frontiers | Frontiers in Microbiology

ORIGINAL RESEARCH
published: 23 June 2022
doi: 10.3389/fmicb.2022.899150

# Transcriptomic and Phenomic Investigations Reveal Elements in Biofilm Repression and Formation in the Cyanobacterium *Synechococcus elongatus* PCC 7942

*Ryan Simkovsky[1†], Rami Parnasa[2†], Jingtong Wang[1], Elad Nagar[2], Eli Zecharia[2], Shiran Suban[2], Yevgeni Yegorov[2], Boris Veltman[2], Eleonora Sendersky[2], Rakefet Schwarz[2‡] and Susan S. Golden[1,3*§]*

[1]Division of Biological Sciences, University of California, San Diego, San Diego, CA, United States, [2]The Mina and Everard Goodman Faculty of Life Sciences, Bar-Ilan University, Ramat Gan, Israel, [3]Center for Circadian Biology, University of California, San Diego, San Diego, CA, United States

Biofilm formation by photosynthetic organisms is a complex behavior that serves multiple functions in the environment. Biofilm formation in the unicellular cyanobacterium *Synechococcus elongatus* PCC 7942 is regulated in part by a set of small secreted proteins that promotes biofilm formation and a self-suppression mechanism that prevents their expression. Little is known about the regulatory and structural components of the biofilms in PCC 7942, or response to the suppressor signal(s). We performed transcriptomics (RNA-Seq) and phenomics (RB-TnSeq) screens that identified four genes involved in biofilm formation and regulation, more than 25 additional candidates that may impact biofilm formation, and revealed the transcriptomic adaptation to the biofilm state. In so doing, we compared the effectiveness of these two approaches for gene discovery.

**Keywords: biofilm, cyanobacteria, transposon library, RNA-Seq, RB-TnSeq**

OPEN ACCESS

***Edited by:***
*Louis S. Tisa, University of New Hampshire, United States*

***Reviewed by:***
*Ilka Maria Axmann, Heinrich Heine University of Düsseldorf, Germany*
*Bodo Philipp, University of Münster, Germany*
*Ross Carlson, Montana State University, United States*

****Correspondence:***
*Susan S. Golden*
*sgolden@ucsd.edu*

*[†]These authors have contributed equally to this work*
*[‡]Correspondence regarding RNA-Seq and biofilm assays*
*[§]Correspondence regarding RB-TnSeq*

***Specialty section:***
*This article was submitted to Systems Microbiology, a section of the journal Frontiers in Microbiology*

***Received:*** *18 March 2022*
***Accepted:*** *23 May 2022*
***Published:*** *23 June 2022*

***Citation:***
*Simkovsky R, Parnasa R, Wang J, Nagar E, Zecharia E, Suban S, Yegorov Y, Veltman B, Sendersky E, Schwarz R and Golden SS (2022) Transcriptomic and Phenomic Investigations Reveal Elements in Biofilm Repression and Formation in the Cyanobacterium Synechococcus elongatus PCC 7942. Front. Microbiol. 13:899150. doi: 10.3389/fmicb.2022.899150*

## INTRODUCTION

Biofilm formation, the adhesion of bacteria to a surface typically involving the production of an extracellular polymeric substance (EPS), is a lifestyle that allows microorganisms to survive in the face of stresses and threats from their environment, including nutrient depletion (Donlan and Costerton, 2002; Harrison et al., 2019), predation (DePas et al., 2014; Chan et al., 2021), and dessication (Lebre et al., 2017). Biofilms can be detrimental to human health and industrial activities, by enabling persistence and antibiotic resistance in patients and on medical surfaces (Flemming et al., 2016), or by causing biofouling as on ship surfaces (de Carvalho, 2018) or in corrosion of industrial piping (Lenhart et al., 2014). Studies aimed at disrupting deleterious biofilms have led to a wealth of knowledge on the molecular mechanisms regulating and forming biofilms in heterotrophic or pathogenic bacteria. This knowledge includes the requirement for pili and adhesins in finding and attaching to surfaces, intracellular cyclic di-GMP secondary messenger systems for controlling activation of biofilm formation, and production and secretion

of exopolymeric matrix materials that can include exopolysaccharides, amyloidgenic proteins, and extracellular DNA (Flemming et al., 2016).

Photosynthetic cyanobacteria, which are among the planet's dominant producers of oxygen and fixed carbon, can also form biofilms, often in microbial mats where a combination of gas, nutrient, signaling, and light gradients determine the composition, density, and organization of the diverse community of organisms in the assemblage (Prieto-Barajas et al., 2018). Cyanobacteria are a promising agricultural crop for sustainable conversion of solar energy and atmospheric carbon into replacements for petroleum products, including biofuels, industrial chemicals, nutraceuticals, and plastics (Nozzi et al., 2013; Farrokh et al., 2019; Sitther et al., 2020). In this context, biofilm formation can be both disadvantageous, due to biofouling of growth vessels and piping, or beneficial for protection against predation, efficiency in harvesting the crop, or minimizing resources for growth and production in a biofilm as opposed to in a planktonic culture (Bruno et al., 2012; Barros et al., 2020). In contrast to health applications that seek to prevent biofilm formation, the goal for agricultural scale industrial production of cyanobacterial biomass is to be able to reliably activate or deactivate biofilm formation to minimize costs and damage to equipment while maximizing productivity.

Only in recent years have studies on the molecular mechanisms of biofilm formation extended to cyanobacteria, where light limitation through self-shading is an inherent consequence of biofilm structure (Barros et al., 2020). For many cyanobacteria, biofilm formation and regulation appears to have a number of features in common with heterotrophic bacteria: type IV pili and S-layer proteins are required for film formation in *Synechocystis* sp. PCC 6803 (Allen et al., 2019; Conradi et al., 2019), cyclic-di-GMP promotes aggregation in *Thermosynechococcus vulcanus* (Enomoto et al., 2015) and biofilm formation in PCC 6803 (Agostoni et al., 2016), a two-component regulatory system comprising a response regulator and a split histidine kinase controls cell aggregation in PCC 6803 (Kera et al., 2020), and extracellular polysaccharides, such as cellulose in *T. vulcanus*, contribute to the extracellular matrix and cellular aggregation (Kawano et al., 2011).

In contrast, our studies have demonstrated that biofilm formation in *Synechococcus elongatus* PCC 7942, which is by default constitutively suppressed in the laboratory (Schatz et al., 2013), does not require pili (Nagar et al., 2017). In fact, most of the mutations discovered to date that relieve suppression and enable biofilm formation (**Supplementary Table S1**) block formation of the Type IV pilus or its assembly apparatus, including the pilin-encoding *pilA1* and *pilA2* genes; the pilus assembly and Type II secretion proteins encoded by *pilB* (formerly referred to as *t2sE*), *pilC*, and *pilN*; and two proteins encoded by the genes *hfq* and *ebsA* that form a tripartite complex with PilB (Schatz et al., 2013; Nagar et al., 2017; Yegorov et al., 2021). For the majority of these mutants, a biofilm containing 90–95% of the cell population forms 3–4 days following inoculation in bubbling tubes. Electron microscopy has confirmed that the biofilm cells lack pili. Furthermore, they have markedly diminished or altered exoproteomes, suggesting that a single Type IV/II pilus and secretion system fills these dual roles in *S. elongatus* (Schatz et al., 2013; Nagar et al., 2017; Yegorov et al., 2021). Removal of this complex is pleiotropic—impeding other ecologically relevant behaviors such as natural competence and phototaxis-directed twitching motility (Yang et al., 2018; Taton et al., 2020).

Wild-type *S. elongatus* PCC 7942 (WT) does not produce biofilms under laboratory conditions due to the production and secretion of a self-suppressor (Schatz et al., 2013). WT growth medium (conditioned medium, CM) contains accumulated suppressor that prevents biofilm formation by the *pilB* mutant, *PilB*::Tn*5* (formerly named T2SEΩ), but CM from *PilB*::Tn*5* does not, indicating that the mutant still responds to the suppressor it does not secrete (Schatz et al., 2013). The derepression of a set of four small secreted proteins (EbfG1-4), each characterized by a secretion motif shared with microcins, enables biofilm formation in *PilB*::Tn*5* (Schatz et al., 2013; Parnasa et al., 2016, 2019). The processing and secretion of these small proteins requires the cysteine peptidase gene *pteB*, which is cistronic with the *ebfG* genes, and the putative processing peptidase gene *ebfE*. Inactivation of *ebfE* significantly alters the exoproteome in a manner that is distinct from changes observed with inactivation of *pilB* (Parnasa et al., 2019).

The robust formation of biofilms by the recent wild isolate *S. elongatus* UTEX 3055 demonstrates that biofilm formation of PCC 7942 mutants uncovers a process that is reversible in the natural environment (Yang et al., 2018). Decades of laboratory culturing techniques, such as decanting, that subtly select for planktonic cells (Golden, 2019) has likely domesticated the PCC 7942 WT strain into a biofilm-suppressed state through genetic drift or epigenetic changes, which now requires biofilm-inducing mutations to activate the biofilm program (Adomako et al., 2022).

The known gene products involved in suppressing or enabling biofilm formation in *S. elongatus* PCC 7942 (**Supplementary Table S1**) represent only pieces of the structural and regulatory system of biofilm formation, which must include the circuitry to synthesize, detect, and respond to the secreted repressor, as well as the repressed components that contribute to biofilm formation once self-suppression is alleviated. To obtain a more comprehensive understanding of genes involved in biofilm formation we leveraged the natural competence (Shestakov and Khyen, 1970) of *S. elongatus* and genetic tools that include a library of transposon mutagenesis vectors for the rapid individual knockout of nearly any nonessential gene in the genome (Holtman et al., 2005; Chen et al., 2012) and a randomly barcoded transposon sequencing (RB-TnSeq) library for quantitatively assessing the impact of gene mutations across the genome on the fitness of the strain under specific conditions, thus directly linking genotypes with phenotypes (Rubin et al., 2015; Price et al., 2018). Analogous TnSeq libraries in other organisms have been used to study complex behaviors, as in biofilm formation in *Bacillus cereus* (Yan et al., 2017; Okshevsky et al., 2018) or in community interactions of bacteria and fungi on the surface of maturing cheese (Morin et al., 2018; Pierce et al., 2021). In addition to phenomics using RB-TnSeq, which identifies genes whose loss either enhances or reduces

a cell's fitness in a biofilm, we employed transcriptomics using RNA-Seq, which identifies genes whose transcription differs between planktonic and sessile cells. These two data sets suggest that RB-TnSeq is more powerful for identifying genes that contribute to a phenotype, whereas transcriptomics describes the underlying biology of a phenotypic state.

# MATERIALS AND METHODS

## Culture Conditions, Plasmids, and Strains

For standard growth of *S. elongatus* without additional $CO_2$, cultures were grown at 30°C in BG-11 medium (Allen, 1968). For biofilm assays performed in test tubes, where cultures were bubbled with 3% $CO_2$, or in 96-well plates, cultures were inoculated into BG-11 media with HEPES and freshly prepared ferric ammonium citrate and citric acid components as previously described (Sendersky et al., 2017). Preparation and collection of conditioned media from WT cultures grown in BG-11 were performed as described previously (Schatz et al., 2013; Parnasa et al., 2016).

All vectors used in this study are listed in **Supplementary Table S10**. Except for the RB-TnSeq experiments, where strains were derived from the Golden lab culture collection, all *S. elongatus* strains were derived from those in the Schwarz lab culture collection, which lack the small plasmid, pANS. Generation of novel insertional knockout *S. elongatus* strains using Unigene Set (UGS; Holtman et al., 2005; Chen et al., 2012) or the constructed vector EZ27 were performed according to standard transformation methods that take advantage of *S. elongatus*' natural competence and ability to perform double homologous recombination (Clerico et al., 2007). In cases in which other mutations were combined with a *pilB* disruption, *pilB* inactivation was introduced after knockout of the other gene of interest due to impaired competence in *PilB*::Tn*5* (Schatz et al., 2013). Double recombination and segregation of mutants were confirmed *via* PCR using primers listed in **Supplementary Table S10**.

An inactivation vector for gene Synpcc7942_0774 was constructed through blunt-end cloning of a PCR-amplified gene fragment, generated using the K273 and K274 primers listed in **Supplementary Table S10**, into the pJet1.2 vector (Thermo Fisher Scientific). A spectinomycin-resistance cassette was inserted at the XcmI site in the gene fragment, yielding the insertional knockout construct (pEZ27) that was used for transformation and homologous recombination into WT.

## Biofilm Formation Assays

Unless noted otherwise, growth of *S. elongatus* PCC 7942 and all derived strains for the purpose of generating biofilms were performed in bubbling tube cultures containing 25 ml of BG-11 with HEPES after inoculation, using fresh starter cultures derived from solid media cultures, at an $OD_{750}$ of 0.5 at 30°C and under low light (5–30 μmol photons $m^{-2}$ $s^{-1}$), as described previously (Sendersky et al., 2017). Biofilm assays under static conditions were grown in 96-well plates, as previously described (Parnasa et al., 2019), or in 5 ml culture volumes in sterile 25 ml Pyrex glass flasks (# 4980) plugged with cotton balls that were incubated under low light (5–30 μmol photons $m^{-2}$ $s^{-1}$) without shaking for 6 days prior to biofilm quantification. Quantification of the biofilm formation as a percentage of chlorophyll in suspension or using ethanol-extracted crystal violet staining was performed as described previously (Sendersky et al., 2017; Parnasa et al., 2019). Quantitative biofilm formation values in the form of absorbance units (A.U.) were summarized into three categories due to the typically binary, but not always reproducible, nature of these assays in multiple labs and under various conditions: successful biofilms (>1 A.U.), weak biofilms (< 1 A.U., >0.5 A.U.), or no biofilms (< 0.5 A.U.). These categorizations and the number of replicates for each are provided in **Supplementary Tables S7** and **S9**.

## RNA-Seq Experiment

WT and *PilB*::Tn*5* strains were grown for biofilm formation in triplicate sets of bubbling tubes as described previously (Parnasa et al., 2016). Samples were taken for RNA extraction from the liquid supernatants 1 and 4 days after inoculation (Days 1 and 4). For the *PilB*::Tn*5* culture samples on Day 4, samples were first taken from the supernatant fraction, after which the liquid was removed by decanting, the tubes were gently washed with deionized water, and the remaining biofilm left behind was scraped to obtain biofilm samples. Total RNA was extracted as previously described by Tu et al. (2004), with the following modifications: phenol was saturated with water; cell extraction was performed by vortexing for 1 min, incubating at 65°C for 10 min with a brief vortex after 5 min, and incubating on ice for 3 min. The beads and unbroken cells were pelleted by centrifugation (Eppendorf centrifuge, maximal speed, 10 min), and the aqueous phase was subjected to one phenol, two to three phenol-chloroform (1:1), and one chloroform extraction, where chloroform was pre-equilibrated with isoamyl alcohol at 24:1 ratio. Nucleic acids were precipitated by adding 1/3 volume of 8 M LiCl and 2 volumes of ethanol followed by gentle mixing, incubation for 40 min at −80°C or for 2 h at −20°C, and centrifugation as described. The pellet was washed with 70% ethanol and resuspended with HPLC-grade water. Forty microgram of total RNA was treated with 4 U of TURBO™ DNase (Ambion, Catalog #: AM2238) at room temperature for 45 min followed by a boost with 4 U and an additional 45-min incubation. The DNase was removed by phenol-chloroform and chloroform extractions and the RNA was precipitated as described above. RNA sample pellets (Schatz et al., 2013), representing triplicates of the seven sampling groups described above and resulting from the ethanol precipitation, were washed with ice cold 70% ethanol. The resulting air-dried pellets were resuspended in RNase-free water and 3 μg of DNA from each sample was depleted of ribosomal RNA using the Bacteria Ribo-Zero rRNA Removal Kit (Epicentre) and purified using the RNeasy miniprep kit (Qiagen), both performed according to the manufacturer's instructions. Recovery rates for each sample fell in the range of 3.6–9.2%, and rRNA depletion and RNA quality were confirmed for a subset of samples using 1% agarose gel electrophoresis. Purified ribosomal RNA-depleted samples were

submitted to the Scripps Research (La Jolla, CA) sequencing core for strand-specific library preparation and sequencing on an Illumina 1×100 TruSeq system.

## RNA-Seq Data Analysis

FASTQ files were trimmed using trimmomatic (Bolger et al., 2014) prior to analysis of the seven sampling groups with two separate RNA-Seq analysis pipelines: (1) Rockhopper 2 (Tjaden, 2015), an all-in-one analysis tool specifically designed for analysis of bacterial transcriptomics data, using default parameters and (2) a custom scripted pipeline written in shell commands and in R (Team R Development Core, 2018) using Bowtie2 (Langmead and Salzberg, 2012) for alignment to the genome, SAMtools (Li et al., 2009) and BEDTools (Quinlan and Hall, 2010) for file format conversion and generation of coverage files for visualization in IGV (Thorvaldsdóttir et al., 2013), htseq-count for generating strand-specific gene counts (Anders et al., 2015), the DeSeq2 (Love et al., 2014) R package for differential gene expression analysis using an alpha value of 0.05 and either a likelihood ratio test for ANOVA analysis of multiple sample groups simultaneously using or the Wald significance test for pairwise sample group comparisons, the ashr (Stephens, 2017) R package for determining effect size shrinkage estimates, and tidyverse (Wickham et al., 2019) R packages for further data organization and visualization. For both pipelines, significantly differentially expressed genes were determined for each of the 21 possible pairwise comparisons of the 7 sample groups using a false detection rate (FDR; *q*-value from Rockhopper2 and FDR-adjusted *p*-values from DESeq2) threshold of $\leq 0.05$ to determine significance and a $|\log_2(\text{Fold Change})| > 1$ for differentiating highly changed genes, labeled "Up" or "Down," from genes with minor expression changes, labeled "Neutral Up" or "Neutral Down." All alignment statistics, gene counts, fold change values, FDR values, and DEG assignment data for the Rockhopper2 analysis are provided in **Supplementary File S1**, while the same data for the DESeq2-based analysis are provided in **Supplementary File S2**. To consolidate the lists of DEGs for each pairwise comparison from each analysis pipeline, a union of the data sets was performed so that the labels "Up" or "Down" were applied to any gene in which the gene was so labeled in at least one of the data sets, while "Neutral Up" or "Neutral Down" was applied to any remaining gene that was otherwise determined as significantly changed using the FDR threshold in either of the analyses. Comparisons of pairwise comparisons were then performed by manually clustering genes into groups of interest with similar patterns of DEG categorization across relevant pairwise experiments. Groupings that consider the "Neutral" genes (labeled "Full") are provided in **Supplementary File S3**, as well as the groupings discussed herein that do not consider the "Neutral" genes for categorization purposes (labeled "Reduced"). These latter groups of interest served as the basis for the enrichment analysis detailed below. All data regarding the pipeline analyses consolidation, DEG categorization, comparisons of pairwise comparisons, and enrichment analysis are provided in **Supplementary File S3**.

## RB-TnSeq Experiment

For the first experiment, an aliquot of the *S. elongatus* RB-TnSeq version 1.0 library (Rubin et al., 2015) was revived and grown in three flasks of 100-ml BG-11, as described previously (Welkie et al., 2018; Taton et al., 2020). The three flasks were pooled to generate a culture with $OD_{750}$ of 0.4. An aliquot of this pooled culture was pelleted by centrifugation and frozen at −80°C to serve as a T0 sample. A culture (93 ml) of the pooled library was centrifuged at 5,000 rpm at room temperature for 15 min. The pelleted cells were resuspended in 75 ml of BG-11 with HEPES media to an $OD_{750}$ of 0.5, split equally into triplicate bubbling tubes, and allowed to form biofilms for 21 days at 30°C and under low fluorescent light (5 μmol photons $m^{-2}$ $s^{-1}$) while bubbled with hydrated 3% $CO_2$.

For the second experiment, an aliquot of the *S. elongatus* RB-TnSeq version 2.0 library was revived, grown, and sampled for a T0 sample as performed for the version 1.0 library. Aliquots of the library were centrifuged to pellet cells and resuspended in either fresh or WT conditioned BG-11 media, both with HEPES, to an $OD_{750}$ of 0.5. Resuspensions were then split into triplicate bubbling tubes containing 25 ml of culture each or into sterile 25-ml Pyrex glass flasks (# 4980), plugged with cotton balls, containing 5 ml of culture each. Bubbling tube cultures for this experiment were placed under the same conditions as for the first experiment for 15 days to allow biofilms to form. Flask cultures were incubated under low fluorescent light (5 μmol photons $m^{-2}$ $s^{-1}$) without shaking at room temperature for 15 days.

After the appropriate incubation periods, biofilm formation was confirmed visually. Planktonic cell samples were collected by decanting the supernatants of the cultures into sterile conical tubes. Each vessel was then gently washed once with water, which was decanted to a sterile conical tube to investigate settler cells. The remaining biofilms were resuspended in a water wash through physical scraping of the vessel and collected into a sterile conical tube. All samples were then centrifuged to generate cell pellets, which were frozen at −80°C prior to genomic DNA extraction (Clerico et al., 2007), barcode PCR amplification using Q5 DNA polymerase (NEB) and Illumina adaptor-encoded primers that include unique 6-bp TruSeq indexes in the forward primer, PCR purification of barcode amplicons with the DNA Clean and Concentrator kit (Zymo Research), and Illumina HiSeq single-end sequencing at LBNL, all as detailed in previously reported RB-TnSeq protocols (Wetmore et al., 2015).

## RB-TnSeq Data Analysis

Barcode sequences were extracted and counted from FASTQ files using PERL scripts previously reported by Wetmore et al. (2015) using an updated genome annotation that includes the known biofilm enhancing *ebfG* genes and other ORFs that are not annotated in publicly available genome files. Strain fitness, gene fitness, and T-value *t*-like statistics were calculated and a *T*-value threshold for significance of ≥4 was applied

as in Wetmore et al. (2015). Significant genes with an absolute fitness score >1 were labeled "Up" or "Down," while those with absolute fitness scores ≤1 were labeled "Neutral Up" or "Neutral Down."

## Gene Enrichment Analysis

Gene categorical information, including known biofilm genes from previous *S. elongatus* PCC 7942 biofilm publications (Schatz et al., 2013; Parnasa et al., 2016, 2019; Nagar et al., 2017; Yegorov et al., 2021), pili genes described by Taton et al. (2020), essentiality and conservation data from Rubin et al. (2015), functional categories from the JGI IGM (Chen et al., 2017) and CyanoBase (Fujisawa et al., 2017) databases, and signal peptide and secretion predictions (Dilks et al., 2003; Bendtsen et al., 2005; Käll et al., 2007; Imam et al., 2011; Petersen et al., 2011; Pundhir and Kumar, 2011), was amassed for all genes in the *S. elongatus* PCC 7942 genome from appropriate sources and are provided in **Supplementary Tables S1–S3**. Significant enrichments of genes in each interest group identified for RNA-Seq and RB-TnSeq data comparison sets for each of the informational categories were determined using custom R scripts to perform two-sided Fisher's exact tests, with FDR-adjusted *p*-values ≤0.05 being designated as significant. Fold enrichment (*F*) was calculated as the number of genes in the interest group that are also in the category ($N_{gc}$) divided by the number of genes expected in the group and category ($E_{gc}$). This expected number was calculated by multiplying the number of total genes in the interest group ($N_g$) by the frequency of all genes in the genome that are found in the category ($f_c$), which was determined as the number of genes in the category ($N_c$) divided by the number of genes in the genome ($N$).

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

# RESULTS AND DISCUSSION

## RNA-Seq Characterization of WT and Biofilm-Forming Strains

Our first approach aimed to identify genes that are differentially regulated between planktonic and biofilm states. To compare the transcript profiles of cells destined to form or actively forming biofilms with those of planktonic cells, as well as the impact that conditioned (biofilm-suppressing) medium has upon these processes, total RNA was purified from WT or *PilB*::Tn*5* cultures inoculated in either fresh or WT-derived conditioned medium (CM) and sampled 1 and 4 days following inoculation. Only the 4-day *PilB*::Tn*5* cultures grown in fresh BG-11 produced biofilms, so transcript profiles were processed from both planktonic and biofilm fractions from these cultures (**Figure 1A**). Twenty-one samples representing triplicate experiments of seven experimental conditions were depleted of ribosomal RNA and sequenced using an Illumina platform, yielding 6–8 million reads per sample. Data were aligned to the *S. elongatus* PCC 7942 genome and analyzed for differentially expressed genes (DEG) using two different software pipelines: (1) Rockhopper 2, a prokaryotic-specific RNA-seq analysis platform (Tjaden, 2015) and (2) a pipeline of published RNA-Seq software including Bowtie2 (Langmead and Salzberg, 2012) for alignment, htseq-count (Anders et al., 2015) for gene counts, and DESeq2 (Love et al., 2014) for DEG analysis, with each analysis pipeline aligning approximately 99% of trimmed reads for nearly all samples, with less than 1% of these reads mapping to ribosomal RNAs (See **Supplementary Files S1** and **S2** for further details). Both the WT and *PilB*::Tn*5* strains used in these experiments lack the smaller plasmid, pANS, and thus, its potential impact on biofilm formation was not assessed.

ANOVA analysis using DESeq across all samples demonstrated that over 65% (1,776 out of 2,729) of the ORFs encoded by the genome, including all known biofilm-related (**Supplementary Table S1**) and pili (**Supplementary Table S2**) genes except for *tsaP*, are differentially expressed between at least two experimental conditions (**Supplementary File S2** Sheets 4 and 8). Additionally, Rockhopper predicted 118 noncoding RNAs (ncRNAs), of which 42 are antisense to annotated ORFs, 55 overlap with or are enclosed by previously reported potential ncRNAs (Vijayan et al., 2011; Billis et al., 2014), and one overlaps an essential intergenic region (Rubin et al., 2015; **Supplementary Table S4**, **Supplementary File S1** Sheet 9). All but three of these predicted RNAs were differentially expressed between at least two experimental conditions. Although RNA-binding by cyanobacterial Hfq homologs so far has not been demonstrated (Schuergers et al., 2014), we hypothesize that these novel differentially expressed RNAs may represent binding targets.

Both Rockhopper and DESeq analyses were performed for all 21 pairwise comparisons of the seven experimental groups (**Supplementary Files S1** Sheet 6 and **S2** Sheet 5). The union of the lists of DEGs from the two analyses generated a final list of DEGs for all pairwise comparisons (**Supplementary File S3** Sheets 1 and 2); the intersection of the lists proved too stringent and excluded known biofilm forming genes that the union did not.

Ten pairwise comparisons (**Figures 1, 2, Supplementary Figure S2**, and **Supplementary File S3** Sheet 4) are relevant to the experimental variables tested: the *pilB* genetic alteration, biofilm state, time, and media conditions. While recognizing that more modest changes in expression may be biologically relevant, focusing on genes with fold changes >2, designated as strong DEGs (sDEGs), limited the total set considered among the 10 pairwise comparisons nearly four-fold from 1,867 DEGs to 508 sDEGs and resulted in a more limited range of sDEGS per pairwise comparison of 10 and 18 for those involving CM to 344 for *pilB*::Tn*5* biofilm formation over time (**Supplementary Figures S1A,C** and **Supplementary File S3** Sheets 2 and 4), in contrast to 15, 67, and 1,332 DEGs, respectively. A summary of all sDEGs resulting from this analysis, as well as hits from the RB-TnSeq experiments discussed below, is provided in **Supplementary Table S5**.

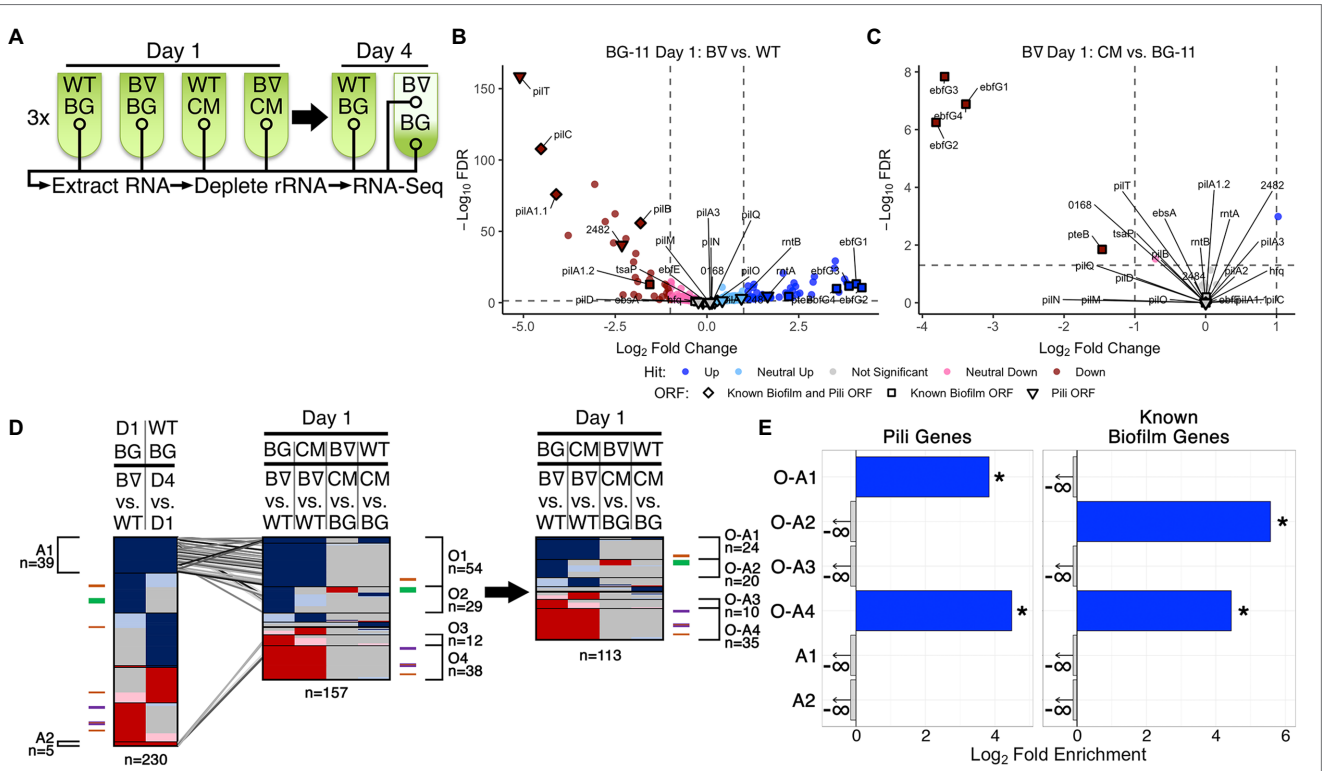



**FIGURE 1 |** RNA-Seq of WT and *PilB*::Tn*5* and Day 1 DEG analysis. **(A)** Diagram of the time course for triplicate samples processed for Wild Type (WT) and *PilB*::Tn*5* (BV) when placed under biofilm forming conditions in either fresh BG-11 (BG) or WT-derived conditioned media (CM). Samples (circles) include planktonic samples from all test tubes and a sample of the biofilm formed by *PilB*::Tn*5* on Day 4 in BG-11. **(B,C)** Volcano plots comparing the $\log_2$ fold change of pairwise comparisons of samples against their $-\log_{10}$ false discovery rate (FDR) as calculated by DESeq2. Data points are colored based on their DEG designation ("Hit"). Labeled shapes indicate known biofilm and pili ORFs. Dashed lines indicate the thresholds used for hit designation. Note that each plot is scaled individually. **(D)** Comparisons of sDEG designations (colored as in **B,C**) of ORFs (*y*-axis) from different relevant pairwise comparisons, as labeled above. The locations in the graph of known pili and biofilm ORFs are indicated by the colored ticks next to each graph (purple = biofilm and pili ORF, orange = pili ORF, green = biofilm ORF) and clusters of genes with common sDEG patterns across the *x*-axis of interest are designated with brackets and labeled with their identifiers and the number of sDEGs in the group. The number of ORFs considered in each graph is provided below the graph. On the left, the Day 1 BG-11 *PilB*::Tn*5* vs. WT pairwise comparison is compared against the WT BG-11 Day 4 vs. Day 1 comparison to identify ORFs related to the stress or density state of the culture. Those ORFs that are similarly up- or downregulated in these two pairwise comparisons (groups A1 and A2) are connected to their location in the middle graph comparing all relevant Day 1 pairwise comparisons by the diagonal lines. These ORFs were then removed from the middle graph to generate the graph on the right. **(E)** Enrichment analyses of pili genes and known biofilm genes in the O-A and A groups of interest labeled in **(D)**. Significant enrichment (FDR ≤ 0.05) is designated by solid bars and stars. Groups that do not have any of the genes analyzed are designated with a negative infinity and arrow, as appropriate for the $\log_2$ scale of the *x*-axis.

Tracking the 29 known biofilm and pili genes (**Supplementary Tables S1** and **S2**, and **Supplementary File S3** Sheet 6) enabled evaluation of the biological significance of these results (Schatz et al., 2013; Parnasa et al., 2016; Nagar et al., 2017; Taton et al., 2020; Yegorov et al., 2021). All except 7 (*tsaP*, *pilD*, *hfq*, *ebfE*, *ebsA*, *pilA2* and the second *pilT* gene) were identified as sDEGs (**Figures 1B–D, Supplementary Figure S2**, and **Supplementary File S3** Sheet 6). While it is reasonable that *tsaP*, *pilD*, and the second *pilT* genes do not have notable impacts on biofilm formation due to degeneracy or their encoded proteins' roles as accessories to the pilus, the genes of the Hfq-EbsA-PilB tripartite complex (Yegorov et al., 2021) or the EbfE-PteB (Parnasa et al., 2016) maturation-secretion system play important roles that likely act at post-transcriptional levels (Zhang et al., 2020; Yegorov et al., 2021). These results demonstrate the limitations of using changes in expression level as a proxy for contribution to a particular phenotype.

## Transcriptome Comparisons of Biofilm and Planktonic Cells Reflect the Biofilm-Proficient *pilB*::Tn*5* Genotype

We initially hypothesized that the Day 4 comparisons between biofilm and planktonic samples would identify novel genes required to regulate and create the biofilm. However, the sDEG patterns instead point to transcriptional consequences of the *pilB*::Tn*5* mutation independent of its biofilm state or to the change in environment associated with existing in the biofilm (**Figure 2, Supplementary Figures S2**, **S3**). None of the known biofilm and pili genes differed sufficiently in expression between *pilB*::Tn*5* biofilms and planktonic cells to be considered as sDEGs (**Figure 2A, Supplementary File S3** Sheet 6), but the majority of them were either upregulated (e.g., *ebfG* genes, named group DF2) or downregulated (e.g., *pilB* operon genes, group DF5) when comparing either of the *PilB*::Tn*5* samples against WT (**Figure 2D**). The sets of similarly up- or

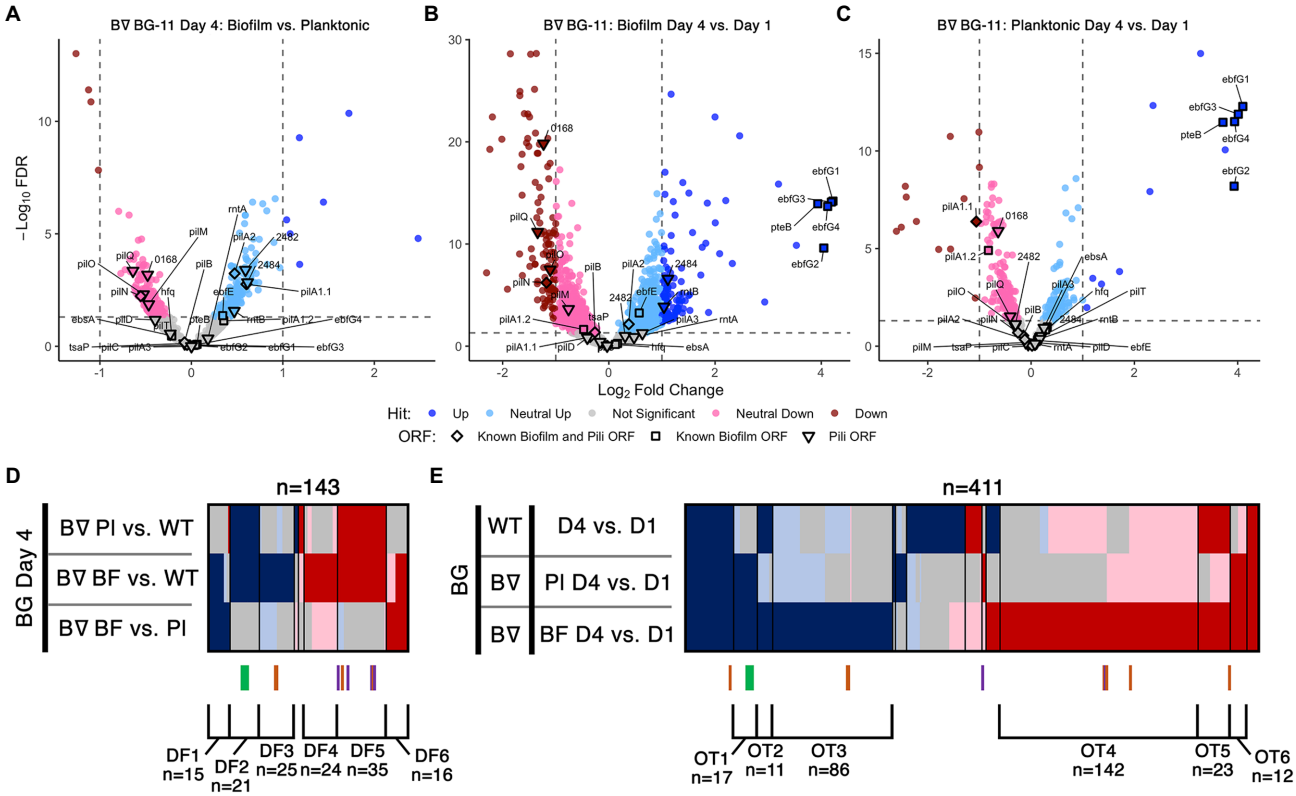



**FIGURE 2 |** Day 4 and time course RNA-Seq analysis. Volcano plots, presented as described for **Figures 1B,C**, of **(A)** Day 4 pairwise comparison of *PilB*::Tn*5* biofilms vs. planktonic cells, **(B)** *PilB*::Tn*5* biofilms on Day 4 vs. *PilB*::Tn*5* on Day 1, and **(C)** *PilB*::Tn*5* planktonic cells on Day 4 vs. *PilB*::Tn*5* on Day 1. Comparison of sDEGs associated with **(D)** Day 4 and **(E)** over time are presented as described for **Figure 1D**.

downregulated sDEGs in DF2 and DF5 are enriched for genes related to defense mechanism, motility, and proteins with predicted secretion signals (**Supplementary Figures S6B,E** and **Supplementary File S3** Sheets 10 and 18), consistent with known alterations in the exoproteome of *PilB*::Tn*5* (Nagar et al., 2017; Parnasa et al., 2019; Yegorov et al., 2021).

Sets of 15 upregulated and 16 downregulated sDEGs correlate with the biofilm state (groups DF1 and DF6, respectively, in **Figure 2D, Supplementary File S3** Sheet 10). The set of downregulated sDEGs (DF6) is enriched for Cyanobase categories of energy and inorganic ion metabolism (**Supplementary Figures S5A**, **S6D**, **Supplementary File S3** Sheet 18), suggesting that biofilm cells damp down core photosynthetic, ATP synthesis, nitrate, and iron nutrient metabolisms. Upregulated sDEGs (DF1) may reflect the resource limitations and self-shaded state of the biofilm. DF1 is dominated by genes encoding hypothetical proteins (**Supplementary File S3** Sheets 10 and 18) and include: *sigD/rpoD3* (sigma factor), which is typically induced under high light in PCC 7942 (Seki et al., 2007) and is also associated with survival under nitrogen deficiency (Antal et al., 2016) and resistance to oxidative stress in PCC 6803 (Li et al., 2004); and the phycobilisome degradation protein *nblA*. Other DF1 biofilm-specific upregulated sDEGs relate to stress responses including metal limitations, such as the high light-inducible *hliA* (Ruffing, 2013) and genes encoding a putative ferric uptake regulator, a cyclic-AMP binding protein, and a cupredoxin.

Day 1 and Day 4 samples were compared to identify genes that might be involved in the process of biofilm formation. Genes in the *ebfG*-cluster and *pteB* are the most upregulated sDEGs in both *PilB*::Tn*5* biofilm and planktonic samples (**Figures 2B,C**) and are the only non-pili biofilm sDEGs over time. Although the distributed pattern of pili sDEGs over time does not highlight any specific cluster of candidates (**Figure 2E**), the set of genes that are upregulated similarly to the *ebfG*-cluster in both *PilB*::Tn*5* samples over time (group OT1) is enriched with nine genes of unknown function (**Supplementary Figure S7B**, **Supplementary File S3** Sheet 12) that may play a role in biofilm formation. Biofilm *PilB*::Tn*5* cells result in a much larger number of sDEGs (148 up; 196 down) than planktonic cells (61 up; 24 down) when compared to Day 1 samples (**Figures 2B,C**). Assessment of all over-time pairwise comparisons (**Figure 2E, Supplementary File S3** Sheet 12) highlights 86 up (group OT3) and 142 down (group OT4) biofilm-specific sDEGs, whose overlapping composition with the sDEGs of the Day 4 comparisons and an OT4-specific enrichment for energy metabolism, motility, and unknowns (**Supplementary Figure S5B**, **Supplementary File S3** Sheet 18) reinforce that the major changes in transcript abundance over time relate to adaptation to the self-shaded,

nutrient-limited biofilm condition, rather than induction of components that control or form the biofilm phenotype.

## Day 1 Transcriptomics Highlight Regulatory and CM-Responsive Genes of Interest

The four pairwise comparisons of all Day 1 samples identified nearly all known biofilm and pili genes among the highest magnitude up- and downregulated sDEGs in the comparison of *PilB*::Tn*5* with WT (**Figure 1B**). Whether in fresh (**Figure 1B**) or conditioned media (**Supplementary Figure S2**), the *pilB*::Tn*5* mutation strongly decreased structural pilin gene expression, including *pilB* and downstream genes. The Day 1 data reliably capture the known genotype–phenotype associations related to biofilm regulation, as the *ebfG* genes were the most upregulated genes in *PilB*::Tn*5* compared to WT in fresh media *ebfG* genes (**Figure 1B** and **Supplementary Figure S2**), and were strongly downregulated (**Figure 1C**) in CM. The 54 upregulated (group O1) and 38 downregulated (group O4) sDEGs that differ between *pilB*::Tn*5* and WT on Day 1 (**Figure 1D, Supplementary File S3** Sheet 8), independent of the media, are comprised of numerous Day 4 sDEGs that are associated with responses to the environment. These include upregulation of heat shock proteins Hsp20 and DnaK and an iron-stress chlorophyll-binding protein, and downregulation of heat shock protein GrpE. Some of these genes are sDEGs for WT when examined over time, suggesting that a pleiotropic effect of the *pilB*::Tn*5* mutation is early establishment of a condition that WT assumes in a more advanced culture. These sDEGs suggest that *PilB*::Tn*5* statically resembles a stationary culture, as these genes are not sDEGs when this mutant is examined over time.

To filter out the sDEGs related to this apparent constitutive-stress state, we compared the sDEGs set for *PilB*::Tn*5* vs. WT in BG-11 on Day 1, which reflects the genetic difference between the strains, against the sDEGs set for WT in BG-11 on Day 4 vs. Day 1, which presumably reflects response of the cells to aging of the culture and diminishing resources in low light-penetration conditions (**Figure 1D**, left). This comparison identified 39 (group A1) up- and 5 (group A2) downregulated sDEGs in both pairwise comparisons, which are enriched in proteins of unknown function and Phobius-predicted secretion signals, respectively (**Supplementary Figures S4A,B**, **Supplementary File S3** Sheets 14 and 18). Group A1 includes stress-related genes, such as *hsp20* and the iron-stress chlorophyll-binding protein previously noted, as well as transcription regulators such as *sigF2*, which regulates dark-induced competence (Taton et al., 2020) and is associated with pili formation and desiccation tolerance in other cyanobacteria (Srivastava et al., 2021), and a Crp/Fnr transcriptional family regulator. These sDEGs support the premise that stress management genes associated with alterations in pili and protein maintenance that occur in a dense, low-light culture are regulated in WT, but are constitutively expressed at stress-response levels at all times in *PilB*::Tn*5*.

Based on this analysis, genes in groups A1 and A2 were removed from the Day 1 comparison and four primary groups of interest were identified as being up- or downregulated in *PilB*::Tn*5* independent of the media (O-A1 and O-A4, respectively) or in fresh BG-11 media only (O-A2 and O-A3, respectively; **Figure 1D, Supplementary File S3** Sheet 16). While O-A1 is not further enriched for any particular functionality beyond that of pili genes (**Figure 1E, Supplementary Figure S4C**, **Supplementary File S3** Sheet 18), this group of 24 sDEGs includes genes encoding two response regulator proteins, a histidine kinase, a glycosyltransferase, and 13 genes of unknown function that may contribute to the regulation of biofilm formation. O-A2 is enriched in known biofilm genes, predicted secreted proteins, and includes a response regulator protein, a lipoyltransferase, and 11 genes of unknown function that may participate in biofilm production along with the EbfG proteins (**Figure 1E, Supplementary Figure S4D**, **Supplementary File S3** Sheet 16). These upregulated clusters of sDEGs highlight potential novel genes involved in biofilm formation.

Group O-A4 is enriched in known pili and biofilm genes and genes encoding proteins of unknown function, proteins involved in motility, and predicted secreted proteins (**Figure 1E, Supplementary Figure S4F**, **Supplementary File S3** Sheets 16 and 18); these properties are consistent with potential novel biofilm-repressing genes. Two genes of interest that encode proteins with cyclic di-GMP-binding domains and GAF domains, Synpcc7942_1148 and Synpcc7942_2096, may contribute to regulation of biofilm formation through cyclic di-GMP signaling, as occurs in *Synechocystis* PCC 6803 (Agostoni et al., 2016).

The pairwise comparisons pertaining to the impact of conditioned media, Day 1 in BG-11 vs. CM for both *PilB*::Tn*5* and WT, resulted in small sets of 10 sDEGs each (**Supplementary Table S6** groups IG1 and IG2), with the most prominent impact of CM being the downregulation of the *ebfG* operon genes in *PilB*::Tn*5* (**Figure 1C**), consistent with previous RT-PCR data (Parnasa et al., 2016). Most of these genes encode proteins of unknown function (**Supplementary File S3** Sheet 16).

Overall, the RNA-Seq experiment provides an overview of the pleiotropic phenotypes associated with the *PilB*::Tn*5* mutation and a large set of genes that likely reflect the impact of the environment of the biofilm, which is self-shaded and has limited gas and nutrient exchange, rather than the active process of biofilm formation. Extensive comparative analysis does identify sDEG clusters of interest that include candidate genes for biofilm regulation, formation, and responses to conditioned media, though the majority of these potential novel biofilm genes encode proteins of unknown function and must be investigated through other means.

## RB-TnSeq Screen for Loci That Affect Biofilm Formation

A complementary approach used an RB-TnSeq assay that measures the contribution of any given gene in the genome to influence the ability of a cell to proliferate in biofilms vs. planktonic growth. We screened a pooled RB-TnSeq library of approximately 150,000 individual barcoded insertions mapped to the *S. elongatus* PCC 7,942 genome that have previously shown negligible impacts from polar effects (Rubin et al., 2015). PCR amplification and high-throughput sequencing of the

barcodes quantifies changes in abundance of individual insertion mutants in the population under varying conditions, directly associating phenotypes with genotypes as has been shown for diurnal growth (Welkie et al., 2018), c-di-AMP utilization (Rubin et al., 2018), and natural competence (Taton et al., 2020). We grew replicate biological samples of the RB-TnSeq library under a variety of biofilm formation conditions, including in bubbling tubes and stationary flasks in fresh BG-11 or CM, in two separate experiments (**Figure 3A**). Conditions were similar to those used for RNA-Seq experiments except that biofilm formation was allowed to proceed for two-weeks to ensure sufficient biomass to perform genome extraction and barcode counting. Because the majority of the mutant cells in the library should still secrete and respond to the biofilm-repressing signal that is characteristic of the WT and its CM (Schatz et al., 2013), it was not expected *a priori* that the library could form biofilms. Therefore, we grew samples of a characterized biofilm forming mutant and a 1:1 mixture of WT and the biofilm-forming mutant in parallel with the library as controls. The transposon-mutant library samples, but not the WT, generated visible biofilms, although never to the extent observed for the biofilm mutant (**Figure 3B**). The 1:1 mixture of the biofilm-forming mutant and WT also produced an intermediate level of visible biofilm.

Biofilm-forming library cultures were fractionated into three samples for barcode counting: (Donlan and Costerton, 2002) planktonic cells that were removed by decanting the media, (Harrison et al., 2019) settled cells that were removed from the emptied test tubes with gentle water washes, and (DePas et al., 2014) sessile cells that required scraping to be removed from the glass culture vessels (**Figure 3A**). The planktonic sample from one replicate experiment did not produce usable sequence data. For all other experiments, the abundance of barcodes of individual transposon insertions were counted and used to calculate fitness values and statistical significance *T*-values for individual genes that reflect the abundance of those mutant genes in each fraction relative to their abundance in the starting culture (Wetmore et al., 2015). The number of genes identified with statistically significant fitness value changes, called hits, in each fraction ranged from 16 to 103 (**Supplementary Figure S1**, **Supplementary File S4** Sheet 3), much smaller than the initial DEG ranges observed for RNA-Seq, but similar to the ranges of sDEGs acquired after filtering for strong fold changes (**Supplementary Figures S1A,C**). In total, 195 genes were classified as hits in at least one of the examined samples, with 108 having absolute fitness scores greater than 1. These hit counts are both substantially reduced compared to, and have little overlap with, the sDEGs in the RNA-Seq experiment (**Supplementary Figure S10**, **Supplementary Table S5**), indicating that the RB-TnSeq data set provides a distinctly different perspective on biofilm formation than does RNA-Seq.

Although the majority of genes, 1,725 or 89.8% of the 1,920 ORFs examined, show no significant fitness changes in any

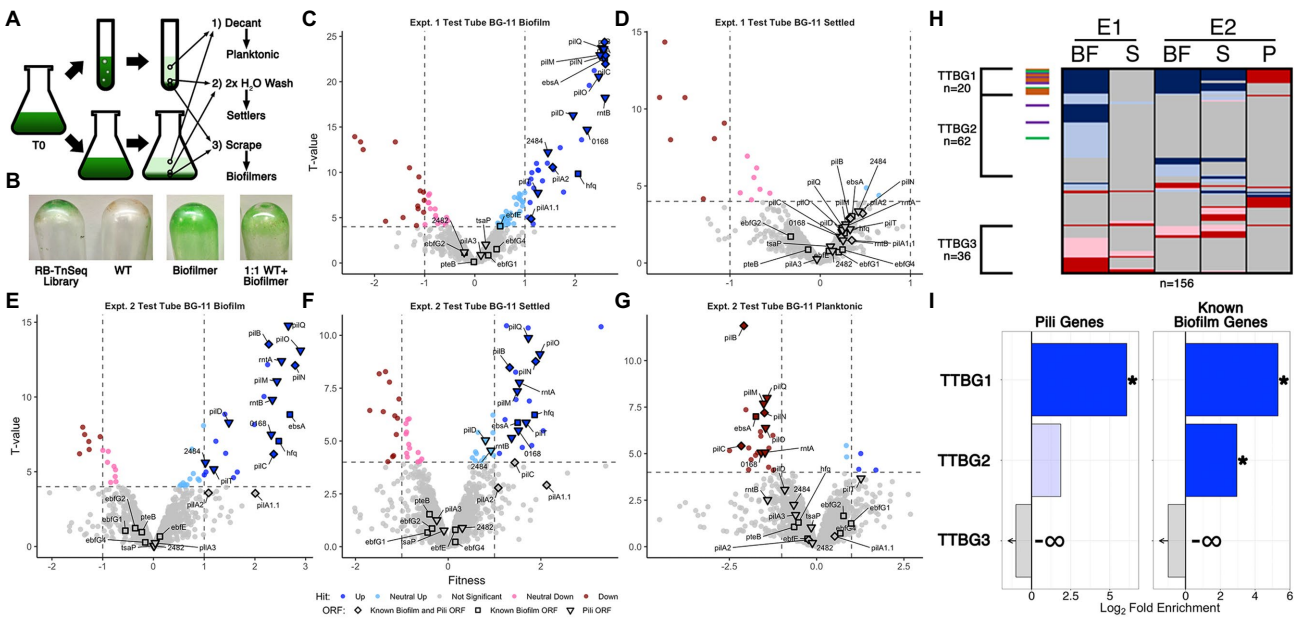



**FIGURE 3 |** RB-TnSeq analysis of biofilm forming cultures of the pooled TnSeq library. **(A)** Diagram of the RB-TnSeq library biofilm formation experiments and collection of samples. Triplicate samples were pooled prior to DNA extraction to ensure enough biomass and material for barcode sequencing. **(B)** Representative biofilms, or lack thereof, left attached to the glass after decanting and washing steps for the RB-TnSeq library, WT, a homogeneous biofilm formation strain, and a 1:1 mixture of WT and the biofilmation strain. Volcano plots of experiment 1 (E1) test tube **(C)** biofilm (BF) fractions and **(D)** settled fractions (S), and experiment 2 (E2) test tube fractions in BG-11 **(E)** biofilm, **(F)** settler, and **(G)** planktonic (P) fractions, presented as described for **Figures 1B,C** except that the *x*-axis is the fitness value and the *y*-axis is the T-value statistic. **(H)** Comparison of hit data for ORFs with significant fitness value changes, presented as described for **Figure 1D**, for all BG-11 test tube fractions. **(I)** Enrichment analyses for pili and known biofilm genes in the clusters of hits identified in **(H)** are presented as described for **Figure 1E**.

fraction examined (**Figures 3, 4, Supplementary File S4** Sheet 3), insertions in these genes were readily detected in all fractions (**Supplementary File S4** Sheet 1). These results indicate that the biofilm is composed not only of the mutants specially enriched in that fraction, but also of recruited or captured mutants that are also present in the planktonic fraction. It follows that a mutant that is under-represented in the biofilm may either be less fit to grow in the matrix or less likely to be recruited. Volcano plots demonstrate that more gene knockouts result in increased prevalence in the biofilm (positive hits) than decreased prevalence (negative hits; **Figures 3C,E, 4A,C,D**). Positive hits in the fresh media test-tube experiments are dominated by known pili and biofilm genes whose loss enable film formation, such as *pilB*, while no pili or biofilm genes are present as negative hits in the biofilm fractions. Notably, the *ebfG* operon and *pteB* genes that prevent film formation when knocked out in a *PilB*::Tn*5* background show no significant changes in fitness in any of the experiments, indicating that these mutants can be incorporated into a biofilm even when they are not producing the critical elements for biofilm formation themselves (**Supplementary File S4** Sheet 5). Analysis of the settled fractions from both test tube experiments show inconsistent results (**Figures 3D,F**). This inconsistency may be due to differences in the strength of the biofilms' attachment to the glass vessel, so that wash steps differed in dislodging some of the biofilm as part of the settler sample. Nonetheless, the depletion of pili-related gene mutants and the *ebsA* mutant, which is known to lack pili (Yegorov et al., 2021), from the planktonic fraction is consistent with the lack of pili resulting in faster rates of sedimentation (Nagar et al., 2017; **Figure 3G, Supplementary File S4** Sheet 5). Of the remaining 13 genes identified as strong negative hits (fitness < −1) in the planktonic fraction that are not known pili or biofilm associated genes (**Supplementary Table S6** group IG3, **Supplementary File S4** Sheet 8), 11 are predicted to encode proteins with secretion motifs (**Supplementary Table S6** group IG4) but were not detected previously in the exoproteomes of WT, *PilB*::Tn*5*, 1,127 Ω, or *PilB*::Tn*5* /1127Ω (Nagar et al., 2017; Parnasa et al., 2019). These genes either play some role in maintaining buoyancy or affect growth in the context of a biofilm.

Comparative analysis of all fresh media test-tube experiments revealed three clusters of genes that represent consistent biofilm formers (TTBG1), and biofilm formers (TTBG2) or biofilm-depleted mutants (TTBG3) identified in one or more but not all of the experiments (**Figure 3H**, **Supplementary File S4** Sheet 9). TTBG2 is enriched only in known biofilm genes (**Figure 3I, Supplementary Figure S8**, **Supplementary File S4** Sheet 14), and TTBG3 lacks any known pili or biofilm genes. TTBG1 includes six new loci in addition to 14 known pili and biofilm mutants. Significantly, known biofilm genes that were never identified as sDEGs in the RNA-Seq analysis are present in TTBG1 (*pilD, hfq,* and *ebsA*) and TTBG2 (*ebfE* and *pilA2*). Thus, the remaining six genes in TTBG1, which include genes encoding a sigma factor, a response regulator, and a two-component sensor histidine kinase, represent likely candidates (**Supplementary Table S6** group IG5) for genes that participate in the self-suppression of biofilm formation, of which only Synpcc7942_0464, encoding a hypothetical protein, was a sDEG in the RNA-Seq analysis. The 59 novel hits in TTBG2 (**Supplementary Table S6** group IG6, **Supplementary File S4** Sheet 9) may also participate in the self-suppression of biofilm formation, prevent sedimentation,

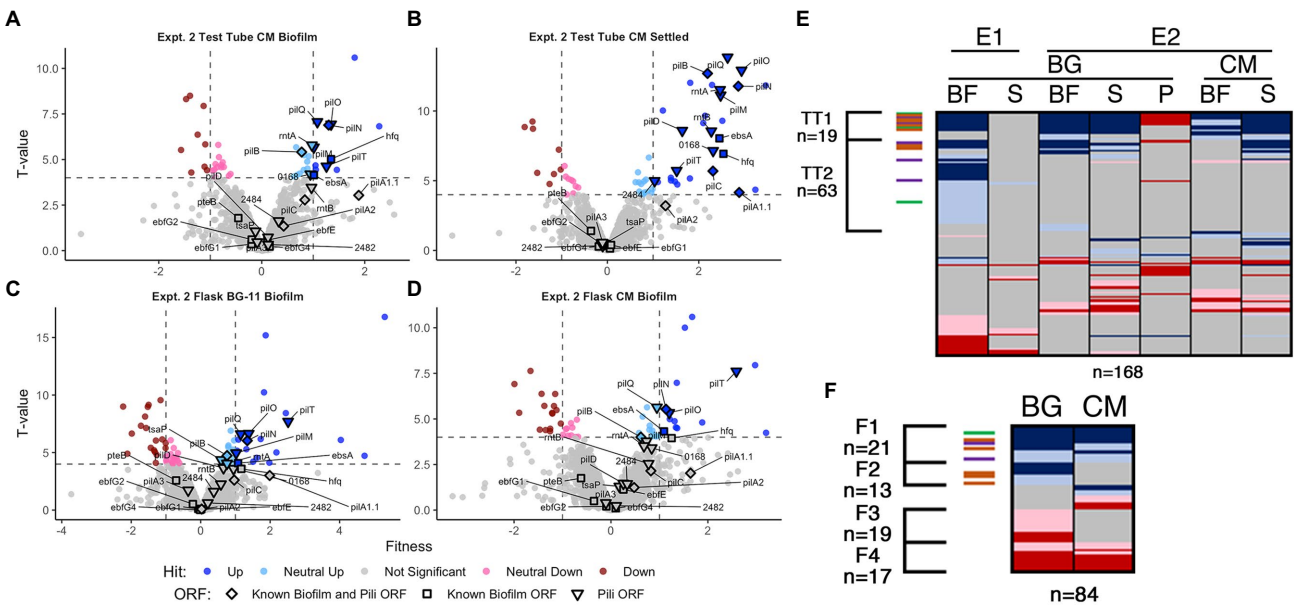



**FIGURE 4 |** Conditioned media and flask RB-TnSeq analysis Volcano plots of CM test tube **(A)** biofilm and **(B)** settled fractions, and biofilm fractions from flask experiments with **(C)** BG-11 and **(D)** CM, are presented as described for **Figures 1B,C**. Comparison of fractions associated with **(E)** all test tube experiments and **(F)** flask experiments presented as described for **Figure 1D**.

or prevent attachment to a growing biofilm because mutations that affect any of these processes would increase propensity for inclusion in the biofilm. TTBG2 is enriched for proteins conserved among cyanobacteria or involved in translation and ribosomal biogenesis (**Supplementary File S4** Sheet 14), and it is unclear how they affect biofilm formation. The 36 genes in TTBG3 (**Supplementary File S4** Sheet 9) whose mutants are absent from the biofilm are enriched for those encoding amino acid transport and metabolism, cell motility and chemotaxis, and protein maintenance genes (**Supplementary File S4** Sheet 14), consistent with phenotypes that may be needed for either recruitment into or survival in the biofilm.

In CM in test tubes, the settled fraction gave similar results to the biofilm fraction, where known biofilm and pili genes are enriched in the positive hits (**Figures 4A,B**, and **Supplementary Figure S8**). Pattern analysis across all test tube experiments identified those that are unresponsive (TT1) and responsive (TT2) to CM (**Figure 4E, Supplementary File S4** Sheet 13). The majority of the known biofilm and pili genes are found in TT1, including *pilB* whose homogenous mutant culture does not produce biofilms in CM (Schatz et al., 2013). We propose two hypotheses to explain the existence of transposon insertion mutants that are capable of producing biofilms, even in the presence of CM: (1) the self-suppressor generated by the WT-like cells in the transposon library experiments and in the mixed WT and mutant cultures was not at a high enough concentration to prevent biofilm formation at early time points, thus indicating a concentration-dependent window for enacting biofilm suppression, or (2) planktonic cells that would not initiate biofilm formation can be recruited into a mutant-seeded biofilm, which in CM would mean recruitment of CM-responsive mutant cells into a biofilm seeded by CM-unresponsive mutants. Although both hypotheses may contribute to the reduced impact of CM on heterogeneous cultures in this experiment, the latter is supported here as this analysis highlights nine novel genes identified as media-independent biofilm formers (**Supplementary Table S6** group IG7), a phenotype that was not previously encountered with *S. elongatus* biofilm mutants. This group includes a glycosyltransferase (Synpcc7942_0388) predicted to function in lipopolysaccharide production (Simkovsky et al., 2012). The identified genes are typified by the potential to impact the properties of the cell surface and thus impact biofilm formation or incorporation. Additionally, the analysis identified three consistent biofilm forming mutants that are responsive to CM, which include *sigF1*, Synpcc7942_0051, and Synpcc7942_B2646, and five mutants that biofilm only in CM, which includes a plasmid-encoded OmpR response regulator (Synpcc7942_B2647; **Supplementary Table S6** groups IG 8 and IG9, **Supplementary File S4** Sheet 13).

Many known loci that were identified in test-tube assays did not display any significant fitness value changes in flasks (**Figures 4C,D, Supplementary File S4** Sheets 5, 11, and 12). Subtle differences between the assays, including shearing stresses and increased gas exchange associated with bubbling, may alter division rates in the biofilm or the overall strength of the biofilm and affect fitness scores. Fresh BG-11 in flasks resulted in 16 novel biofilm formation mutants that did not form biofilms in BG-11 test tubes (**Supplementary Table S6** group IG10, **Supplementary File S4** Sheet 11), including *tsaP* and three mutants that had reduced prevalence in one of the test tube biofilms (**Supplementary Table S6** group IG11). Of these, five are genes involved in O-antigen production and susceptibility to grazing by amoeba, and were previously reported to result in cell aggregation, but not biofilm formation (Piechura et al., 2017; Okshevsky et al., 2018; **Supplementary Table S6** group IG12). We hypothesize that under the flask biofilm formation conditions, surface alterations that enhance aggregation or settling can more frequently enable incorporation into the heterogeneous biofilm.

Of the 13 flask biofilm forming mutants that were responsive to CM (group F2, **Figure 4F, Supplementary File S4** Sheet 12), only *pilM*, *rntA*, and *tsaP* are known pili genes and nine are specific to flask biofilm formation (**Supplementary Table S6** group IG13). Eight mutants are specifically enriched in the flask biofilm in the presence of CM that were not otherwise enriched in test tube biofilms or in fresh BG-11 in flasks (**Supplementary Table S6** group IG14), including three co-transcribed chemotaxis-like genes of the *tax2* operon (**Supplementary Table S6** group IG15), whose function is unknown except that they are not involved in phototaxis in PCC 3055 (Yang et al., 2018).

## Validation of Novel Genes Enabling or Repressing Biofilm Formation

Together, the RNA-Seq and RB-TnSeq analyses identify hundreds of potential genes of interest (**Supplementary Table S5**). To evaluate the efficacy of these two approaches toward novel target identification, we selected a subset of previously untested genes to knock out in WT or in *PilB*::Tn*5* for biofilm evaluation, prioritizing those loci for which Unigene Set (UGS) transposon insertion vectors are available (Holtman et al., 2005; Chen et al., 2012). Of the 91 mutants attempted in WT, 79 produced segregated knockout mutants of the intended gene of interest (**Supplementary Tables S7** and **S8**). Of these, we assayed candidates that were identified by: RNA-Seq only (30, 38%), RB-TnSeq only (38, 48%), or by both RNA-Seq and RB-TnSeq (11, 14%; **Supplementary Table S7**). Biological replicates of mutants were tested in bubbling test tubes, 96-well plates, and sometimes in flasks to increase the throughput of the biofilm assays and because different vessels impact the inclusion of a mutant in the biofilm. Most of the mutants created in a WT background did not form biofilms (49 of the 79, or 62%). Candidates based on RNA-Seq alone accounted for a higher false hit rate with 23 of the 30 assayed (77%) compared to 21 of the 38 assayed (55%) based on RB-TnSeq data alone.

The remaining 30 mutants formed biofilms to some degree and were classified based on the strength, reproducibility, and diversity of vessels in which biofilms were formed. Sixteen resulted in more replicates without any biofilms than with and were deemed "mostly WT-like." Of the remaining 14 mutants, most produced biofilms only sporadically or in a vessel-dependent

manner. Only four mutants consistently produced strong biofilms in multiple vessel assays—the predicted TPR-repeat lipoprotein with an O-linked N-acetylglucosamine transferase domain (Synpcc7942_0051) that was identified in the TTBG1 RB-TnSeq group; a gene predicted without experimental confirmation by Taton et al. (2020) to encode the pilus assembly protein PilP (Synpcc7942_0168), whose impact on biofilm formation was not previously known; the gene encoding the RNA polymerase sigma factor SigF (Synpcc7942_1510); and the gene encoding the pilus assembly protein PilQ (Synpcc7942_2450). Notably, all genes identified from the TTBG1 RB-TnSeq group of interest whose validated mutations produced strong biofilm phenotypes were either never identified as sDEGs in the RNA-Seq analysis or had patterns of expression that were inconsistent with previous biofilm mutants (**Supplementary Tables S6** group IG16, and **S7**).

Similarly, 13 mutations were chosen for investigation in a *PilB*::Tn*5* background and assayed for mutants that no longer produce biofilms. Although none of the mutations completely abolished biofilm formation, knockout of a deacetylase (Synpcc7942_1393; DEG in RNA-Seq group O-A4 full, **Supplementary Table S9**, **Supplementary File S3** Sheet 15) or a pimeloyl-ACP methyl ester carboxylesterase (Synpcc7942_0774) decreased the reliability or strength of biofilm formation.

## Transcriptomics and Phenomics Reveal Novel Information on the Biofilm State and the Quality of the Genome Model

The RNA-Seq time course data and enrichment analysis results highlight phenotypic changes associated with adjusting to the self-shaded, nutrient and gas exchange-reduced biofilm lifestyle, including general stress responses, changes in iron, sulfur, and nutrient transportation and metabolism, protein maintenance, and energy generation through photosynthesis (**Figure 5**). The absence of mutants affected in many of these same genes in the biofilm fractions of the RB-TnSeq experiments support the hypothesis that these functions are essential for survival in biofilms. Given the substantial change in the cells' niche as it forms biofilms, it is not surprising that a large portion of the transcriptome changes to properly adapt the cell for survival in the biofilm state, as has been observed in other systems (Nagar and Schwarz, 2015; Yan et al., 2017; Okshevsky et al., 2018).

Strand-specific transcriptional data provided here and in previous publications (Vijayan et al., 2011; Billis et al., 2014) can improve the genome models of PCC 7942 and related strains through resolving discrepancies between the observed transcriptome and the current genome annotation. The data confirmed that *pilB* is the first of four genes expressed as a single transcriptional unit (**Supplementary Figure S9A**; Vijayan et al., 2011). PilB::Tn5 transcript levels were reduced over 27 fold for the portion of *pilB* following the insertion site and the downstream genes including *pilT* and *pilC*, suggesting a polar effect. Phenotypic complementation, however, of PilB::Tn5 by introduction of *pilB in trans* (Yegorov et al., 2021) indicates sufficient expression of the downstream genes in the mutant. Transcriptional coverage of the *hfq* gene does not exceed one read count until positions annotated as codons 25 or 27, depending upon the strain analyzed, suggesting that an alternative GTG start codon previously designated as amino acid position 28 begins the open reading frame (**Supplementary Figure S9B**). This shorter annotation is consistent with the NCBI conserved domain database' identification of an Sm-like RNA binding domain from codons 35 to 95 (Marchler-Bauer et al., 2015).

## RNA-Seq Compared With RB-TnSeq

As the results for *hfq* and *ebsA* highlight, some key genes may not be identified by transcriptomic approaches because protein levels in cyanobacteria are poorly correlated with transcript levels (Xiong et al., 2015) or protein functions are regulated post-transcriptionally. In contrast, RB-TnSeq directly probes the connection between genotype and phenotype in a high-throughput manner (Wetmore et al., 2015; Price et al., 2018). In this project, we leveraged the self-fractionation of the sample that occurs through biofilm formation as an application for an RB-TnSeq library. In comparison to experiments where changes in fitness score reflect mutants' rate of division relative to the rest of the library's population in the presence of a selective condition (Wetmore et al., 2015; Price et al., 2018; Rubin et al., 2018; Welkie et al., 2018; Taton et al., 2020), this application's fitness scores are also affected by the capacity to reproducibly fractionate the cultures as well as mutation- and niche-dependent variations in rates of growth, survival, recruitment of WT-like cells into the biofilm, and settling, which is more reflective of the ability to remain planktonic rather than the capacity to generate biofilm structures.

Nonetheless, the RB-TnSeq strategy provides a more direct identification of both known and novel gene products that participate in a complex behavior such as biofilm formation, even with fewer experimental samples, than does RNA-Seq. This conclusion is based on the higher rate of validation of hits for the RB-TnSeq data than the RNA-Seq data and the lack of known biofilm regulatory genes in the transcriptomics data (**Supplementary Figures S1**, **S10**). Although the differences in incubation times (4 days vs. 2 weeks), between homogenous and heterogenous mutant cultures, and the inherent questions that these two methods address means that this comparison is a complicated one, the data supports a greater success rate for target identification using RB-TnSeq than for RNA-Seq. Because RB-TnSeq experiments require less sequence coverage than RNA-Seq samples, this strategy is also a less expensive methodology for identifying novel candidates involved in nuanced and complex behavior.

Because the WT *S. elongatus* PCC 7942 background constitutively represses biofilm formation, the experimental design would not identify mutants that are defective in components that are required for biofilm formation, such as the *ebfG-pteB* genes. These genes show the highest fold changes in transcript expression related to biofilm formation, but their inactivation in the RB-TnSeq library showed no significant changes in fitness values, likely due to the secretion of their encoded proteins by other biofilm forming mutants in the library. Future applications of Interaction RB-TnSeq (IRB-Seq), in which a second mutation (e.g., *pilB*::Tn*5*) is introduced into the RB-TnSeq library (Rubin et al., 2018) to stimulate biofilm formation, or an RB-TnSeq library constructed in a native biofilm forming strain such as UTEX 3055, is more

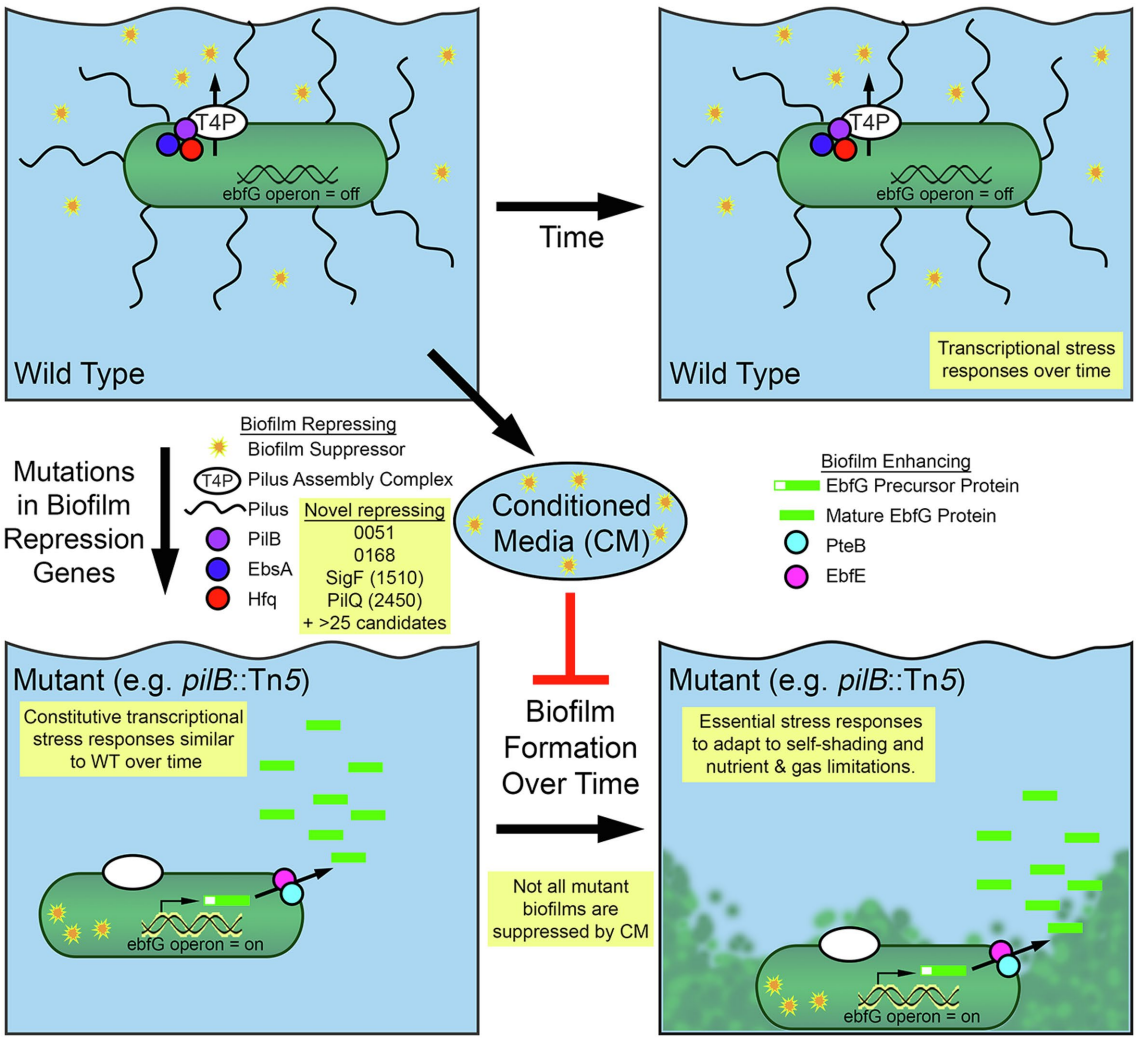



**FIGURE 5 |** Diagram of biofilm formation for *Synechococcus elongatus* PCC 7942. Self-suppression of biofilm formation in WT *S. elongatus* PCC 7942 is mediated by secretion through the type IV pilus (T4P) apparatus and its association with pili and a tripartite complex of the PilB, EbsA, and Hfq proteins. Suppressor present in conditioned media (CM) represses biofilm formation in some biofilm forming mutants. Removal of suppressing elements enables biofilm formation through the expression of the EbfG proteins and their processing and secretion, which involves PteB and EbfE. RNA-Seq and RB-TnSeq findings that add to this model are highlighted in yellow boxes.

likely to identify genes that are involved in biofilm formation process rather than the self-suppression mechanism.

## Novel Genes of Biofilm Regulation in *Synechococcus elongatus*

Of the large number of potential candidates for participating in regulation or formation of biofilms (**Supplementary Table S5**), only four resulted in strong and consistent biofilm phenotypes: Synpcc7942_0051, Synpcc7942_0168, *sigF* (Synpcc7942_1510), and *pilQ* (Synpcc7942_2450). Synpcc7942_0051 shares a domain with proteins in heterotrophic bacteria that are transcriptionally regulated by a sigma-54-interacting response regulator and a histidine kinase and participate in exopolysaccharide production and protein export in biofilm forming bacteria (Haft et al., 2006). Experiments based on the data presented here have recently demonstrated that this protein is in fact a glycosyltransferase that glycosylates PilA (Suban et al., 2022). If regulation of this protein occurs in a similar manner in PCC 7942 biofilms, then histidine kinases and response regulators that were identified as potential genes of interest (**Supplementary Table S6** group IG17) are candidates for participating in the regulatory circuit for controlling biofilm self-suppression. These validated hits, combined with the large set of genes with unknown function highlighted by the transcriptome and phenome data sets, serve as a starting point for future investigations of novel mechanisms of biofilm regulation in *S. elongatus* PCC 7942.

## DATA AVAILABILITY STATEMENT

The data presented in the study are deposited in the National Center for Biotechnology Information (NCBI) Gene

Expression Omnibus (GEO) repository, accession number GSE205445.

## AUTHOR CONTRIBUTIONS

RSi, RP, RSc, and SG contributed to the conception and design of the study. RP and RSi performed the RNA-Seq experiments. RSi and JW performed the RB-TnSeq experiments. RSi performed the data analysis of these experiments. RP, EN, EZ, SS, YY, BV, and ES performed and analyzed the mutant validation studies. RSi wrote the first draft of the manuscript. RSi, RSc, and SG edited the manuscript for submission. All authors contributed to the article and approved the submitted version.

## FUNDING

Studies were supported by the program of the National Science Foundation and the US-Israel Binational Science Foundation (NSF-BSF 2012823 to RSc and SG). This work was also supported by grants from the Israel Science Foundation (ISF 1406/14 and 2494/19) to RSc.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.899150/full#supplementary-material

**Supplementary Table S1 |** List of genes known to impact biofilm formation.

**Supplementary Table S2 |** List of pilin genes from Taton et al. (2020).

**Supplementary Table S3 |** Functional categorization, essentiality, conservation, and signal peptide prediction for ORFs in *Synechococcus elongatus* PCC 7942.

**Supplementary Table S4 |** Rockhopper predicted novel RNAs.

**Supplementary Table S5 |** Summary of hits of interest.

**Supplementary Table S6 |** Groups of genes discussed in the text not listed in a named interest group.

**Supplementary Table S7 |** Biofilm assay results for single mutants in a WT background.

**Supplementary Table S8 |** Mutations attempted in WT that did not produce segregated clones.

**Supplementary Table S9 |** Biofilm assay results for *PilB*::Tn5 double mutants.

**Supplementary Table S10 |** Vectors and associated primer pairs used in this study.

**Supplementary Figure S1 |** Comparing the results of RNA-Seq and RB-TnSeq analyses. Combined scatter and box plots of **(A)** the number of DEGs for the pairwise comparisons presented in figure or the number of hits from all RB-TnSeq fractions, **(B)** the percent of genes investigated in each experiment that are DEGs or RB-TnSeq hits, **(C)** the number of sDEGs or RB-TnSeq hits, and **(D)** the percent of genes investigated in each experiment that are sDEGs or RB-TnSeq hits. Boxed letters indicate groups that are not significantly different ($p > 0.05$) based on pairwise two-tailed t-tests assuming unequal variances; RB-TnSeq results are significantly different from those of RNA-Seq when considering all DEGs, but not when limiting the analysis to sDEGs.

**Supplementary Figure S2 |** Volcano plots for all 10 pairwise comparisons relevant to the tested RNA-Seq experimental variables. Plots are presented as described in **Figures 1B,C** and are derived from either the **(A)** DESeq analysis or **(B)** Rockhopper analysis of RNA-Seq data. For the Rockhopper analysis, the software returns FDR values of 0.0 for a number of ORFs, which become infinite values on the *y*-axis' log-scale. These infinite data points are accumulated at the top of the graph.

**Supplementary Figure S3 |** Enrichment analyses of **(A)** pili genes and **(B)** known biofilm genes in all named RNA-seq clusters of interest, as identified in **Figures 1, 2** and **Supplementary File S3**. The graphs are presented as described for **Figure 1E**.

**Supplementary Figure S4 |** Enrichment analyses of sDEGs groups, as identified in **Figure 1D**. The graphs are presented as described for **Figure 1E**, except only categories of information accumulated in **Supplementary Table S3** with at least one gene present in the interest group are shown.

**Supplementary Figure S5 |** Day 4 and time course RNA-Seq enrichment analyses. Functional enrichment analysis of Cyanobase COGs for **(A)** Day 4 and **(B)** over time groups of interest as identified in **Figures 2D,E**. Colors identify the data for the appropriate group of interest, as provided in the graph legend.

**Supplementary Figure S6 |** Enrichment analyses of all Day 4 sDEGs groups, as identified in **Figure 2D**. The graphs are presented as described for **Figure 1E**, except only categories of information accumulated in **Supplementary Table S3** with at least one gene present in the interest group are shown.

**Supplementary Figure S7 |** Enrichment analyses of all over time sDEGs groups, as identified in **Figure 2E**. The graphs are presented as described for **Figure 1E**, except only categories of information accumulated in **Supplementary Table S3** with at least one gene present in the interest group are shown.

**Supplementary Figure S8 |** Enrichment analyses of **(A)** pili genes and **(B)** known biofilm genes in all named RB-Tnseq clusters of interests, as identified in **Figures 3, 4** and **Supplementary File S4**. The graphs are presented as described for **Figure 1E**.

**Supplementary Figure S9 |** RNA-Seq Coverage of *pilB* operon and *hfq*. **(A)** Strand-specific transcriptional coverages in the region of the *PilB*::Tn*5* insertion and the downstream operon in WT and *PilB*::Tn*5* samples in fresh BG-11 on Day 1. Both coverage tracks are at the same scale, with the highest peak being approximately 1,000×. **(B)** Strand-specific transcriptional coverages in the region of *hfq* (Synpcc7942_1926) in WT and *PilB*::Tn*5* samples in fresh BG-11 on Day 1. Both coverage tracks are at the same scale, with the highest peak being approximately 70×. Transcriptional coverage starts in the middle of the canonical annotation (black arrow) but before the proposed ORF based on an alternative start codon (black brackets) based on the location of a predicted Sm-like RNA binding domain (purple).

**Supplementary Figure S10 |** Venn Diagrams of Hit Sets for RNA-Seq and RB-TnSeq data sets. **(A)** Strong DEGs (sDEGs) for RNA-Seq vs. RB-TnSeq Hits, **(B)** sDEGs for RNA-Seq vs. RB-TnSeq strong hits, **(C)** DEGs for RNA-Seq vs. RB-TnSeq hits, and **(D)** DEGs for RNA-Seq vs. RB-TnSeq strong hits. Areas are proportional to numbers provided for overlaps, RNA-Seq only DEGs, or RB-TnSeq hits.

**Supplementary File S1 |** Rockhopper RNA-Seq Analysis. Sheets include (1) alignment statistics, (2) all data output by the Rockhopper software, (3) raw counts, (4) normalized counts, (5) RPKM and expression values, (6) fold and statistical values for all pairwise comparisons of samples, (7) a summary of all DEG assignments, (8) lists of all DEGs in each pairwise comparison, (9) a

summary of DEG assignments for predicted RNAs, and (10) a summary of DEG assignments for known biofilm and pili genes.

**Supplementary File S2 |** DESeq RNA-Seq Analysis. Sheets include (1) alignment statistics, (2) raw counts, (3) normalized counts, (4) ANOVA analysis for DEGs across all samples, (5) fold and statistical values for all pairwise comparisons of samples, (6) a summary of all DEG assignments, (7) lists of all DEGs in each pairwise comparison, and (8) a summary of DEG assignments for known biofilm and pili genes.

**Supplementary File S3 |** RNA-Seq Analysis. Sheets include (1) Consistency analysis or the data for generating the union of all pairwise comparisons from Rockhopper and DESeq data, (2) a summary of all DEG assignments, (3) lists of all DEGs in each pairwise comparison, (4) a summary of sDEGs for the 10 relevant pairwise comparisons, (5) lists of sDEGs in each of the 10 relevant pairwise comparisons, (6) a summary of DEG assignments for known biofilm and pili genes, (7) comparison data of DEGs for Day 1 pairwise comparisons, (8) comparison data of sDEGs for Day 1 pairwise comparisons used to produce the middle graph of **Figure 1D**, (9) comparison data of DEGs for Day 4 pairwise comparisons, (10) comparison data of sDEGs for Day 4 pairwise comparisons used to produce **Figure 2D**, (11) comparison data of DEGs for Day 4 vs. Day 1 pairwise comparisons, (12) comparison data of sDEGs for Day 4 vs. Day 1 pairwise comparisons used to produce **Figure 2E**, (13) comparison data of DEGs for WT BG-11 Day 4 vs. Day 1 against *PilB*::Tn*5* BG-11 Day 1 vs. WT BG-11 Day1, (14) comparison data of sDEGs for WT BG-11 Day 4 vs. Day 1 against *PilB*::Tn*5* BG-11 Day 1 vs. WT BG-11 Day1 used to produce the left graph of **Figure 1D**, (15) the reduced set of DEGs for Day 1 comparison data, (16) the reduced set of sDEGs for Day 1 comparison data used to produce the right graph of **Figure 1D**, (17) enrichment analysis data for named DEG clusters of interest for all gene categories listed in **Supplementary Table S3**, and (18) enrichment analysis data for named sDEG clusters of interest for all gene categories listed in **Supplementary Table S3**.

**Supplementary File S4 |** RB-TnSeq Analysis. Sheets include (1) Gene insertion counts, (2) Fitness and *T*-values, (3) a summary of all hit assignments, (4) lists of all hits in each sample, (5) a summary of hit assignments for known biofilm and pili genes, (6) a listing of genes as per a traditional Venn diagram comparing various sample results, (7) comparison of experiment 1 samples, (8) comparison of experiment 2 BG-11 test tube samples, (9) comparison all test tube BG-11 samples, data used to produce **Figure 3H**, (10) comparison of all experiment 2 test tube samples, (11) comparison of all BG-11 biofilm samples, (12) comparison of experiment 2 flask samples, data used to produce **Figure 4F**, (13) comparison of all test tube samples, data used to produce **Figure 4E**, and (14) enrichment analysis data for named hit clusters of interest for all gene categories listed in **Supplementary Table S3**.

# REFERENCES


Adomako, M., Ernst, D., Simkovsky, R., Chao, Y. Y., Wang, J., Fang, M., et al. (2022). Comparative genomics of *Synechococcus elongatus* explains the phenotypic diversity of the strains. *mBio* e0086222. doi: 10.1128/MBIO.00862-22 [Epub ahead of print].

Agostoni, M., Waters, C. M., and Montgomery, B. L. (2016). Regulation of biofilm formation and cellular buoyancy through modulating intracellular cyclic di-GMP levels in engineered cyanobacteria. *Biotechnol. Bioeng.* 113, 311–319. doi: 10.1002/bit.25712

Allen, M. M. (1968). Simple conditions for growth of unicellular blue-green algae on plates. *J. Phycol.* 4, 1–4. doi: 10.1111/j.1529-8817.1968.tb04667.x

Allen, R., Rittmann, B. E., and Curtiss, R. (2019). Axenic biofilm formation and aggregation by *Synechocystis* sp. strain PCC 6803 are induced by changes in nutrient concentration and require cell surface structures. *Appl. Environ. Microbiol.* 85:e02192–18. doi: 10.1128/AEM.02192-18

Anders, S., Pyl, P. T., and Huber, W. (2015). HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics* 31, 166–169. doi: 10.1093/bioinformatics/btu638

Antal, T., Kurkela, J., Parikainen, M., Kårlund, A., Hakkila, K., Tyystjärvi, E., et al. (2016). Roles of group 2 sigma factors in acclimation of the Cyanobacterium *Synechocystis* sp. PCC 6803 to nitrogen deficiency. *Plant Cell Physiol.* 57, 1309–1318. doi: 10.1093/pcp/pcw079

Barros, A. C., Gonçalves, A. L., and Simões, M. (2020). "Microalgal and cyanobacterial biofilms," in *Recent Trends in Biofilm Science and Technology* (London, UK: Elsevier), 127–156.

Bendtsen, J. D., Kiemer, L., Fausbøll, A., and Brunak, S. (2005). Non-classical protein secretion in bacteria. *BMC Microbiol.* 5:58. doi: 10.1186/1471-2180-5-58

Billis, K., Billini, M., Tripp, H. J., Kyrpides, N. C., and Mavromatis, K. (2014). Comparative transcriptomics between *Synechococcus* PCC 7942 and Synechocystis PCC 6803 provide insights into mechanisms of stress acclimation. *PLoS One* 9:e109738. doi: 10.1371/journal.pone.0109738

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Bruno, L., di Pippo, F., Antonaroli, S., Gismondi, A., Valentini, C., and Albertano, P. (2012). Characterization of biofilm-forming cyanobacteria for biomass and lipid production. *J. Appl. Microbiol.* 113, 1052–1064. doi: 10.1111/j.1365-2672.2012.05416.x

Chan, S. Y., Liu, S. Y., Seng, Z., and Chua, S. L. (2021). Biofilm matrix disrupts nematode motility and predatory behavior. *ISME J.* 15, 260–269. doi: 10.1038/s41396-020-00779-9

Chen, Y., Holtman, C. K., Taton, A., and Golden, S. S. (2012). "Functional analysis of the *Synechococcus elongatus* PCC 7942 genome," in *Functional Genomics and Evolution of Photosynthetic Systems*. eds. R. Burnap and W. Vermaas (Dordrecht, Netherlands: Springer), 119–137.

Chen, I. M. A., Markowitz, V. M., Chu, K., Palaniappan, K., Szeto, E., Pillay, M., et al. (2017). IMG/M: integrated genome and metagenome comparative data analysis system. *Nucleic Acids Res.* 45, D507–D516. doi: 10.1093/nar/gkw929

Clerico, E. M., Ditty, J. L., and Golden, S. S. (2007). Specialized techniques for site-directed mutagenesis in cyanobacteria. *Methods Mol. Biol.* 362, 155–171. doi: 10.1007/978-1-59745-257-1_11

Conradi, F. D., Zhou, R. Q., Oeser, S., Schuergers, N., Wilde, A., and Mullineaux, C. W. (2019). Factors controlling floc formation and structure in the cyanobacterium synechocystis sp. strain pcc 6803. *J. Bacteriol.* 201:e00344-19. doi: 10.1128/JB.00344-19

de Carvalho, C. C. C. R. (2018). Marine biofilms: a successful microbial strategy with economic implications. *Front. Mar. Sci.* 5:126. doi: 10.3389/fmars.2018.00126

DePas, W. H., Syed, A. K., Sifuentes, M., Lee, J. S., Warshaw, D., Saggar, V., et al. (2014). Biofilm formation protects *Escherichia coli* against killing by Caenorhabditis elegans and *Myxococcus xanthus*. *Appl. Environ. Microbiol.* 80, 7079–7087. doi: 10.1128/AEM.02464-14

Dilks, K., Rose, R. W., Hartmann, E., and Pohlschröder, M. (2003). Prokaryotic utilization of the twin-arginine translocation pathway: a genomic survey. *J. Bacteriol.* 185, 1478–1483. doi: 10.1128/JB.185.4.1478-1483.2003

Donlan, R. M., and Costerton, J. W. (2002). Biofilms: survival mechanisms of clinically relevant microorganisms. *Clin. Microbiol. Rev.* 15, 167–193. doi: 10.1128/CMR.15.2.167-193.2002

Enomoto, G., Ni-Ni-Win, N. R., and Ikeuchi, M. (2015). Three cyanobacteriochromes work together to form a light color-sensitive input system for c-di-GMP signaling of cell aggregation. *Proc. Natl. Acad. Sci. U. S. A.* 112, 8082–8087. doi: 10.1073/pnas.1504228112

Farrokh, P., Sheikhpour, M., Kasaeian, A., Asadi, H., and Bavandi, R. (2019). Cyanobacteria as an eco-friendly resource for biofuel production: a critical review. *Biotechnol. Prog.* 35:e2835. doi: 10.1002/btpr.2835

Flemming, H. C., Wingender, J., Szewzyk, U., Steinberg, P., Rice, S. A., and Kjelleberg, S. (2016). Biofilms: a$n emergent form of bacterial life. *Nat. Rev. Microbiol.* 14, 563–575. doi: 10.1038/nrmicro.2016.94

Fujisawa, T., Narikawa, R., Maeda, S. I., Watanabe, S., Kanesaki, Y., Kobayashi, K., et al. (2017). CyanoBase: a large-scale update on its 20th anniversary. *Nucleic Acids Res.* 45, D551–D554. doi: 10.1093/nar/gkw1131

Golden, S. S. (2019). The international journeys and aliases of *Synechococcus elongatus*. *N. Z. J. Bot.* 57, 70–75. doi: 10.1080/0028825X.2018.1551805

Haft, D. H., Paulsen, I. T., Ward, N., and Selengut, J. D. (2006). Exopolysaccharide-associated protein sorting in environmental organisms: The PEP-CTERM/SpsH system. Application of a novel phylogenetic profiling heuristic. *BMC Biol.* 4:29. doi: 10.1186/1741-7007-4-29

Harrison, A., Hardison, R. L., Wallace, R. M., Fitch, J., Heimlich, D. R., Bryan, M. O., et al. (2019). Reprioritization of biofilm metabolism is associated with nutrient adaptation and long-term survival of *Haemophilus influenzae*. *NPJ Biofilms Microbiomes* 5, 1–14. doi: 10.1038/s41522-019-0105-6

Holtman, C. K., Chen, Y., Sandoval, P., Gonzales, A., Nalty, M. S., Thomas, T. L., et al. (2005). High-throughput functional analysis of the *Synechococcus elongatus* PCC 7942 genome. *DNA Res.* 12, 103–115. doi: 10.1093/dnares/12.2.103

Imam, S., Chen, Z., Roos, D. S., and Pohlschröder, M. (2011). Identification of surprisingly diverse type IV Pili, across a broad range of gram-positive bacteria. *PLoS One* 6:e28919. doi: 10.1371/journal.pone.0028919

Käll, L., Krogh, A., and Sonnhammer, E. L. L. (2007). Advantages of combined transmembrane topology and signal peptide prediction-the Phobius web server. *Nucleic Acids Res.* 35, W429–W432. doi: 10.1093/nar/gkm256

Kawano, Y., Saotome, T., Ochiai, Y., Katayama, M., Narikawa, R., and Ikeuchi, M. (2011). Cellulose accumulation and a cellulose synthase gene are responsible for cell aggregation in the cyanobacterium *Thermosynechococcus vulcanus* RKN. *Plant Cell Physiol.* 52, 957–966. doi: 10.1093/pcp/pcr047

Kera, K., Yoshizawa, Y., Shigehara, T., Nagayama, T., Tsujii, M., Tochigi, S., et al. (2020). Hik36–Hik43 and Rre6 act as a two-component regulatory system to control cell aggregation in *Synechocystis* sp. PCC6803. *Sci. Rep.* 10:19405. doi: 10.1038/s41598-020-76264-2

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923

Lebre, P. H., De Maayer, P., and Cowan, D. A. (2017). Xerotolerant bacteria: surviving through a dry spell. *Nat. Rev. Microbiol.* 15, 285–296. doi: 10.1038/nrmicro.2017.16

Lenhart, T. R., Duncan, K. E., Beech, I. B., Sunner, J. A., Smith, W., Bonifay, V., et al. (2014). Identification and characterization of microbial biofilm communities associated with corroded oil pipeline surfaces. *Biofouling* 30, 823–835. doi: 10.1080/08927014.2014.931379

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352

Li, H., Singh, A. K., McIntyre, L. M., and Sherman, L. A. (2004). Differential gene expression in response to hydrogen peroxide and the putative PerR regulon of *Synechocystis* sp. strain PCC 6803. *J. Bacteriol.* 186, 3331–3345. doi: 10.1128/JB.186.11.3331-3345.2004

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15:550. doi: 10.1186/s13059-014-0550-8

Marchler-Bauer, A., Derbyshire, M. K., Gonzales, N. R., Lu, S., Chitsaz, F., Geer, L. Y., et al. (2015). CDD: NCBI's conserved domain database. *Nucleic Acids Res.* 43, D222–D226. doi: 10.1093/nar/gku1221

Morin, M., Pierce, E. C., and Dutton, R. J. (2018). Changes in the genetic requirements for microbial interactions with increasing community complexity. *Elife* 7:e37072. doi: 10.7554/eLife.37072

Nagar, E., and Schwarz, R. (2015). To be or not to be planktonic? Self-inhibition of biofilm development. *Environ. Microbiol.* 17, 1477–1486. doi: 10.1111/1462-2920.12583

Nagar, E., Zilberman, S., Sendersky, E., Simkovsky, R., Shimoni, E., Gershtein, D., et al. (2017). Type 4 pili are dispensable for biofilm development in the cyanobacterium *Synechococcus elongatus*. *Environ. Microbiol.* 19, 2862–2872. doi: 10.1111/1462-2920.13814

Nozzi, N. E., Oliver, J. W. K., and Atsumi, S. (2013). Cyanobacteria as a platform for biofuel production. *Front. Bioeng. Biotechnol.* 1:7. doi: 10.3389/fbioe.2013.00007

Okshevsky, M., Louw, M. G., Lamela, E. O., Nilsson, M., Tolker-Nielsen, T., and Meyer, R. L. (2018). A transposon mutant library of *Bacillus cereus* ATCC 10987 reveals novel genes required for biofilm formation and implicates motility as an important factor for pellicle-biofilm formation. *Microbiology* 7:e00552. doi: 10.1002/mbo3.552

Parnasa, R., Nagar, E., Sendersky, E., Reich, Z., Simkovsky, R., Golden, S., et al. (2016). Small secreted proteins enable biofilm development in the cyanobacterium *Synechococcus elongatus*. *Sci. Rep.* 6:32209. doi: 10.1038/srep32209

Parnasa, R., Sendersky, E., Simkovsky, R., Waldman Ben-Asher, H., Golden, S. S., and Schwarz, R. (2019). A microcin processing peptidase-like protein of the cyanobacterium *Synechococcus elongatus* is essential for secretion of biofilm-promoting proteins. *Environ. Microbiol. Rep.* 11, 456–463. doi: 10.1111/1758-2229.12751

Petersen, T. N., Brunak, S., Von Heijne, G., and Nielsen, H. (2011). SignalP 4.0: Discriminating signal peptides from transmembrane regions. *Nat. Methods* 8, 785–786. doi: 10.1038/nmeth.1701

Piechura, J. R., Amarnath, K., and O'Shea, E. K. (2017). Natural changes in light interact with circadian regulation at promoters to control gene expression in cyanobacteria. *Elife* 6:e32032. doi: 10.7554/eLife.32032

Pierce, E. C., Morin, M., Little, J. C., Liu, R. B., Tannous, J., Keller, N. P., et al. (2021). Bacterial–fungal interactions revealed by genome-wide analysis of bacterial mutant fitness. *Nat. Microbiol.* 6, 87–102. doi: 10.1038/s41564-020-00800-z

Price, M. N., Wetmore, K. M., Waters, R. J., Callaghan, M., Ray, J., Liu, H., et al. (2018). Mutant phenotypes for thousands of bacterial genes of unknown function. *Nature* 557, 503–509. doi: 10.1038/s41586-018-0124-0

Prieto-Barajas, C. M., Valencia-Cantero, E., and Santoyo, G. (2018). Microbial mat ecosystems: structure types, functional diversity, and biotechnological application. *Electron. J. Biotechnol.* 31, 48–56. doi: 10.1016/j.ejbt.2017.11.001

Pundhir, S., and Kumar, A. (2011). SSPred: a prediction server based on SVM for the identification and classification of proteins involved in bacterial secretion systems. *Bioinformation* 6, 380–382. doi: 10.6026/97320630006380

Quinlan, A. R., and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842. doi: 10.1093/bioinformatics/btq033

Rubin, B. E., Huynh, T. A. N., Welkie, D. G., Diamond, S., Simkovsky, R., Pierce, E. C., et al. (2018). High-throughput interaction screens illuminate the role of c-di-AMP in cyanobacterial nighttime survival. *PLoS Genet.* 14:e1007301. doi: 10.1371/journal.pgen.1007301

Rubin, B. E., Wetmore, K. M., Price, M. N., Diamond, S., Shultzaberger, R. K., Lowe, L. C., et al. (2015). The essential gene set of a photosynthetic organism. *Proc. Natl. Acad. Sci. U. S. A.* 112, e6634–e6643. doi: 10.1073/pnas.1519220112

Ruffing, A. M. (2013). RNA-Seq analysis and targeted mutagenesis for improved free fatty acid production in an engineered cyanobacterium. *Biotechnol. Biofuels* 6:113. doi: 10.1186/1754-6834-6-113

Schatz, D., Nagar, E., Sendersky, E., Parnasa, R., Zilberman, S., Carmeli, S., et al. (2013). Self-suppression of biofilm formation in the cyanobacterium *Synechococcus elongatus*. *Environ. Microbiol.* 15, 1786–1794. doi: 10.1111/1462-2920.12070

Schuergers, N., Ruppert, U., Watanabe, S., Nürnberg, D. J., Lochnit, G., Dienst, D., et al. (2014). Binding of the RNA chaperone Hfq to the type IV pilus base is crucial for its function in *Synechocystis* sp. PCC 6803. *Mol. Microbiol.* 92, 840–852. doi: 10.1111/mmi.12595

Seki, A., Hanaoka, M., Akimoto, Y., Masuda, S., Iwasaki, H., and Tanaka, K. (2007). Induction of a group 2 σ factor, RPOD3, by high light and the underlying mechanism in *Synechococcus elongatus* PCC 7942. *J. Biol. Chem.* 282, 36887–36894. doi: 10.1074/jbc.M707582200

Sendersky, E., Simkovsky, R., Golden, S., and Schwarz, R. (2017). Quantification of chlorophyll as a proxy for biofilm formation in the Cyanobacterium *Synechococcus elongatus*. *Bio Protoc.* 7, 1–12. doi: 10.21769/BioProtoc.2406

Shestakov, S. V., and Khyen, N. T. (1970). Evidence for genetic transformation in blue-green alga *Anacystis nidulans*. *Mol. Gen. Genet.* 107, 372–375. doi: 10.1007/BF00441199

Simkovsky, R., Daniels, E. F., Tang, K., Huynh, S. C., Golden, S. S., and Brahamsha, B. (2012). Impairment of O-antigen production confers resistance to grazing in a model amoeba-cyanobacterium predator-prey system. *Proc. Natl. Acad. Sci. U. S. A.* 109, 16678–16683. doi: 10.1073/pnas.1214904109

Sitther, V., Tabatabai, B., Fathabad, S. G., Gichuki, S., Chen, H., and Arumanayagam, A. C. S. (2020). *Cyanobacteria as a Biofuel Source: Advances and Applications*. eds. P. K. Singh, A. Kumar, V. K. Singh and A. K. Shrivastava (Advances in Cyanobacterial Biology, Academic Press), 269–289.

Srivastava, A., Varshney, R. K., and Shukla, P. (2021). Sigma factor modulation for Cyanobacterial metabolic engineering. *Trends Microbiol.* 29, 266–277. doi: 10.1016/j.tim.2020.10.012

Stephens, M. (2017). False discovery rates: A new deal. *Biostatistics* 18, 275–294. doi: 10.1093/biostatistics/kxw041

Suban, S., Sendersky, E., Golden, S. S., and Schwarz, R. (2022). Impairment of a cyanobacterial glycosyltransferase that modifies a pilin results in biofilm development. *Environ. Microbiol. Rep.* 14, 218–229. doi: 10.1111/1758-2229.13050

Taton, A., Erikson, C., Yang, Y., Rubin, B. E., Rifkin, S. A., Golden, J. W., et al. (2020). The circadian clock and darkness control natural competence in cyanobacteria. *Nat. Commun.* 11:1688. doi: 10.1038/s41467-020-15384-9

Team R Development Core (2018). A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. vol. 2. Available at: https://www.R-project.org (Accessed January 27, 2022).

Thorvaldsdóttir, H., Robinson, J. T., and Mesirov, J. P. (2013). Integrative genomics viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.* 14, 178–192. doi: 10.1093/bib/bbs017

Tjaden, B. (2015). De novo assembly of bacterial transcriptomes from RNA-seq data. *Genome Biol.* 16:1. doi: 10.1186/s13059-014-0572-2

Tu, C. J., Shrager, J., Burnap, R. L., Postier, B. L., and Grossman, A. R. (2004). Consequences of a deletion in dspA on transcript accumulation in Synechocystis sp. strain PCC6803. *J. Bacteriol.* 186, 3889–3902. doi: 10.1128/JB.186.12.3889-3902.2004

Vijayan, V., Jain, I. H., and O'Shea, E. K. (2011). A high resolution map of a cyanobacterial transcriptome. *Genome Biol.* 12:R47. doi: 10.1186/gb-2011-12-5-r47

Welkie, D. G., Rubin, B. E., Chang, Y. G., Diamond, S., Rifkin, S. A., LiWang, A., et al. (2018). Genome-wide fitness assessment during diurnal growth reveals an expanded role of the cyanobacterial circadian clock protein KaiA. *Proc. Natl. Acad. Sci. U. S. A.* 115, E7174–E7183. doi: 10.1073/pnas.1802940115

Wetmore, K. M., Price, M. N., Waters, R. J., Lamson, J. S., He, J., Hoover, C. A., et al. (2015). Rapid quantification of mutant fitness in diverse bacteria by sequencing randomly bar-coded transposons. *MBio* 6, 1–15. doi: 10.1128/mBio.00306-15

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., et al. (2019). Welcome to the Tidyverse. *J. Open Source Softw.* 4:1686. doi: 10.21105/joss.01686

Xiong, Q., Feng, J., Li, S. T., Zhang, G. Y., Qiao, Z. X., Chen, Z., et al. (2015). Integrated transcriptomic and proteomic analysis of the global response of Synechococcus to high light stress. *Mol. Cell. Proteomics* 14, 1038–1053. doi: 10.1074/mcp.M114.046003

Yan, F., Yu, Y., Gozzi, K., Chen, Y., Guo, J. H., and Chai, Y. (2017). Genome-wide investigation of biofilm formation in *Bacillus cereus*. *Appl. Environ. Microbiol.* 83:e00561-17. doi: 10.1128/AEM.00561-17

Yang, Y., Lam, V., Adomako, M., Simkovsky, R., Jakob, A., Rockwell, N. C., et al. (2018). Phototaxis in a wild isolate of the cyanobacterium *Synechococcus elongatus*. *Proc. Natl. Acad. Sci. U. S. A.* 115, E12378–E12387. doi: 10.1073/pnas.1812871115

Yegorov, Y., Sendersky, E., Zilberman, S., Nagar, E., Waldman Ben-Asher, H., Shimoni, E., et al. (2021). A cyanobacterial component required for pilus biogenesis affects the exoproteome. *MBio* 12, 1–15. doi: 10.1128/mBio.03674-20

Zhang, N., Chang, Y. G., Tseng, R., Ovchinnikov, S., Schwarz, R., and LiWang, A. (2020). Solution NMR structure of Se0862, a highly conserved cyanobacterial protein involved in biofilm formation. *Protein Sci.* 29, 2274–2280. doi: 10.1002/pro.3952

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

*Copyright © 2022 Simkovsky, Parnasa, Wang, Nagar, Zecharia, Suban, Yegorov, Veltman, Sendersky, Schwarz and Golden. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*