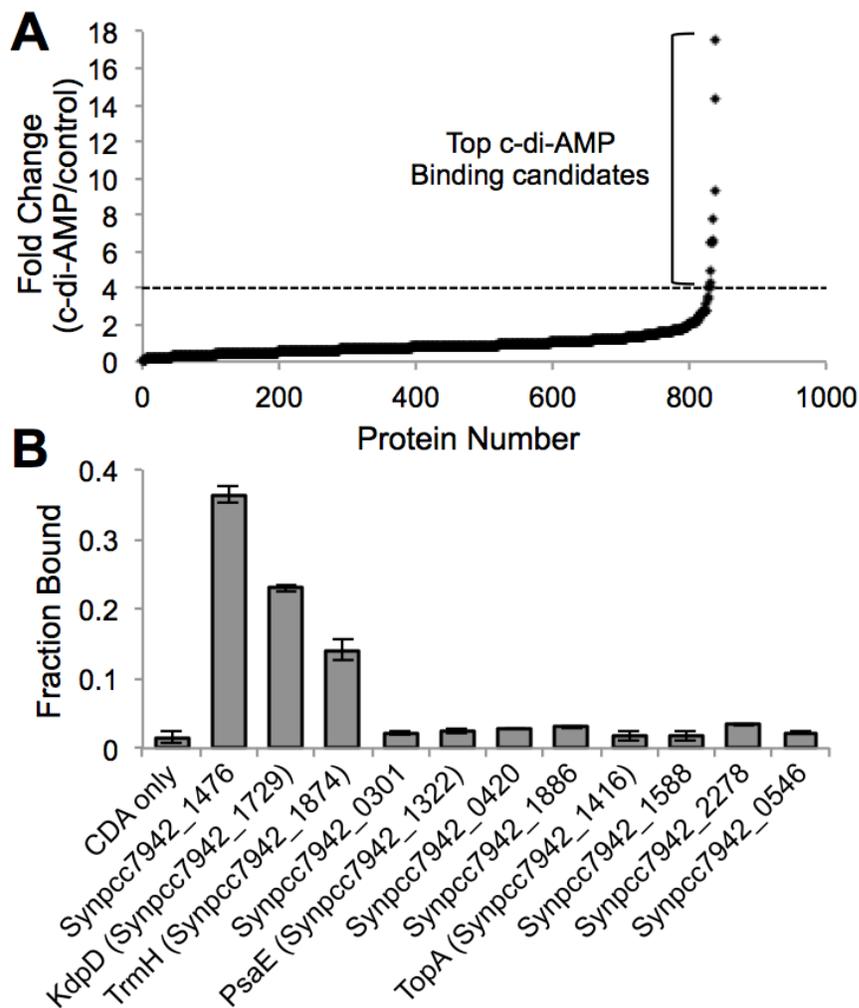


Fig. 4.3-3. Identifying c-di-AMP binding proteins. (A) Protein interaction with c-di-AMP bound beads ordered by fold change (c-di-AMP sepharose beads/control sepharose beads). Top binding candidates (>4 fold change) all have $fdr < .05$. (B) Direct binding of candidate proteins expressed in *E. coli* determined by DRaCALA on cell lysate. Error bars indicate SE of two replicates.

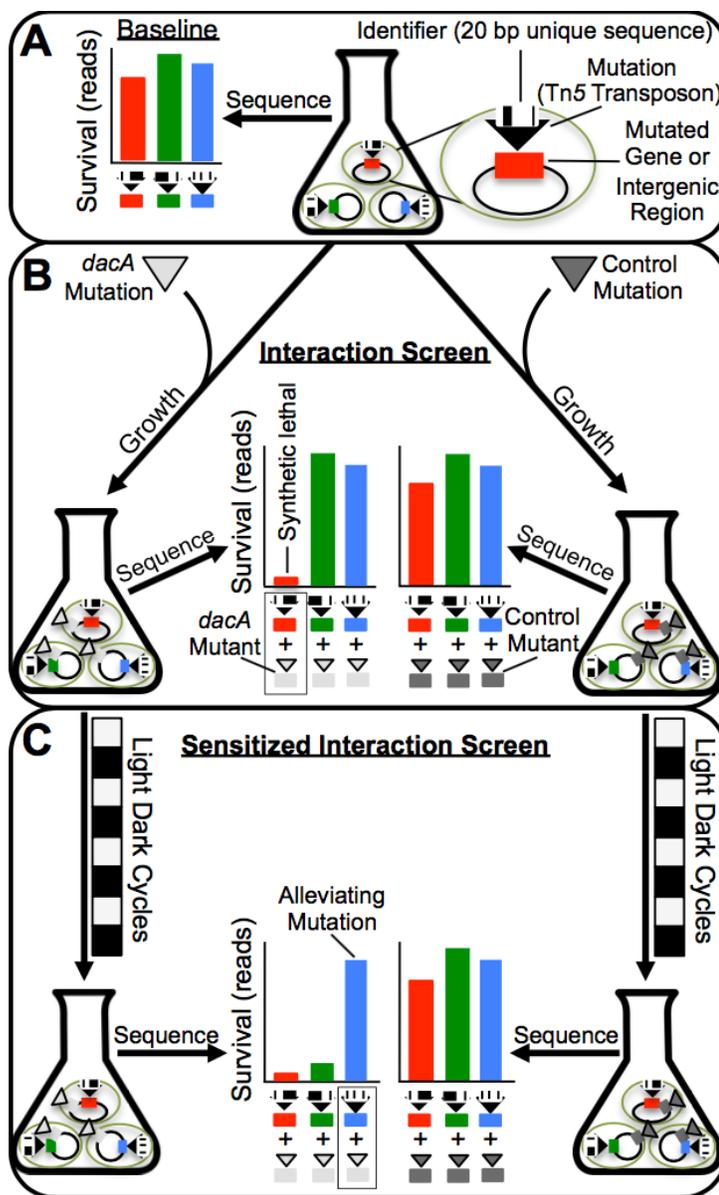


Fig. 4.3-4. IRB-Seq approach to genetic interaction screens. (A) Each mutant in the starting library (Rubin et al., 2015) contains a loss of function mutation with a unique identifier sequence, “barcode”, that has been previously linked to the mutation’s locus. After the library is thawed the barcodes present in each mutant are sequenced using next-generation sequencing to determine their baseline level. (B) The library is then split into two aliquots with one receiving an experimental mutation, and one receiving a control mutation. After outgrowth, these two aliquots are sequenced for barcodes, which allow for identification of genetic interactions between the experimental mutation and the constituent mutants of the library. (C) The double mutant library is again outgrown under a condition of stress for the introduced mutant, and again sequenced to determine genetic interactions under this sensitized condition.

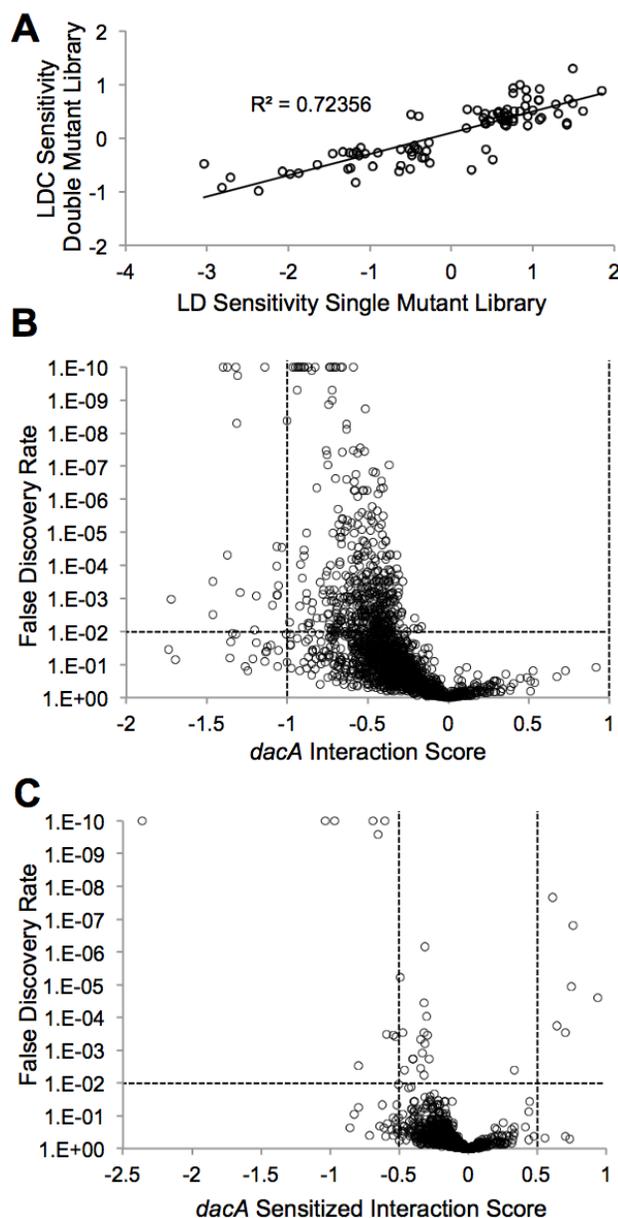


Fig. 4.3-5. DacA genetic interactions using IRB-Seq. (A) Validation of double mutation screening by comparison to previous LDC sensitivity screen. Each circle represents a gene's score for LDC sensitivity from a previously conducted screen on LDCs in the single mutant library (x-axis) (Welkie et al., n.d.), compared to a similar screen conducted in the double mutant library containing the control mutation (y-axis). A linear regression analysis was used to determine correlation. (B and C) Plots of (B) genetic interactions and (C) LDC sensitized genetic interaction of library genes with *dacA*. Genes above horizontal line dashed line have $FDR < .01$ (Linear mixed-effects model). Genes with interaction scores greater than absolute value (B) 1 or (C) .5 are indicated by vertical dashed lines. All points with $FDR < 10^{-10}$ are plotted as $FDR = 10^{-10}$.

References.

- Agostoni, M., Montgomery, B.L., 2014. Survival strategies in the aquatic and terrestrial world: the impact of second messengers on cyanobacterial processes. *Life (Basel)* 4, 745–769.
- Allahverdiyeva, Y., Mustila, H., Ermakova, M., Bersanini, L., Richaud, P., Ajlani, G., Battchikova, N., Cournac, L., Aro, E.-M., 2013. Flavodiiron proteins Flv1 and Flv3 enable cyanobacterial growth and photosynthesis under fluctuating light. In: Presented at the Proceedings of the
- Angermayr, S.A., Gorchs Rovira, A., Hellingwerf, K.J., 2015. Metabolic engineering of cyanobacteria for the synthesis of commodity products. *Trends Biotechnol.* 33, 352–361.
- Brochado, A.R., Typas, A., 2013. High-throughput approaches to understanding gene function and mapping network architecture in bacteria. *Current Opinion in Microbiology*.
- Bryant, D.A., 2003. The beauty in small things revealed. *Proc Natl Acad Sci USA* 100, 9647–9649.
- Campeotto, I., Zhang, Y., Mladenov, M.G., Freemont, P.S., Gründling, A., 2015. Complex structure and biochemical characterization of the *Staphylococcus aureus* cyclic diadenylate monophosphate (c-di-AMP)-binding protein PstA, the founding member of a new signal transduction protein family. *J. Biol. Chem.* 290, 2888–2901.
- Cao, Y., Pan, Y., Huang, H., Jin, X., Levin, E.J., Kloss, B., Zhou, M., 2013. Gating of the TrkH ion channel by its associated RCK protein TrkA. *Nature* 496, 317–322.
- Checchetto, V., Segalla, A., Allorent, G., 2012. Thylakoid potassium channel is required for efficient photosynthesis in cyanobacteria. In: Presented at the Proceedings of the
- Checchetto, V., Segalla, A., Sato, Y., Bergantino, E., Szabo, I., Uozumi, N., 2016. Involvement of Potassium Transport Systems in the Response of *Synechocystis* PCC 6803 Cyanobacteria to External pH Change, High-Intensity Light Stress and Heavy Metal Stress. *Plant Cell Physiol.* 57, 862–877.
- Choi, P.H., Sureka, K., Woodward, J.J., Tong, L., 2015. Molecular basis for the recognition of cyclic-di-AMP by PstA, a PII-like signal transduction protein. *MicrobiologyOpen* 4, 361–374.

- Commichau, F.M., Dickmanns, A., Gundlach, J., Ficner, R., Stülke, J., 2015. A jack of all trades: the multiple roles of the unique essential second messenger cyclic di-AMP. *Molecular Microbiology* 97, 189–204.
- Corrigan, R.M., Abbott, J.C., Burhenne, H., Kaefer, V., Gründling, A., 2011. c-di-AMP is a new second messenger in *Staphylococcus aureus* with a role in controlling cell size and envelope stress. *PLoS Pathog.* 7, e1002217.
- Corrigan, R.M., Bowman, L., Willis, A.R., Kaefer, V., Gründling, A., 2015. Cross-talk between two nucleotide-signaling pathways in *Staphylococcus aureus*. *J. Biol. Chem.* 290, 5826–5839.
- Corrigan, R.M., Campeotto, I., Jeganathan, T., Roelofs, K.G., Lee, V.T., Gründling, A., 2013. Systematic identification of conserved bacterial c-di-AMP receptor proteins. *Proc Natl Acad Sci USA* 110, 9084–9089.
- Corrigan, R.M., Gründling, A., 2013. Cyclic di-AMP: another second messenger enters the fray. *Nat Rev Micro* 11, 513–524.
- Dejesus, M.A., Nambi, S., Smith, C.M., Baker, R.E., Sasseti, C.M., Ioerger, T.R., 2017. Statistical analysis of genetic interactions in Tn-Seq data. *Nucleic Acids Res.* 1–11.
- Diamond, S., Jun, D., Rubin, B.E., Golden, S.S., 2015. The circadian oscillator in *Synechococcus elongatus* controls metabolite partitioning during diurnal growth. *Proc Natl Acad Sci USA* 112, E1916–25.
- Diamond, S., Rubin, B.E., Shultzaberger, R.K., Chen, Y., Barber, C.D., Golden, S.S., 2017. Redox crisis underlies conditional light-dark lethality in cyanobacterial mutants that lack the circadian regulator, RpaA. *Proc Natl Acad Sci USA* 114, E580–E589.
- Finn, R.D., Coghill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L., Potter, S.C., Punta, M., Qureshi, M., Sangrador-Vegas, A., Salazar, G.A., Tate, J., Bateman, A., 2016. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* 44, D279–D285.
- Flombaum, P., Gallegos, J.L., 2013. Present and future global distributions of the marine Cyanobacteria *Prochlorococcus* and *Synechococcus*. In: Presented at the Proceedings of the
- Gray, A.N., Koo, B.-M., Shiver, A.L., Peters, J.M., Osadnik, H., Gross, C.A., 2015. High-throughput bacterial functional genomics in the sequencing era. *Current Opinion in Microbiology* 27, 86–95.
- Gundlach, J., Dickmanns, A., Schröder-Tittmann, K., Neumann, P., Kaesler, J., Kampf, J., Herzberg, C., Hammer, E., Schwede, F., Kaefer, V., Tittmann, K.,

- Stülke, J., Ficner, R., 2015. Identification, Characterization, and Structure Analysis of the Cyclic di-AMP-binding P II-like Signal Transduction Protein DarA. *Journal of Biological Chemistry* 290, 3069–3080.
- Gutu, A., O'Shea, E.K., 2013. Two antagonistic clock-regulated histidine kinases time the activation of circadian gene expression. *Mol. Cell* 50, 288–294.
- Hood, R.D., Higgins, S.A., Flamholz, A., Nichols, R.J., Savage, D.F., 2016. The stringent response regulates adaptation to darkness in the cyanobacterium *Synechococcus elongatus*. *Proc Natl Acad Sci USA* 113, E4867–76.
- Huynh, T.N., Luo, S., Pensinger, D., Sauer, J.-D., Tong, L., Woodward, J.J., 2015. An HD-domain phosphodiesterase mediates cooperative hydrolysis of c-di-AMP to affect bacterial growth and virulence. *Proc Natl Acad Sci USA* 112, E747–56.
- Johnson, L.S., Eddy, S.R., Portugaly, E., 2010. Hidden Markov model speed heuristic and iterative HMM search procedure. *BMC Bioinformatics* 11, 431.
- Käll, L., Krogh, A., Sonnhammer, E.L.L., 2004. A Combined Transmembrane Topology and Signal Peptide Prediction Method. *J. Mol. Biol.* 338, 1027–1036.
- Markson, J.S., Piechura, J.R., Puszynska, A.M., O'Shea, E.K., 2013. Circadian control of global gene expression by the cyanobacterial master regulator RpaA. *Cell* 155, 1396–1408.
- Matsunaga, T., Burgess, J.G., Yamada, N., 1993. An ultraviolet (UV-A) absorbing biopterin glucoside from the marine planktonic cyanobacterium *Oscillatoria* sp. *Applied microbiology*
- Meeske, A.J., Sham, L.-T., Kimsey, H., Koo, B.-M., Gross, C.A., Bernhardt, T.G., Rudner, D.Z., 2015. MurJ and a novel lipid II flippase are required for cell wall biogenesis in *Bacillus subtilis*. *Proc Natl Acad Sci USA* 112, 6437–6442.
- Moon, Y.-J., Kim, S., Chung, Y.-H., 2012. Sensing and Responding to UV-A in Cyanobacteria. *IJMS* 13, 16303–16332.
- Müller, M., Hopfner, K.-P., Witte, G., 2015. c-di-AMP recognition by *Staphylococcus aureus* PstA. *FEBS Lett.* 589, 45–51.
- Nakanishi, K., Nureki, O., 2005. Recent progress of structural biology of tRNA processing and modification. *Mol. Cells* 19, 157–166.
- Nelson, J.W., Sudarsan, N., Furukawa, K., Weinberg, Z., Wang, J.X., Breaker, R.R., 2013. Riboswitches in eubacteria sense the second messenger c-di-AMP. *Nature Chemical Biology* 9, 834–839.

- Oppenheimer-Shaanan, Y., Wexselblatt, E., Katzhendler, J., Yavin, E., Ben-Yehuda, S., 2011. c-di-AMP reports DNA integrity during sporulation in *Bacillus subtilis*. *EMBO Rep.* 12, 594–601.
- Price, M.N., Wetmore, K.M., Waters, R.J., Callaghan, M., Ray, J., Kuehl, J.V., Melnyk, R.A., Lamson, J.S., Suh, Y., Esquivel, Z., Sadeeshkumar, H., Chakraborty, R., Rubin, B.E., Bristow, J., Blow, M.J., Arkin, A.P., Deutschbauer, A.M., 2016. Deep Annotation of Protein Function across Diverse Bacteria from Mutant Phenotypes. *bioRxiv*.
- Roelofs, K.G., Wang, J., Sintim, H.O., Lee, V.T., 2011. Differential radial capillary action of ligand assay for high-throughput detection of protein-metabolite interactions. *Proc Natl Acad Sci USA* 108, 15528–15533.
- Rubin, B.E., Wetmore, K.M., Price, M.N., Diamond, S., Shultzaberger, R.K., Lowe, L.C., Curtin, G., Arkin, A.P., Deutschbauer, A., Golden, S.S., 2015. The essential gene set of a photosynthetic organism. *Proc Natl Acad Sci USA* 201519220–10.
- Sakata, S., Mizusawa, N., Kubota-Kawai, H., Sakurai, I., Wada, H., 2013. Psb28 is involved in recovery of photosystem II at high temperature in *Synechocystis* sp. PCC 6803. *BBA - Bioenergetics* 1827, 50–59.
- Schindelin, J., Rueden, C.T., Hiner, M.C., Eliceiri, K.W., 2015. The ImageJ ecosystem: An open platform for biomedical image analysis. *Molecular Reproduction and Development* 82, 518–529.
- Shaku, K., Shimakawa, G., Hashiguchi, M., Miyake, C., 2015. Reduction-Induced Suppression of Electron Flow (RISE) in the Photosynthetic Electron Transport System of *Synechococcus elongatus* PCC 7942. *Plant Cell Physiol.* pcv198–11.
- Sureka, K., Choi, P.H., Precit, M., Delince, M., Pensinger, D.A., Huynh, T.N., Jurado, A.R., Goo, Y.A., Sadilek, M., Iavarone, A.T., Sauer, J.-D., Tong, L., Woodward, J.J., 2014. The cyclic dinucleotide c-di-AMP is an allosteric regulator of metabolic enzyme function. *Cell* 158, 1389–1401.
- Terauchi, K., Ohmori, M., 2004. Blue light stimulates cyanobacterial motility via a cAMP signal transduction system. *Molecular Microbiology* 52, 303–309.
- Tsinoremas, N.F., Ishiura, M., Kondo, T., Andersson, C.R., Tanaka, K., Takahashi, H., Johnson, C.H., Golden, S.S., 1996. A sigma factor that modifies the circadian expression of a subset of genes in cyanobacteria. *EMBO J.* 15, 2488–2495.
- van Opijnen, T., van Opijnen, T., Bodi, K.L., Bodi, K.L., Camilli, A., Camilli, A.,

2009. Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. *Nature Methods* 6, 767–772.
- Wachi, Y., Burgess, J.G., Iwamoto, K., Yamada, N., 1995. Effect of ultraviolet-A (UV-A) light on growth, photosynthetic activity and production of biopterin glucoside by the marine UV-A resistant cyanobacterium *Oscillatoria* sp. ... et *Biophysica Acta* (BBA ... 1244, 165–168.
- Welkie, D.G., Rubin, B.E., Chang, Y., Smith, S., LiWang, A., Golden, S., n.d. KaiA's essential role in day-night survival in *Synechococcus elongatus*.
- Wetmore, K.M., Price, M.N., Waters, R.J., Lamson, J.S., He, J., Hoover, C.A., Blow, M.J., Bristow, J., Butland, G., Arkin, A.P., Deutschbauer, A., 2015. Rapid Quantification of Mutant Fitness in Diverse Bacteria by Sequencing Randomly Bar-Coded Transposons. *mBio* 6, e00306–15–15.
- Whiteley, A.T., Pollock, A.J., Portnoy, D.A., 2015. The PAMP c-di-AMP Is Essential for *Listeria monocytogenes* Growth in Rich but Not Minimal Media due to a Toxic Increase in (p)ppGpp. *Cell Host and Microbe* 17, 788–798.
- Witte, G., Hartung, S., Büttner, K., Hopfner, K.-P., 2008. Structural biochemistry of a bacterial checkpoint protein reveals diadenylate cyclase activity regulated by DNA recombination intermediates. *Mol. Cell* 30, 167–178.
- Xayarath, B., Freitag, N.E., 2015. Uncovering the Nonessential Nature of an Essential Second Messenger. *Cell Host and Microbe* 17, 731–732.

4.4 Acknowledgments

Chapter 4, in part, is made up of reprints of two manuscripts in preparation. The first manuscript in preparation in section **4.2 RB-TnSeq Under Light-Dark Cycles** is: Welkie DG, Rubin BE, Chang YG, Diamond S, LiWang A, Golden SS. The essential role of the circadian clock protein KaiA for survival of *Synechococcus elongatus* sp. PCC 7942 in alternating light-dark environments. (In Preparation). The dissertation author will be the secondary author of this paper. The second manuscript in preparation in section **4.3 The Role of c-di-AMP in Light-Dark Cycles** is: Rubin BE, Huynh TA, Welkie DG, Diamond S, Lowe LC, Jenny JL, Woodward JJ, Golden SS. High-throughput interaction screens illuminate the role of c-di-AMP in cyanobacterial nighttime survival. (In Preparation). The dissertation author will be the primary author of this paper.

CHAPTER 5: Conclusion

5.1 Discussion

In this work we applied traditional genomics, biochemical assays, metabolomics, proteomics, and *in silico* modeling, and developed RB-TnSeq in *S. elongatus* to expand our understanding of two central unknowns in cyanobacterial biology. The first is the physiology of light-dark cycle (LDC) survival. In this area we identified the full set of genes that cause impair dark survival under LDCs when mutated. In addition, we delved in to the mechanism of LDC involvement for the circadian clock genes: *kaiC*, *rpaA*, and *kaiA*. Finally, we discovered the presence and activity of c-di-AMP in cyanobacteria, the cyclase responsible for its synthesis, and the signaling nucleotide's role in LDC physiology.

The second fundamental unknown in cyanobacterial biology that we have explored is the large number of genes without functional annotations. To address this gap we have used RB-TnSeq to conduct more than 200 quantitative whole-genome mutant screens under approximately 50 experimental conditions. The data produced has allowed approximately ~40% of genes to be connected to phenotypes.

Common themes in the four studies presented here allow us to identify overarching principles of LDC survival. Mutants sensitive to LDCs consistently show increased reactive oxygen species (ROS). In the LDC-sensitive mutants where we have quantified ROS, we find that values are

significantly higher than WT during the day, but cause death only upon the onset of darkness. These findings are supplemented by RB-TnSeq and metabolomics data which together suggest that increased ROS is caused by metabolic imbalance, which is likely not lethal during the day because of ample antioxidant in the form of NADPH produced by photosynthesis. However, at night upon the immediate cessation of NADPH-producing photosynthesis, but in the presence of the leftover ROS caused by it, mutants that are unable to properly scavenge ROS are no longer viable. One likely source of this metabolic imbalance that is common to many LDC-sensitive mutants is an inability to activate the oxidative pentose phosphate pathway, the sole source of NADPH at nighttime (Knoop et al., 2013; Waldbauer et al., 2012). Therefore, the studies contained in this dissertation implicate inability to clear ROS at nighttime as a key factor in LDC sensitivity.

The RB-TnSeq approach to whole genome screens in *S. elongatus* was developed here with the original aim of contributing to LDC research. It has achieved this purpose, but also emerged as a powerful tool for broader functional annotation. From it we have measured the contributions to fitness of 96% of protein-coding genes, approximately half of noncoding RNAs (ncRNAs), and intergenic regions genome-wide. We have been able to contextualize these data by overlaying them on a genome-wide metabolic model. This comparison revealed members of the essential gene study that were surprising in their essential or non-essential nature. In the process we were able to improve the accuracy of the metabolic model by taking into account the RB-TnSeq *in vivo*

data. In addition, the hundreds of screens we have conducted on the library have allowed us to connect more than 40% of genes with experimental conditions in which they are important for survival, extending functional predictions. Thus, the RB-TnSeq approach has proven a powerful lens for viewing gene function in the light-driven lifestyle.

5.2 Future Directions

From this work we have proposed a model of LDC physiology where in the management of oxidative stress at night largely through the oxidative pentose phosphate pathway is key to survival. The findings that suggest this mechanism, however, are largely indirect or correlation based. It will be important in future studies to determine the mechanism and casual nature of oxidative stress on LDC survival. This goal could be accomplished by artificially increasing antioxidant availability, and specifically NADPH, at night in LDC-sensitive mutants and assaying for suppression. This step along with more observational work into the aberrations in metabolism in LDC-sensitive mutants will lead to a more mechanistic understanding of the role of oxidative stress in LDC physiology.

There is great potential for future applications of RB-TnSeq and IRB-Seq in *S. elongatus*. The most obvious is further whole-genome fitness screens. With the pipeline now in place for RB-TnSeq in *S. elongatus*, each whole-genome mutant screen now requires a PCR reaction and 1/50th of an Illumina HiSeq 4000 lane. As a result, for any project in which the survival of loss-of-function mutants under a spectrum of experimental conditions would likely be revealing, the hurdle

for conducting a quantitative whole-genome mutant screen is extremely low. As an example, testing which non-essential mutations cause hypersensitivity or resistance to a panel of 50 grazers would be straightforward experimentally and cost less than \$1000 in sequencing. Regardless of the condition, each new screen is valuable in itself because it provides new phenotype-to-genotype connections, which in turn provide the raw material for useful meta-analysis of genes across conditions (Price et al., 2016)

If the phenotype of interest is not fitness, the library can still have utility. As shown in our screen for biofilm formers, when mutants that display the phenotype of interest can be separated from those that are not, RB-TnSeq can be used to discover contributing genes. While biofilms provide inherent methods for their separation, other phenotypes such as cell size or protein expression may be harder to assay using RB-TnSeq. However, separation techniques such as cell sorting may allow screens to be conducted with more complex phenotypes. In this case, characteristics distinguishable by fluorescence such as size and shape can be screened. In addition, having now shown that secondary mutations can be added to the library, it is feasible to cell sort based on expression using a reporter. In this way, screens may be conducted for a broad range of phenotypes beyond survival.

Another use of Rb-TnSeq developed here is whole genome high-throughput quantitative interaction screens (IRB-Seq). This approach enables the identification and quantification of alleviating or synthetic interactions with a mutant of interest on a genome-wide scale. IRB-Seq is considerably quicker and

less expensive than previous transposon sequencing-based methods of interaction screening (Dejesus et al., 2017). In addition, the value of IRB-Seq's sensitivity and quantitative nature is clear for mutants, such as *dacA*, for which the phenotype is not strong. Finally, the high-throughput nature of the approach makes it feasible to test genetic interactions under many different conditions. The ability to inexpensively and easily discover the global network of interactions with a gene of interest, as well as the pathway information inherent to this data, has obvious value to many projects. Based on early interest in the experimental and computational pipeline developed here for IRB-Seq, we expect it to be valuable in our *S. elongatus* library and applied in other RB-TnSeq libraries.

Data analysis is another area of RB-TnSeq where there is great potential for improvement. While we have mined RB-TnSeq for essential loci, conditionally important genes, and metabolic modeling, there are other valuable perspectives through which to view the data. Perhaps the most exciting of these is considering mutant fitness information beyond protein-coding genes. Conditional fitness for mutants in the hundreds of screens conducted is available genome-wide, but has only been analyzed for those genes that encode proteins. There is much more to the genome than these protein-coding genes. Consider that there are approximately 1,600 ncRNAs in the *S. elongatus* genome that are non-ribosomal and are not tRNAs (Vijayan et al., 2011). Only three of these currently have functional annotations in the National Center for Biotechnology Information (NCBI). Connecting even a small portion of these ncRNAs to the conditions in which they are important for survival represents a massive step forward in

assigning meaning to this genetic material. The same analysis could also be done for un-annotated intergenic regions, which similarly represent complete unknowns. In this way, the library could be applied to the genomic regions between protein-coding genes to further illuminate functional knowledge gaps in the *S. elongatus* genome.

Another valuable use of RB-TnSeq data is its overlay onto other whole-genome datasets. This approach is taken here by using the essential-gene dataset to create an improved metabolic model in *S. elongatus*. However, there is now a repository of conditional data of hundreds of screens that could guide further improvement to metabolic models, as well as insights into the physiology of *S. elongatus*. As an example, a model built for the survival of *S. elongatus* in LDCs could be compared to the set of RB-TnSeq screens done in these conditions. This comparison would improve the accuracy of that model, which would in turn provide a better understanding of the essential metabolic functions for LDC survival.

Moving beyond *S. elongatus*, there is likely to be great value in the development of RB-TnSeq libraries in other cyanobacteria and other photosynthetic organisms. Just as the first genome became much more useful because of the presence of a second, comparisons among species of the genotype-to-phenotype landscapes created by RB-TnSeq have already proven valuable (Price et al., 2016). One concrete outcome of having more photosynthetic RB-TnSeq libraries would be a better understanding of the core functionalities required for photosynthesis. Among other applications, these data

would be invaluable to *de novo* synthesis of a photosynthetic organism, just as similar datasets have provided the foundation for building a minimal heterotrophic organism (Hutchison et al., 2016). More generally, RB-TnSeq data across a diversity of photosynthetic organisms would allow processes that are currently understood only through model heterotrophic paradigms to be better adapted to the photosynthetic lifestyle.

References.

- Dejesus, M.A., Nambi, S., Smith, C.M., Baker, R.E., Sasseti, C.M., Ioerger, T.R., 2017. Statistical analysis of genetic interactions in Tn-Seq data. *Nucleic Acids Res.* 1–11.
- Hutchison, C.A., Chuang, R.-Y., Noskov, V.N., Assad-Garcia, N., Deerinck, T.J., Ellisman, M.H., Gill, J., Kannan, K., Karas, B.J., Ma, L., Pelletier, J.F., Qi, Z.-Q., Richter, R.A., Strychalski, E.A., Sun, L., Suzuki, Y., Tsvetanova, B., Wise, K.S., Smith, H.O., Glass, J.I., Merryman, C., Gibson, D.G., Venter, J.C., 2016. Design and synthesis of a minimal bacterial genome. *Science* 351, aad6253–aad6253.
- Knoop, H., Gründel, M., Zilliges, Y., Lehmann, R., Hoffmann, S., Lockau, W., Steuer, R., 2013. Flux balance analysis of cyanobacterial metabolism: the metabolic network of *Synechocystis* sp. PCC 6803. *PLoS Comput. Biol.* 9, e1003081.
- Price, M.N., Wetmore, K.M., Waters, R.J., Callaghan, M., Ray, J., Kuehl, J.V., Melnyk, R.A., Lamson, J.S., Suh, Y., Esquivel, Z., Sadeeshkumar, H., Chakraborty, R., Rubin, B.E., Bristow, J., Blow, M.J., Arkin, A.P., Deutschbauer, A.M., 2016. Deep Annotation of Protein Function across Diverse Bacteria from Mutant Phenotypes. *bioRxiv*.
- Vijayan, V., Jain, I.H., O'Shea, E.K., 2011. A high resolution map of a cyanobacterial transcriptome. *Genome Biol.* 12, R47.
- Waldbauer, J.R., Rodrigue, S., Coleman, M.L., Chisholm, S.W., 2012. Transcriptome and Proteome Dynamics of a Light-Dark Synchronized Bacterial Cell Cycle. *PLoS ONE* 7, e43432–13.