THE UNIVERSITY OF CHICAGO

REGULATION OF SURFACE ATTACHMENT BY A SUITE OF TWO-COMPONENT SYSTEMS AND TRANSCRIPTION FACTORS IN *CAULOBACTER CRESCENTUS*

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES

AND THE PRITZKER SCHOOL OF MEDICINE

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

COMMITTEE ON GENETICS, GENOMICS AND SYSTEMS BIOLOGY

BY

LEILA MARIE REYES RUIZ

CHICAGO, ILLINOIS

AUGUST 2019

Table of contents

List of figures

List of tables

Acknowledgments

I wish every graduate student could have the chance to have an advisor as Sean Crosson and a mentor as Aretha Fiebig. Sean has been the most supportive mentor I could have asked for. Thank you for believing in my potential as a scientist, for giving me the opportunity to work in your lab for all these years, and for your time when I needed advice no matter how busy you were. I will always admire the good things he sees in every student and all his knowledge about almost every scientific topic. Aretha was my second mentor during graduate school. Thank you for trusting me with the *hfiA* project and for all the thoughtful training provided during these years. More than just being my mentor, thank you for all your wisdom and advice about baking, health, and just any daily life topic. Thank you Sean and Aretha for all your kindness, support, and guidance, and for being such wonderful humans. I want to thank my thesis committee, Lucia Rothman-Denes, Mike Rust, and Ed Munro, for the advice and scientific input about my thesis research project.

Success during graduate school would not have been possible without past and present members of the Crosson lab. Matthew Tien, thank you for being such an amazing bench mate and friend; I missed the daily conversations with you this past year. You have been my support and you always made me feel deserving of my position as a scientist. Julien Herrou and Jonathan Willet, thank you for being my mentors for many years, teaching me everything about molecular biology and biochemistry. Lydia Varesio, thank you for being such a supportive friend. You always try your best to make me and everyone in the lab feel happy. Ben Stein and David Hershey, I have had such a great time with

you these past years. Thank you for all the jokes, illogical conversations, advice, help, and friendship. Above all, thanks for all those times we had amazing and not so amazing food. I will miss these memories with everyone, and I wish I can come across with more people as wonderful as you all.

Thanks to my aunts that have always treated me as their own daughter. Especially Berta, who has been a support for me and my parents, and always lighted her candles to pray for my success during graduate school. Thanks to my lovely big and little cousins, because going back to Puerto Rico and spending time with you always cheered me up for the next coming months in Chicago. I want to thank Yasha, my best friend, who over the past 14 years has been supportive and genuinely cared about everything in my life. Thank you for always being there no matter the physical distance between us.

Most importantly, I want to thank my sisters and parents. I am lucky to have two sisters that have been my rock in many different ways. Shirley, my older sister, I am proud of the strength you have demonstrated throughout your life and great things will come to your life. Thank you, because you always say yes when I need your help and advice. I want to thank you, because even in the distance I always felt happy knowing that my parents were safe. Thanks for taking care of mom and dad so much, and doing everything we could not do for them in the distance. Valeria, my young sister, you have always been there for all the ups and downs. I could not be happier to have gone through the graduate school experience together with you; everything was so much easier with your guidance, help, and advice. I am proud of everything you have accomplished and I know there are

a lot of great things coming your way. Thanks for being the amazing person that you are, for being my best friend, and for all your words of wisdom during difficult times.

My parents have been role models of perseverance and kindness for me and many people. My dad has always been an amazing human, man, son, husband, and father. I have never met a person with such patience and love for everyone. He always used to tell us “Patience is a virtue”, and he is right. He only got to 7th grade, but he fought so hard for the amazing life he had. Even though he is not capable of celebrating this achievement with me, I know he was and will always be proud of what I have accomplished as a student and as a person. To my mother, the person that had the most faith in me and my biggest supporter. She is the most perseverant women I have met and an example of hard work. Coming from a time and place where education was rare, she overcame every obstacle to get her associate degree, job, own home, and family. Most importantly, she never forgot her good heart and has always been in line to helping others. Thanks for pushing us to do better in our education and in our lives. I cannot thank her enough for all the days that she stayed up late in the phone talking with me, how many times I called her for grocery shopping advice, and how many times she listened while I complained about how difficult the day was. For everything that you two have done for me, I thank you.

Abstract

Two-component systems (TCS) are broadly employed by bacteria to appropriately sense and respond to stimuli. TCS are typically thought of as insulated, linear pathways. However, emerging data provide evidence that complex cellular decisions requiring integration of multiple signals can use interconnected networks of TCS proteins. Many bacteria utilize TCS to regulate their transition to a surface-associated community lifestyle, known as a biofilm. While a surface-associated lifestyle can have advantages, shifts in the physiochemical state of the environment may result in conditions in which attachment has a negative fitness impact. Therefore, bacteria employ numerous mechanisms to control the complex surface attachment decision.

The Alphaproteobacterium *Caulobacter crescentus* secretes a polar polysaccharide adhesin known as the holdfast, which enables permanent attachment to surfaces. The small protein, HfiA, is a potent developmental inhibitor of holdfast synthesis. Multiple environmental cues influence expression of *hfiA*, but mechanisms of *hfiA* regulation remain largely undefined. The TCS LovK/LovR has previously been shown to repress *hfiA* transcription, which results in an increase in holdfast development. However, LovR lacks a DNA-binding output domain, suggesting that regulation of *hfiA* transcription by LovK/LovR is indirect. Through a forward genetic selection, I sought to identify other regulatory proteins that play a role in LovK/LovR-dependent regulation of *hfiA* transcription. I have discovered a multi-gene network encoding a suite of two-component system (TCS) proteins and transcription factors that coordinately control *hfiA* transcription, holdfast development and surface adhesion. The hybrid HWE-family

histidine kinase, SkaH, is central among these regulators and forms heteromeric complexes with the kinases, LovK and SpdS. The response regulator SpdR indirectly inhibits *hfiA* expression by activating two XRE-family transcription factors that directly bind the *hfiA* promoter to repress its transcription. My thesis provides evidence for a model in which a consortium of environmental sensors and transcriptional regulators integrate environmental cues at the *hfiA* promoter to control the attachment decision.

Chapter I: Introduction

Overview

In the environment, bacteria are exposed to a wide variety of changes in physical and chemical properties of the niches they inhabit and must be able to sense these changes to mount an appropriate response. Signal transduction systems allow bacteria to sense an extracellular or intracellular stimulus and transduce that signal through effectors to control proper physiological adaptations. In bacteria, a widely conserved signal transduction mechanism that allows for adaptation to intracellular or extracellular stimuli are two-component systems (TCS) (Fig 1.1A). This dissertation characterizes the transcriptional regulation of a critical surface adhesion regulator, *hfiA*, by a network of TCS and transcription factors in the bacterium *Caulobacter crescentus.* Because the signaling pathway I identified over the course of my thesis research is unorthodox, I will start this introduction by exposing the diversity in TCS, followed by highlighting several physiological decisions employing complex mechanisms of TCS regulation, and I will end by outlining what is known about the regulation of surface adhesion in *C. crescentus.*

Two-component systems: A conserved signal transduction system in bacteria

TCS employ phosphotransfer reactions that are directly linked to the detection of environmental signal(s). Canonical TCS are comprised of a single histidine kinase (HK) and its cognate response regulator (RR) (Fig 1.1A). In the most typical form, the HK autophosphorylates at a conserved histidine residue and this phosphoryl group is subsequently transferred to its cognate RR at a conserved aspartate residue (Stock et

al., 2000) (Fig 1.1A). Phosphorylation of the RR results in a conformational change that activates its output response (typically DNA binding) that allows for physiological adaptation to the activating signal (Stock et al., 2000). In this section, I will focus in the exquisite variability in TCS domain organization and function, which evidence the wide range of roles that TCS proteins play in adaptation to intracellular or extracellular stimuli.

Histidine kinases

The orthodox domain organization of a HK includes a sensor domain, a two-helix dimerization and histidine phosphotransfer (DHp) domain, and a catalytic and ATP binding (CA) domain (Fig 1.1A). The DHp and CA domains together will be referred to as the kinase domain.

The sensor domain is the most variable between HKs and it specifies the external or internal stimuli sensed. Some HKs display more than one sensory domain, indicating their potential ability to sense and respond to several stimuli. There is a wide variety of sensor modules and sensing mechanisms for HKs (Cheung and Hendrickson, 2010). In some HKs, the sensory region is a motif in a transmembrane loop able to bind a molecule or a protein for regulation of its activity (i.e. RegB in *Rhodobacter capsulatus*) (Swem et al., 2006). In other membrane bound HKs, a periplasmic domain serves as the sensory domain, as is the case for the PAS-like domain of PhoQ (Cheung et al., 2008) or the α-helical fold domain of NarX (Cheung and Hendrickson, 2009). Furthermore, cytoplasmic domains can link the transmembrane and kinase domains of membrane bound HKs, and can also be found in soluble (i.e. not membrane bound) HKs. Some of these cytoplasmic

domains include HAMP, PAS, and GAF (Mascher et al., 2006). HAMP domains consist of a four-helix bundle (two helices from each HK monomer, discussed below) and lies between the transmembrane and DHp domain. Several models propose that HAMP domains transmit an extracellular signal to the DHp domain by structural changes of the four-helix bundle (Jacob-Dubuisson et al., 2018). PAS domains can bind small molecules or ions to sense changes in light, oxygen, and redox potential, among other stimuli (Henry and Crosson, 2011). These domains are highly variable at the sequence level and are formed by a central β-sheet core flanked by α-helices (Jacob-Dubuisson et al., 2018). GAF domains are structurally similar to PAS domains and often found in phosphodiesterases and adenyl cyclases (Bhate et al., 2015). Even though a number of structures for sensor domains have been obtained, the variability in their ligand binding properties make the identification of the specific stimuli perceived by HKs very challenging.

Perception of a stimulus usually activates the HK, triggering autophosphorylation at the conserved histidine residue in the DHp domain. However, some HKs also have the ability to dephosphorylate their cognate RR, which is referred to as phosphatase or dephosphorylase activity (Dutta and Inouye, 1996;Hsing and Silhavy, 1997;Jiang et al., 2000b;Zhu et al., 2000;Carmany et al., 2003;Chamnongpol et al., 2003). Besides being the domain where histidine phosphorylation occurs, the DHp domain is the site of HK homodimerization through a four-helix bundle comprised of two helices from each monomer. Historically, HKs have been thought to require homodimerization to be active; an evolutionary study on the model HK, EnvZ, has identified conserved residues in the

DHp domain that determine the homodimerization of HKs (Ashenberg et al., 2011). However, examples of HKs that do not require homodimerization to be active have recently emerged. Specifically, the full-length version of the blue-light sensor HK EL346 from *Erythrobacter litoralis* HTCC2595 is active as a monomer, and the regulation of kinase activity by blue light does not require dimerization (Rivera-Cancel et al., 2014). *Brucella abortus* LOV-HK, which contains a blue-light sensory domain as well, has also been shown to be an active monomer in solution (Rinaldi et al., 2016). Furthermore, the HK ExsG from *Rhizobium* NT-26 adopts a hexameric oligomeric state (Wojnowska et al., 2013). I want to emphasize that these three non-homodimeric HKs are part of the same non-canonical family of HKs known as HisKA_2/HWE, shown to often form part of complex TCS networks and unusual oligomeric states (Herrou et al., 2017). This kinase family differs from other HK families in the region surrounding the conserved histidine phosphorylation site, in the conserved catalytic residues in the CA domain, and in inter-domain contact sites (Herrou et al., 2017).

Response regulators

The RR typically consists of a receiver (REC) domain and an output domain. Upon phosphorylation at a conserved aspartate present in the REC domain, the phosphorylated RR (RR~P) can then exert its function through its output domain (Gao et al., 2007). About 70% of output domains are DNA-binding domains, which are presumed to directly regulate gene expression (Zschiedrich et al., 2016). This reflects the very important role for transcriptional regulation during physiological adaptations. Another class of RRs,

which consists of about 8% of output domains, contains enzymatic domains (Zschiedrich et al., 2016). A major group of RRs with enzymatic domains are diguanylate cyclases (GGDEF) or phosphodiesterases (EAL, HD-GYP) (Romling et al., 2005), which synthesize, hydrolyze or simply bind c-di-GMP. Other enzymatic domains are the methylesterase domain of CheB (Djordjevic et al., 1998) and the phosphatase domain of RsbY (de Been et al., 2010). Finally, an interesting class of RRs are the single-domain RRs (SDRRs), which consist of just the REC domain, lacking any output domain as an effector. These have been shown to have roles in allosteric regulation, as in the case of CheY regulation of the FliM flagellar motor switch. Interestingly, the CheY-like proteins in *Caulobacter crescentus,* named Cle, act as RRs that are not controlled by phosphorylation but instead by binding c-di-GMP to regulate the flagellar motor and surface adhesion (Nesper et al., 2017). In addition, in *Rhizobium meliloti* and *Rhodobacter spheroids*, CheY homologues have been suggested to act as phosphate sinks to regulate phosphorelays (Sourjik and Schmitt, 1996;1998;Porter and Armitage, 2002). SDRRs have also been shown to act as hubs in complex TCS phosphorelays, as is the case for SdrG and MrrA in the regulation of the General Stress Response (GSR) in *Sphingomonas melonis* Fr1 and *C. crescentus,* respectively (Kaczmarczyk et al., 2014;Lori et al., 2018). SDRRs have been shown to also function as allosteric regulators of HKs as is the case of DivL in *C. crescentus* cell cycle regulation (Wu et al., 1998;Childers et al., 2014) and the cell cycle protease adapter CtrA (Lau et al., 2015).

Hybrid histidine kinases and histidine phosphotransferases

Beyond the canonical TCS with cognate HK and RR partners, a group of TCS named hybrid histidine kinases (HHKs) consist of the kinase and REC domain in the same polypeptide. HHKs are often involved in phosphorelays of multiple phosphotransfer events (i.e. His1→Asp1→His2→Asp2) (Fig 1.1B). In these phosphorelays, not all phosphorylation sites belong to the same protein; usually an additional SDRR and/or histidine-phosphotransferase (Hpt) are involved. Hpts are structurally similar to the DHp domain of HKs, though there are differences between the Hpts. Some Hpts are monomeric four-helix bundles that resemble a dimerized DHp domain (Kato et al., 1997;Xu and West, 1999). Other examples, e.g. Hpts that regulate sporulation in *Bacillus subtilis*, resemble the dimeric structure of a classical HK with a four-helix bundle (two from each monomer) of the DHp domain next to a pseudo CA domain that does not bind ATP (Varughese et al., 1998). In short, Hpts do not catalyze autophosphorylation reactions and typically facilitate phosphoryltransfer in a phosphorelay. Hpts often provide the His2 in the His1->Asp1->His2->Asp2 phosphorelay scheme outlined above (Fig 1.1B).

Complex TCS networks regulate critical physiological and developmental decisions in bacteria

Historically, TCS were thought to form simple linear pathways. Moreover, TCS are generally considered to be insulated from each other (i.e. there is little cross-regulation). The paradigm of insulated TCS arises from genome-wide phosphotransfer profiling of a panel of HKs against a panel of RRs, where it was shown that HKs have a strong kinetic preference for their cognate RR (Skerker et al., 2005). Furthermore, residues in the HK

and RR have been shown to mediate the specificity for phosphotransfer between cognates (Skerker et al., 2008). However, complex cellular decisions requiring integration of multiple signals may require networks of TCS proteins. Because this dissertation is focused on characterization of a complex network of TCS proteins regulating surface adhesion, in this section I will briefly highlight some well-studied examples where complex networks of TCS regulate critical physiological and developmental decisions in a variety of bacteria. In particular, I want these examples to illustrate different mechanisms employed by TCS to undergo a physiological transition. The mechanisms of regulation and physiological responses presented will be: *1)* a phosphorelay system in *Bacillus subtilis* that regulates sporulation, *2)* a spatial and temporal regulation of asymmetric cell cycle control in *Caulobacter crescentus*, *3)* a multi-kinase network regulating a single output module in the General Stress Response of *Bacillus subtilis* and *Caulobacter crescentus,* and *4)* atypical heteromeric interactions of HKs in *Pseudomonas aeruginosa* regulation of biofilm.

Sporulation

A developmental program regulated by complex signaling systems in different bacteria is endospore formation (i.e. sporulation). Spores are highly resistant to stresses like DNA-damage, high temperatures, and nutrient limitation, making it a mechanism for bacteria to escape life-threatening conditions (Hutchison et al., 2014). Therefore, it is not surprising that complex signal transduction mechanisms tightly regulate this developmental decision. A sporulation system that has been extensively studied in terms

of regulation by TCS is the Gram-positive bacterium *Bacillus subtilis* (Fig 1.2). In *B. subtilis*, the master RR Spo0A is responsible of activating and repressing transcription of genes involved in sporulation initiation (Hoch, 2017). Five HKs, KinA-KinE, have been proposed to regulate Spo0A phosphorylation (Hoch, 2017). The initially identified phosphorelay regulating Spo0A phosphorylation consists of *1)* autophosphorylation of the conserved histidine in KinA, *2)* phosphate-transfer to the aspartate in the SDRR Spo0F, *3)* phosphate-transfer to the histidine in the Hpt Spo0B and *4)* a final transfer of the phosphoryl group to Spo0A ($KinA_{His} \rightarrow Spo0F_{Asp} \rightarrow Spo0B_{His} \rightarrow Spo0A_{Asp}$) (Fig 1.2) (Burbulys et al., 1991). The specific stimuli sensed by the HKs KinA-KinE are still not well understood. Overexpression of *kinA* or *kinB* during exponential growth is sufficient to induce entry into sporulation (Fujita and Losick, 2005). KinC responds to the potassium leakage caused by surfactin exposure (Lopez et al., 2009). The HK KinD was shown to bind pyruvate in a crystal structure (Wu et al., 2013b) but it is still unknown if this is a physiologically-relevant ligand. Furthermore, the bifunctional KinD has been proposed to inhibit Spo0A phosphorylation, probably acting as a phosphatase instead of a kinase, until a specific signal that activates it as kinase is sensed (Aguilar et al., 2010). The decision to initiate sporulation in *B. subtilis* is then a mechanism regulated by a phosphorelay that receives information by a number of HKs that potentially integrate a number of environmental signals that induce sporulation.

In addition to the central sporulation phosphorelay, other regulatory proteins and phosphatases play a role in regulation of Spo0A (Fig 1.2). The protein KipI inhibits the autophosphorylation activity of KinA (Wang et al., 1997). The association of KipI with KinA

is inhibited by the anti-anti-kinase regulatory protein KipA (Wang et al., 1997). However, the mechanisms or signals that control KipI and KipA regulation of KinA are not understood. In addition, the phosphatases Rap A, E, B and H, and Spo0E control the levels of phosphorylated Spo0A, either by dephosphorylating Spo0A directly or by dephosphorylating Spo0F (Fig 1.2) (Hoch, 2017). Negative regulation of the sporulation phosphorelay is potentially a mechanism to turn off the system when needed and to prevent phosphorylation of Spo0A under non-sporulating conditions.

Cell cycle control

The *Caulobacter crescentus* cell cycle yields two distinct cell types: a flagellated swarmer cell and a sessile stalked cell (Fig 1.3) (Poindexter, 1964). These two cells progeny not only differ in their morphological and motility characteristics but also in chromosome replication and gene expression programs. The swarmer cell is arrested in G1 phase and is unable to initiate chromosome replication. This cell type transitions to a stalked cell, shedding its flagellum, retracting its type IV pili and developing a membrane extension known as the stalk (Fig 1.3). The stalked cell is in S phase, is fully competent to replicate and segregate the chromosome (Degnen and Newton, 1972), and divides asymmetrically to spawn a newborn swarmer cell (Poindexter, 1964). The newborn swarmer cell swims away and initiates the cell cycle again. Progression through the *C. crescentus* cell cycle requires coordination of both morphological and cell cycle events and this is tightly regulated by a number of TCS that are temporally and spatially coordinated (Fig 1.4).

CtrA is an essential RR with a DNA-binding domain and is the master cell cycle transcriptional regulator. CtrA is regulated at the levels of transcription, phosphorylation and proteolysis (Domian et al., 1997;Domian et al., 1999). However, in cells where *ctrA* is constitutively transcribed, the cell cycle proceeds without significant problems, suggesting that transcriptional control of *ctrA* is not the main factor of cell cycle control (Domian et al., 1997). Phosphorylated CtrA (CtrA~P) binds to the origin of replication, inhibiting chromosome replication in swarmer cells (Quon et al., 1998) and directly regulates transcription of a number of genes involved in cell morphogenesis and cell division (Laub et al., 2002).

How is CtrA differentially regulated between swarmer and stalked cells? In summary, CtrA must be stable and phosphorylated in the swarmer cell and cleared from the stalked cell (Fig 1.4). During cell division, CtrA is specifically phosphorylated in the nascent swarmer cell while it is specifically proteolyzed from the stalked cell compartment of late predivisional cells (Domian et al., 1997;Jenal and Fuchs, 1998). This temporal and spatial regulation of CtrA is centrally regulated by the bifunctional HK CckA, which is part of two interconnected phosphorelays that regulate CtrA phosphorylation and proteolysis (Collier, 2016) (Fig 1.4). In swarmer cells, CckA phosphorylates the ChpT phosphotransferase and ChpT~P can then transfer its phosphate group to two RRs: CtrA and CpdR (Fig 1.4) (Biondi et al., 2006;Iniesta et al., 2006). CtrA~P leads to inactivation of chromosome replication and phosphorylation, while CpdR~P prevents CtrA degradation by the ClpXP protease (Quon et al., 1998;Lau et al., 2015). c-di-GMP levels increase during the transition from swarmer to stalked cell and CckA switches from kinase

to phosphatase activity when it binds to c-di-GMP (Abel et al., 2013;Lori et al., 2015). CckA then dephosphorylates the RRs CpdR and CtrA, leading to CtrA proteolysis.

How is CtrA regulated in a cell compartment specific manner during cell division? Differences in polar localization of a number of TCS proteins during the late predivisional stage of the cell cycle determine the differences in CtrA phosphorylation between the swarmer and stalked cell types after cell division (Fig 1.4). DivK is a SDRR that dictates cell cycle progression and cellular asymmetry by impacting CckA's bifunctional roles. In late predivisional cells, DivJ, the primary kinase for DivK, is located at the stalked pole. On the other hand, PleC, a phosphatase for DivK~P, resides at the swarmer pole (Hecht et al., 1995;Wu et al., 1998;Wheeler and Shapiro, 1999). This opposite localization of DivJ and PleC promotes accumulation of DivK~P at the stalked cell pole and unphosphorylated DivK at the swarmer cell pole (Fig 1.4) (Jacobs et al., 2001;Lam et al., 2003). DivL is a repurposed pseudo-HK (it lacks a phosphorylated histidine) that specifically binds and stabilizes DivK~P but not unphosphorylated DivK (Tsokos et al., 2011;Childers et al., 2014). Free DivL stimulates CckA kinase activity while the DivL-DivK~P complex stimulates phosphatase activity of CckA (Fig 1.4) (Childers et al., 2014;Mann and Shapiro, 2018). The mechanisms by which DivL can differentially regulate the activities of CckA is an active area of study, but a recent study suggests that differential structural changes in DivL depend on the DivK~P bound state (Mann and Shapiro, 2018).

To summarize, spatial compartmentalization of signaling modules comprised of multiple TCS is a critical feature of the complex *C. crescentus* cell cycle (Fig 1.4). PleC

dephosphorylates DivK at the swarmer cell pole (Hecht et al., 1995;Wu et al., 1998;Wheeler and Shapiro, 1999). This ensures that DivL cannot bind DivK (Childers et al., 2014) and that DivL is free to activate the CckA kinase (Mann and Shapiro, 2018), ultimately leading to CtrA phosphorylation. On the other hand, formation of the DivL-DivK~P complex at the stalked cell pole (after phosphorylation of DivK by DivJ) is proposed to directly inhibit the CckA kinase phosphorelay and therefore CtrA phosphorylation (Tsokos et al., 2011;Childers et al., 2014;Mann and Shapiro, 2018).

General Stress Response

Unlike signaling pathways dedicated to respond to a specific stimulus, General Stress Response (GSR) systems are broadly conserved in bacteria and enable adaptation to a variety of stressful conditions. This is achieved mainly through transcriptional activation of genes involved in stress tolerance and survival (Hecker et al., 2007;Fiebig et al., 2015;Francez-Charlot et al., 2015). I will highlight the complex signal transduction systems that regulate the GSR in the Gram-positive bacterium, *B. subtilis*, and more broadly in the diverse Gram-negative class *Alphaproteobacteria.* Though the general players of the GSR in these bacteria are different, both pathways rely on the activation of an alternative sigma factor (σ) by a protein partner switch mechanism (Hecker et al., 2007;Fiebig et al., 2015;Francez-Charlot et al., 2015). Sigma factors are modular proteins required for initiation of transcription by binding to the core RNA polymerase (Davis et al., 2017).

First of all, the signal transduction system proteins regulating the GSR in *B. subtilis* are not HKs and RRs, but instead serine/threonine kinases and phosphatases. However, this system is of special interest and worth discussing in this introduction because it highlights the role of high order multi-kinase complexes that regulate a physiological response. The GSR in *B. subtilis* relies on an alternative sigma factor (σ^B) that controls transcription of genes important for stress survival (Helmann et al., 2001;Petersohn et al., 2001;Price et al., 2001). The activity of σ^B is determined by a complex signal transduction cascade that terminates in a partner-switch mechanism regulated by the anti-σ^B factor RsbW (a serine kinase) and the anti-anti-σ^B factor RsbV (Fig 1.5) (Yang et al., 1996;Vijay et al., 2000;Hecker et al., 2007). RsbW and RsbV compete for binding to σ^B. During normal growth, RsbW efficiently phosphorylates RsbV (Benson and Haldenwang, 1993). Phosphorylated RsbV cannot bind to σ^B, allowing for RsbW to bind to σ^B and prevent σ^B binding to the RNA polymerase core enzyme (Fig 1.5) (Benson and Haldenwang, 1993).

The regulatory networks upstream of the partner-switch controlling σ^B are far more complex. A wide variety of stresses regulate σ^B-dependent transcription through at least two different signal transduction pathways (Fig 1.5). The first pathway is involved in sensing intracellular energy stresses such as starvation for glucose and phosphate (Hecker et al., 2007). The main player of this pathway is RsbP, a phosphatase that dephosphorylates RsbV (Voelker et al., 1995;Vijay et al., 2000;Moore et al., 2004;Zhang and Haldenwang, 2005). Unphosphorylated RsbV can sequester RsbW to release σ^B and activate transcription of GSR genes (Fig 1.5). The second pathway regulating σ^B activity is involved in sensing environmental stresses such as exposure to heat, acid, salt or light

(Hecker et al., 2007). This pathway is of special interest because it employs a higher order complex of different serine/threonine kinases to regulate the GSR. The phosphatase RsbU is central in this pathway and it is subject of another partner-switch mechanism that involves the switch protein RsbT and its antagonists RsbS/RsbR (Yang et al., 1996). In the absence of stimuli, RsbT is sequestered in a 25S multi-kinase protein complex known as the stressosome, which consists of the antagonists RsbS, RsbR and other paralogous co-antagonists (Fig 1.5) (Chen et al., 2003;Kim et al., 2004b;Delumeau et al., 2006). Upon stimuli sensing, the kinase RsbT is activated and phosphorylates RsbR/RsbS antagonists at conserved threonine and serine residues, respectively (Kim et al., 2004a). Inactivation of RsbS after phosphorylation allows for the release of RsbT from the stressosome complex. RsbT in turns binds to and activates RsbU's phosphatase activity towards RsbV, allowing for activation of σ^B. In summary, although both pathways vary in the stimuli sensed and complexity of players involved in sensing these stimuli, they both employ the activation of phosphatases (RsbP and RsbU) to dephosphorylate RsbV, which in turn sequesters RsbW to then release σ^B and activate the GSR (Fig 1.5).

In the *Alphaproteobacteria*, regulation of the partner-switch mechanism that controls activity of the GSR alternative sigma factor (σ^{EcfG}) has been more extensively studied than the upstream players perceiving and transducing the stimuli (Fiebig et al., 2015;Francez-Charlot et al., 2015). The GSR partner-switch module in *Alphaproteobacteria* consists of σ^{EcfG}, the anti-σ^{EcfG} NepR and the anti-anti-σ^{EcfG} RR PhyR (Fig 1.6). PhyR, contains a σ-like domain and a REC domain. Under non-stressful conditions, NepR binds to σ^{EcfG} and thereby prevents binding of σ^{EcfG} to the RNA

polymerase core enzyme. Upon phosphorylation of the PhyR REC domain after a stress stimulus, NepR switches its partner and binds to the σ-like domain of PhyR. This releases σ^{EcfG} to bind the RNA polymerase core (Fig 1.6) (Francez-Charlot et al., 2009;Gourion et al., 2009;Herrou et al., 2010). Though the partner switch mechanism is well conserved and has been extensively studied in different *Alphaproteobacteria*, the stress stimuli and factors regulating the phosphorylation of PhyR are less well understood; this is currently a very active area of research. In fact, since the *Alphaproteobacteria* inhabit a wide variety of niches, it is not surprising that the stimuli to which the GSR responds and the regulatory factors employed vary between species (Fiebig et al., 2015;Francez-Charlot et al., 2015).

HKs encoded within the GSR chromosomal locus were initially clear candidates to regulate PhyR phosphorylation. Though the GSR locus is variable, it usually contains *ecfG, nepR* and a HK gene adjacent to *phyR* (Fiebig et al., 2015). In several *Alphaproteobacteria* species, more than one HK distal from the GSR locus regulates the GSR (Kaczmarczyk et al., 2014;Kim et al., 2014;Lori et al., 2018;Fiebig, 2019b). Direct phosphotransfer from these kinases to PhyR has been shown in some cases (Kaczmarczyk et al., 2014;Lori et al., 2018). In the case of *Sphingomonas melonis* Fr1 and *Caulobacter crescentus,* the cognate RRs of GSR HKs have been proposed to regulate GSR transcription by modulating phosphoryl flow to PhyR (Foreman et al., 2012;Kaczmarczyk et al., 2014;Lori et al., 2018). Though most of these HKs activate PhyR phosphorylation, some HKs regulating the GSR repress GSR transcription, suggesting that these enzymes have PhyR phosphatase activity (Kaczmarczyk et al., 2011;Kim et al., 2014). Interestingly, HisKA_2/HWE-family HKs are the main sensory

regulators of the GSR (Herrou et al., 2017). In summary, the large number of HKs and RRs that regulate the GSR in *Alphaproteobacteria* implies that these species have the capacity to integrate a wide variety of stimuli through the RR PhyR to control GSR gene transcription.

Biofilm

Bacteria are most often found in the environment living in multicellular communities known as biofilms (Costerton et al., 1995). A biofilm lifestyle can provide a number of advantages including: enhanced resistance to toxic compounds, promoting sorption of nutrients, and facilitating exchange of genes and gene products (Flemming et al., 2016). *Pseudomonas aeruginosa* is known as an opportunistic pathogen and a model for studying biofilm formation. Though a number of regulatory pathways modulate biofilm formation in *P. aeruginosa* (Lee and Yoon, 2017), I will focus on the GacS/GacA system because it provides an example of a well-studied, complex environmental response system involving unusual HKs protein interactions (Fig 1.7).

GacA is an orthodox RR with REC and DNA-binding domains and central to the pathway presented here. The regulatory pathway involving several HHKs upstream of GacA is atypical, integrating a combination of phosphorelays, phosphoryl-transfer crosstalk and atypical heteromeric interactions between the HHKs RetS, PA1611, LadS and GacS (Fig 1.7) (Willett and Crosson, 2017). This multi-TCS complex regulates GacA phosphorylation directly or indirectly, which in turns controls transcription of the small

regulatory RNAs RsmY and RsmZ (Chambonnier et al., 2016). RsmY and RsmZ are central regulators of genes involved in biofilm formation (Lapouge et al., 2008).

Only GacS has been shown to phosphorylate GacA (Goodman et al., 2009), but the other three HHKs (RetS, PA1611, LadS) are proposed to modulate GacS kinase activity directly or indirectly (Fig 1.7). RetS was initially shown to form heteromeric interactions with GacS, which led to the proposal that the role of RetS is to prevent GacS autophosphorylation (Goodman et al., 2009). Later, different phosphorylation site mutants of RetS and GacS were carefully studied *in vivo* and *in vitro* to better understand the role of RetS in this pathway. Three mechanisms for RetS-dependent inactivation of GacS were proposed: *1)* RetS removes phosphoryl groups from the autophosphorylated histidine of GacS, *2)* RetS has phosphatase activity towards the REC domain of GacS, and *3)* RetS inhibits autophosphorylation activity of GacS (Francis et al., 2018).

While RetS appears to act directly through GacS, PA1611 modulation of GacS activity is indirect. RetS-dependent inhibition of GacS has been proposed to be relieved by interactions between PA1611 and RetS (Kong et al., 2013). In other words, PA1611 antagonizes the inhibitory interaction between RetS and GacS, and thereby activates signaling through the GacS/GacA system (Fig 1.7). On the other hand, modulation of GacS/GacA by LadS is achieved through phosphoryl crosstalk between the HHKs. Specifically, GacS contains a kinase, a REC and a Hpt domain while LadS contains only kinase and REC domains. LadS is reported to transfer a phosphoryl group to GacA through the Hpt domain of GacS, comprising the following phosphorelay: $LadS_{kinase}$→ $LadS_{REC}$→$GacS_{Hpt}$→$GacA_{REC}$ (Fig 1.7) (Chambonnier et al., 2016). As with almost every

TCS, the specific molecules that activate these HHKs are not understood. However, the HHK LadS has been shown to bind calcium through its periplasmic sensor domain, which activates its signaling to the GacS/GacA cascade (Broder et al., 2016). Notably, RetS has been shown to respond to lysates of kin cells, though the exact molecule sensed is still unknown (LeRoux et al., 2015).

Finally, RetS and PA1611 are also part of a branched pathway that involves other TCS proteins: two HKs (ErcS and SagS), the histidine phosphotransferase HptB and the RR HsbR (Lin et al., 2006;Hsu et al., 2008). This TCS network modulates biofilm formation through a partner switch mechanism that terminates in regulation of a diguanylate cyclase that influences cellular c-di-GMP (Francis et al., 2017), a signaling molecule known to regulate biofilm in a number across the bacterial kingdom (Boyd and O'Toole, 2012). The integration of multiple HKs and RRs, including two branched pathways, illustrates the complexity of biofilm regulation in *P. aeruginosa* and serves as control system that can potentially integrate a number of stimuli to influence a physiological process.

Regulation of surface adhesion in *Caulobacter crescentus*: my model of study

I will now describe the current state of knowledge on the biosynthesis and regulation of the *C. crescentus* holdfast adhesin, which was the main system of study for my thesis research. As outlined above, *C. crescentus* exhibits a dimorphic cell cycle that results in two cell types after division: a flagellated swarmer cell and a sessile stalked cell (Fig 1.3) (Poindexter, 1964). At the end of the stalk, *C. crescentus* synthesizes a polar

polysaccharide adhesin known as the holdfast, that allows for cell-to-cell or surface attachment (Poindexter, 1964;Merker and Smit, 1988;Smith et al., 2003;Toh et al., 2008). Adherence to solid surfaces mediated by the holdfast is permanent (Ong et al., 1990;Li et al., 2005;Tsang et al., 2006); the holdfast is one of the strongest biological adhesives with an adhesive force in the μNewton range (Li et al., 2005;Tsang et al., 2006). Furthermore, the holdfast enables cell attachment to a chemically diverse range of materials (Berne et al., 2013), and facilitates partitioning of cells to the air-liquid interface in aqueous medium (Fiebig, 2019a). Given the permanence of holdfast synthesis, it needs to be regulated at multiple levels (discussed below) to ensure that cells are not restricted to a sub-optimal niche and promote cell dispersal.

Holdfast synthesis machinery and composition

Two main gene clusters are required for holdfast development: the holdfast synthesis (*hfs*) gene cluster and the holdfast anchoring (*hfa)* gene cluster. The holdfast synthesis genes encode enzymes involved in biosynthesis and secretion of the holdfast polysaccharide (Smith et al., 2003;Toh et al., 2008). *hfsJ* encodes a glycosyltransferase outside of the main *hfs* locus and likely catalyzes the committed step in holdfast synthesis (Fiebig et al., 2014). The holdfast anchoring genes encode proteins important for attaching the holdfast to the cell envelope and mutations in this locus cause shedding of the holdfast into the culture supernatant (Hardy et al., 2010). Though a number of enzymes involved in holdfast biosynthesis, export and anchoring have been identified,

the exact chemical composition of the holdfast is unknown, partly because of challenges in acquiring high quantities of holdfast for analysis.

Wheat-germ agglutinin (WGA), which binds to *N*-acetylglucosamine (GlcNac), specifically binds to the holdfast, indicating that the holdfast contains GlcNac (Merker and Smit, 1988). This sugar has been proposed to provide elasticity to the holdfast (Li et al., 2005). Recently, an *N*-acetyltransferase, HfsK, was identified and shown to contribute to the cohesiveness of the holdfast (Sprecher et al., 2017). Furthermore, a polysaccharide deacetylase encoded by *hfsH* affects the holdfast adhesiveness and cohesiveness (Wan et al., 2013). These findings suggest that holdfast polysaccharide may be modified to influence its adhesive and cohesive properties. A recent study aimed at identifying the composition of the holdfast polysaccharide has shown that it contains 1,4-linked backbone of glucose, mannose, N-acetyglucosamine and xylose decorated with branches at the C-6 positions of glucose and mannose (Hershey, 2019). These data provide a new framework for studying the composition, biosynthesis and regulation of the holdfast.

Cell cycle control of holdfast development

Holdfast biosynthesis is cell cycle regulated, which ensures that holdfast development is largely limited to the stalked cell (Levi and Jenal, 2006). A number of regulatory mechanisms that determine cell cycle control of holdfast synthesis have been identified. Specifically, transcription of the *hfs* and *hfa* genes is regulated as a function of the cell cycle. Transcription of a potent holdfast inhibitor – *hfiA* -- is also cell cycle

regulated in a manner that ensures the appropriate timing of holdfast development (Fiebig et al., 2014). The essential cell cycle regulators, CtrA and GcrA, occupy the *hfiA* promoter and regulate its transcription (Fiebig et al., 2014). Upon division, the daughter swarmer cell is equipped with the proteins required for holdfast synthesis, as neither RNA nor protein synthesis are required for the newborn swarmer cell to develop a holdfast (Levi and Jenal, 2006;Li et al., 2012;Ellison et al., 2017). HfiA has been proposed to serve as a checkpoint in these cells, preventing premature holdfast development by directly repressing the glycosyltransferase HfsJ required for holdfast synthesis (Fiebig et al., 2014).

Physical and environmental stimulus regulating holdfast development

Though the holdfast is cell cycle regulated, physical and environmental stimuli also serve to enhance or inhibit holdfast development. Though holdfast is typically not synthesized until the late swarmer cell stage (Levi and Jenal, 2006), if *C. crescentus* contacts a surface it elaborates a holdfast through a mechanosensing mechanism involving flagellum and pili (Li et al., 2012;Hoffman et al., 2015;Ellison et al., 2017;Hug et al., 2017). Two studies have shown that the diguanylate cyclases DgcB and PleD are also required for the mechanosensing mechanism employed by the flagellum (Hug et al., 2017;Berne et al., 2018), suggesting interplay between c-di-GMP levels and flagellum dynamics. Interestingly, the holdfast synthesis protein HfsJ binds and is activated by c-di-GMP (Hug et al., 2017). Transcription of *hfiA* also responds to changes in the nutritional environment (Fiebig et al., 2014), though the mechanism is not understood. These

findings indicate that holdfast development is cell cycle regulated but physicochemical changes in the niche encountered by *C. crescentus* overrides this regulation.

Finally, the photosensory TCS, LovK/LovR, affects the probability of holdfast development at the single cell level. *lovK/lovR* overexpression results in an increase in holdfast development and surface adhesion (Purcell et al., 2007;Fiebig et al., 2014) by repressing transcription of the holdfast inhibitor *hfiA* (Fiebig et al., 2014). LovK is a soluble, cytoplasmic photosensory HK that contains a N-terminal LOV/PAS sensor domain (Herrou and Crosson, 2011) and an C-terminal HWE/HisKA_2 family kinase domain (Karniol and Vierstra, 2004;Herrou et al., 2017). The LOV/PAS domain binds a flavin mononucleotide (FMN) cofactor (Purcell et al., 2010), which confers the ability to sense blue light and the redox state of the environment (Purcell et al., 2007;Purcell et al., 2010). *lovR* is adjacent to *lovK* on the chromosome and encodes a single domain response regulator. The regulation of *hfiA* transcription by LovK-LovR must be indirect as LovR has no DNA-binding domain.

The initial motivation of my thesis research was to understand the mechanism by which the LovK/LovR system controls *hfiA* transcription (considering that LovR is a single domain receiver that completely lacked a DNA-binding domain). The bulk of my experiments focused on defining molecular and genetic determinants of *hfiA* transcriptional control by LovK/LovR. Using a combination of genetics, genomics, biochemistry and molecular biology I have identified a network of TCS proteins and transcription factors that function in parallel or downstream of LovK/LovR to regulate *hfiA* transcription and holdfast development.

Chapter II: Identification of a network of two-component system proteins that regulate *hfiA* transcription and holdfast development in *Caulobacter crescentus*

The work presented in this chapter has been adapted from the manuscript: Regulation of bacterial surface attachment by a network of sensory transduction proteins by the authors Leila M. Reyes Ruiz, Aretha Fiebig and Sean Crosson. This manuscript was published in PLoS Genetics in May 2019 (DOI: 10.1371/journal.pgen.1008022).

Introduction

In natural ecosystems, bacteria often live in surface attached, multicellular communities known as biofilms (Costerton et al., 1995). Biofilms enhance resistance to toxic compounds, promote sorption of nutrients, and facilitate exchange of genes and gene products (Flemming et al., 2016). In aquatic systems, metabolizable substrates accumulate at surfaces (Loeb and Neihof, 1975), and thus attachment can provide a fitness advantage (Marshall, 1996). However, it is important for cells to be able to disperse in cases when surfaces accumulate toxins, or become depleted of substrates or electron acceptors. It is therefore not surprising that surface attachment is a highly regulated process in bacteria.

The alphaproteobacterium *Caulobacter crescentus* exhibits a dimorphic developmental cycle that results in two cell types after division: a flagellated swarmer cell and a sessile stalked cell (Poindexter, 1964) (Fig 1.3). The stalked cell is the reproductive cell type and divides asymmetrically, giving birth to the non-reproductive swarmer cell

(Poindexter, 1964). Swarmer cells differentiate into stalked cells, whereupon they can secrete a unipolar polysaccharide adhesin known as the holdfast (Poindexter, 1964;Merker and Smit, 1988;Smith et al., 2003;Toh et al., 2008) (Fig 1.3). Holdfast-mediated adherence to solid surfaces is permanent (Ong et al., 1990;Li et al., 2005;Tsang et al., 2006). The holdfast enables cell attachment to a variety of chemically diverse materials (Berne et al., 2013), and also facilitates partitioning of cells to the air-liquid interface in aqueous media (Fiebig, 2019a). *C. crescentus* tightly regulates holdfast synthesis, which helps to ensure that cells can take advantage of resources in particular niches, while not becoming restricted to sub-optimal environments.

The small protein HfiA is a potent inhibitor of *C. crescentus* holdfast synthesis (Fiebig et al., 2014). HfiA directly targets HfsJ, a glycolipid glycosyltransferase required to build the holdfast (Fiebig et al., 2014). The promoter of *hfiA* can integrate a number of regulatory cues that modulate its transcription and thereby affect holdfast production (Fig 2.1). Multiple cell cycle regulators bind and regulate the *hfiA* promoter (Fiebig et al., 2014). In addition, flagellum assembly indirectly affects *hfiA* transcription by an unknown mechanism (Berne et al., 2018). Environmental signals such as nutrient quality and sensory transduction proteins including the two-component system (TCS) proteins LovK-LovR (Fiebig et al., 2014) and MrrA (Lori et al., 2018) influence *hfiA* transcription by unknown and independent mechanisms. CleA, a CheY-like response regulator that tunes flagellar rotation in response to cyclic-di-GMP binding, also regulates holdfast synthesis, though it is not known if CleA-mediated holdfast regulation involves HfiA (Nesper et al., 2017). In an effort to better understand the complex regulatory network that controls

expression of the holdfast inhibitor, HfiA, I have investigated the molecular basis by which the LovK-LovR TCS controls *hfiA* transcription.

TCS are commonly used by bacteria to sense environmental signals and regulate adaptive changes in cell physiology (Stock et al., 2000). Archetypal TCS consist of a sensor histidine kinase (HK) and a DNA-binding response regulator (RR), which generally form simple linear pathways that couple environmental perturbation to a change in gene expression. There is good experimental support for the paradigm that TCS form insulated signaling systems, with little cross-regulation between non-cognate kinases and responses regulators (Skerker et al., 2005;Skerker et al., 2008;Capra and Laub, 2012;Capra et al., 2012), though there are examples of branched TCS pathways in which sensor kinases signal to more than one regulator, or in which signals from multiple kinases converge on a single regulator (reviewed in (Laub and Goulian, 2007;Willett and Crosson, 2017)). Complex cellular processes including sporulation (Jiang et al., 2000a), stress responses (Kaczmarczyk et al., 2014;Mike et al., 2014;Fiebig et al., 2015;Francez-Charlot et al., 2015;Lori et al., 2018), biofilm formation (Norsworthy and Visick, 2015), and cell cycle (Tsokos et al., 2011;Childers et al., 2014) are often controlled by multiple TCS in branched pathways. In this chapter, I have defined a regulatory system composed of multiple TCS proteins that regulates holdfast development in *C. crescentus.*

The Crosson lab has previously shown that the TCS LovK-LovR regulates holdfast development and surface adhesion by influencing the transcription of *hfiA* (Purcell et al., 2007;Fiebig et al., 2014). LovK is a soluble, cytoplasmic photosensory HK that contains a LOV (light, oxygen, voltage) sensor domain (Herrou and Crosson, 2011) N-terminal to

a HWE/HisKA_2 family (Karniol and Vierstra, 2004;Herrou et al., 2017) kinase domain. The LOV domain binds a flavin mononucleotide (FMN) cofactor (Purcell et al., 2010), which confers the ability to sense blue light and the redox state of the environment (Purcell et al., 2007;Purcell et al., 2010). LovK is encoded adjacent to LovR, a single domain response regulator (SDRR). The regulation of *hfiA* transcription by LovK-LovR must be indirect as LovR has no DNA-binding domain. Aretha Fiebig and I developed a genetic selection aimed at identifying genes that function together with or downstream of LovK-LovR to control *hfiA* transcription. I found that the hybrid histidine kinase SkaH and the canonical TCS SpdS-SpdR function with LovK-LovR to control transcription of *hfiA*. LovK and SpdS form heteromeric complexes with SkaH *in vitro* and *in vivo*, providing evidence for cross-regulation between these different TCS.

Results

A *lovK*(H180A) mutant phenocopies a *lovK-lovR* overexpression strain

Coordinate overexpression of *lovK-lovR* increases the probability that any single cell will develop a holdfast, and therefore enhances adhesion of a cell population to surfaces (Purcell et al., 2007;Fiebig et al., 2014). This phenotype is due to the repressive effect of *lovK-lovR* expression on *hfiA* transcription (Fiebig et al., 2014). During the course of this work, Aretha Fiebig discovered that mutating the conserved histidine phosphorylation site of LovK to alanine - *lovK*(H180A) - resulted in decreased *hfiA* transcription, and a concomitant increase in surface attachment and holdfast development (Fig 2.2). In other words, a strain in which the chromosomal copy of *lovK* is

replaced with *lovK*(H180A) phenocopies a *lovK-lovR* overexpression strain. To further investigate the role of the LovK-LovR TCS proteins, I expressed either the wild-type *lovK* allele or the *lovK*(H180A) allele from a plasmid in a Δ*lovK-lovR* background (Fig 2.3A). In the absence of *lovR*, expression of wild-type *lovK* had no effect on surface attachment. On the other hand, expression of *lovK*(H180A) increased surface adhesion (Fig 2.3A). These results indicate that expressing the unphosphorylatable *lovK*(H180A) allele is sufficient to enhance surface attachment, and that *lovR* is not required for this phenotype. I further tested whether hyperadhesion in the *lovK*(H180A) strain was restored to wild-type levels upon ectopic expression of the wild-type *lovK* allele. Ectopic expression of wild-type *lovK* in a strain carrying *lovK*(H180A) at the chromosomal locus reduced cell surface attachment to wild-type levels (Fig 2.3B). In the converse experiment, ectopic expression of *lovK*(H180A) in a wild-type background did not result in a hyperadhesive phenotype (Fig 2.3B). I conclude that wild-type *lovK* is dominant over *lovK*(H180A).

I chose to use the strain with the *lovK*(H180A) allele expressed from its native chromosomal locus to investigate the molecular connection between the LovK sensor kinase and *hfiA* transcription as it *1)* provided a cleaner and more stable genetic background than the *lovK-lovR* plasmid overexpression strain, *2)* expressed *lovK*(H180A) from the native *lovK* promoter, and *3)* did not require the addition of metabolizable inducers.

<u>A forward genetic selection identifies genes that function with *lovK-lovR* to repress *hfiA* expression</u>

The SDRR LovR lacks a DNA-binding output domain, and thus transcriptional regulation of *hfiA* by LovK-LovR is likely indirect. Moreover, LovR is not required to regulate *hfiA* transcription when LovK lacks its histidine phosphorylation site. We hypothesized that additional regulator(s) function either downstream or in concert with LovK-LovR to repress *hfiA* transcription. To identify such regulator(s), Aretha Fiebig and I designed a forward genetic selection to isolate mutants in which the repressive regulatory connection between LovK and *hfiA* was disrupted (Fig 2.4). Specifically, we fused the *hfiA* promoter (P_{hfiA}) to the chloramphenicol acetyltransferase (*cat*) gene and transformed the *lovK*(H180A) strain with this reporter plasmid (P_{hfiA}-*CAT*). I then used growth of this strain in the presence of chloramphenicol as a proxy for P_{hfiA} activity. By serially passaging this reporter strain, I selected for spontaneous mutants with enhanced growth in the presence of chloramphenicol (Fig 2.4). We predicted that this approach would enable us to identify mutants in which the genetic connection between *lovK* and *hfiA* transcription was disrupted.

After several passages, I isolated dozens of mutants that grew faster than the parent strain. In these isolates, I first sequenced the P_{hfiA}-*CAT* promoter fragment to identify clones with *cis* mutations in the *hfiA* promoter itself that increase promoter activity. In mutants with an intact P_{hfiA}-*CAT* promoter fragment, I assessed P_{hfiA} activity in two independent secondary screens: I measured activity of a second *hfiA* promoter fusion (P_{hfiA}-*lacZ*), and I assessed activity of the native *hfiA* promoter by measuring the surface adhesion phenotype of these mutants. Two classes of mutants emerged from this selection: *1)* strains that grew faster in chloramphenicol as a result of P_{hfiA}-*CAT* promoter

mutations, but that displayed unaltered surface adhesion and no change in P_{hfiA}-*lacZ* activity compared to the parental *lovK*(H180A) strain (Fig 2.5), and *2)* mutants with an intact P_{hfiA}-*CAT* reporter plasmid that displayed fast growth in chloramphenicol, reduced surface adhesion, and increased β-galactosidase activity from the P_{hfiA}-*lacZ* plasmid (Table 1). I mapped polymorphisms in nine spontaneous mutants from this latter class, derived from three independent selections, by whole genome sequencing.

Two mutants derived from the same flask shared a ~10kb deletion that included gene *CC_3560* (*CCNA_03675*), annotated as a hybrid histidine kinase (Fig 2.6A and Table 1). A single mutant had a 19bp deletion that resulted in a frameshift in the transmembrane histidine kinase gene *spdS (CC_0248)* (Fig 2.6A and Table 1). The remaining six mutants had lesions in either *tipR* or both *acrB2* and *tipR* (Fig 2.6A and Table 1). TipR is a repressor of the *acrAB2-nodT* operon, which encodes the components of the AcrAB2-NodT antibiotic efflux pump (Kirkpatrick and Viollier, 2014). As several of my mutants harbored lesions in *tipR*, I hypothesized that enhanced growth in chloramphenicol in these strains was due to de-repressed expression of the efflux pump. Deletion of *tipR* in either wild-type or the *lovK*(H180A) background increased growth rate in the presence of chloramphenicol (Fig 2.7). I concluded that disruption of *tipR* is a general genetic solution to combat chloramphenicol toxicity, and that these mutants were confounding the efficiency of my selection. I therefore modified the selection strategy and ectopically expressed a second copy of *tipR* from its native promoter in an effort to minimize enrichment of *tipR* mutants. Using this strategy, I identified nine mutants from three independent selections with a point mutation resulting in a *CC_3560*(H82R) allele

(Table 1). Altogether, my results provided evidence that the sensor kinases CC_3560 and SpdS play a role in *hfiA* repression by LovK-LovR. Though I likely did not saturate the selection, I decided to move forward with my initial hits of the TCS proteins CC_3560 and SpdS-SpdR.

CC_3560 (skaH), spdR and *spdS* regulate *hfiA* transcription, holdfast development, and adhesion downstream of *lovK*

The genetic selection identified mutations in two sensor kinase genes: *CC_3560* and *spdS. CC_3560* encodes a cytoplasmic hybrid histidine kinase (Fig 2.6B) that, like LovK, belongs to the non-canonical HWE/HisKA_2 family of histidine kinases (Herrou et al., 2017). I have named this kinase *skaH* (*s*ensor *k*inase *a*ssociated *h*ybrid) for reasons I describe in a later section. *spdS* is the transmembrane sensor kinase component of the SpdS-SpdR TCS, which has been reported to regulate genes involved in stationary phase adaptation in *C. crescentus* (da Silva et al., 2010;da Silva et al., 2016) (Fig 2.6B). Homologs of SpdS-SpdR in other species are more commonly known as RegB-RegA, and have been reported to detect changes in cellular redox state, and regulate a variety of energy generating and energy consuming processes (Reviewed in (Elsen et al., 2004)).

To validate the results of my genetic selection, I tested whether the TCS genes *skaH* and *spdS* are involved in regulation of *hfiA* transcription, holdfast development and surface adhesion. Though mutations in *spdR* were not identified in my selection, I also tested its role in *hfiA* regulation and adhesion since it is the cognate RR of SpdS. I first generated in-frame deletions of *skaH, spdS* or *spdR* in the *lovK*(H180A) genetic

background to determine if any of these genes were required for LovK(H180A)-dependent control of *hfiA* transcription, holdfast development and surface adhesion. Repression of *hfiA* transcription by *lovK*(H180A) required *skaH, spdR* and *spdS* (Fig 2.8A). Similarly, *skaH*, *spdR* and *spdS* are required for the increase in holdfast development and surface adhesion by *lovK*(H180A) (Fig 2.8C and Fig 2.8D). I noted that the surface adhesion phenotype of the *lovK*(H180A)Δ*spdS* strain varied depending on the growth phase of the culture. At early time points, surface adhesion of *lovK*(H180A)Δ*spdS* cultures is lower than *lovK*(H180A) cultures. When cultures begin to enter stationary phase, I observed a sharp increase in surface adhesion in the *lovK*(H180A)Δ*spdS* cultures that eventually matched the levels I observed in *lovK*(H180A) (Fig 2.8D). Importantly, the *skaH, spdR* and *spdS* deletion phenotypes I observed in the *lovK*(H180A) background were complemented by ectopic expression of these deleted genes from a xylose-inducible promoter (Fig 2.9). Together, these data provide evidence that *skaH and spdR* are required for *lovK*(H180A)-dependent regulation of *hfiA* transcription, holdfast development and surface adhesion. *lovK*(H180A)-dependent regulation of *hfiA* transcription and adhesion requires *spdS* at low culture density.

I next asked whether this set of TCS genes affect *hfiA* transcription, holdfast development and surface adhesion in wild-type cells. Activity from the P_{hfiA}-*lacZ* transcriptional reporter in Δ*skaH*, Δ*spdR* and Δ*spdS* was indistinguishable from wild type (Fig 2.8B), though transcription of *hfiA* is high under these conditions and additional increases are difficult to measure. Deletion of *spdR* (but not *skaH* or *spdS*) resulted in decreased holdfast development compared to wild type (Fig 2.8C). I further measured a

reproducible and statistically significant decrease in adhesion in Δ*skaH* and Δ*spdR* cultures (Fig 2.10), though I note that baseline surface adhesion in wild-type cultures is low under these assay conditions and additional decreases are difficult to measure (Fig 2.8D). Deletion of *spdS* in a wild-type background resulted in a similar adhesion phenotype to the *lovK*(H180A)Δ*spdS* strain, where I observed a marked increase in surface attachment relative to wild type as the culture entered stationary phase (Fig 2.8D). To better understand this Δ*spdS* hyperadhesive phenotype during stationary phase, I first tested whether the phenotype requires holdfast production. Δ*spdS* hyperadhesion requires the glycosyltransferase gene *hfsJ*, which is necessary for holdfast synthesis (Fig 2.11A). I next tested if this phenotype requires the cognate RR *spdR*. Both hyperadhesion and the increase in holdfast development during stationary phase in Δ*spdS* cells require *spdR* (Fig 2.11B and Fig 2.11C).These data provide evidence that, under these growth conditions, SkaH and SpdR play subtle roles in regulation of adhesion in wild-type cells. The role of SpdS in the regulation of holdfast development and cell adhesion depends strongly on growth phase.

Conserved SkaH and SpdR phosphorylation sites are required for *hfiA* transcriptional regulation

I next addressed the role of the conserved *skaH, spdS,* and *spdR* phosphorylation sites in regulation of *hfiA* transcription by *lovK*(H180A). The hybrid histidine kinase SkaH has a histidine phosphorylation site in its kinase domain and an aspartyl phosphorylation site in its receiver (REC) domain. I individually mutated these residues to alanine

(resulting in H285A and D550A alleles, respectively) in a *lovK*(H180A) background. The *skaH*(H285A) mutation had no effect on *lovK*(H180A) repression of *hfiA* transcription (Fig 2.8A). The *skaH*(D550A) mutation ablated *hfiA* repression by *lovK*(H180A) (Fig 2.8A) and ablated its hyperadhesive phenotype (Fig 2.12) similar to the Δ*skaH* mutation. Mutating the aspartyl phosphorylation site of the response regulator, SpdR, to alanine (D64A) also blocked *hfiA* repression by *lovK*(H180A) (Fig 2.8A) and prevented hyperadhesion (Fig 2.12) similar to the Δ*spdR* mutation. Similar to the null alleles, the H→A or D→A phosphosite mutations in *skaH* and *spdR* in a wild-type background had no detectable effect on *hfiA* promoter activity (Fig 2.8B), and a minor effect on surface adhesion (Fig 2.12). The transcription and adhesion phenotypes of *spdS*(H248A) were highly variable from day to day for reasons I do not yet understand, and thus I cannot presently draw any conclusions about this strain. Overall, I conclude that intact phosphorylation sites on the *skaH* and *spdR* REC domains are required for *lovK*(H180A) dependent regulation of *hfiA* transcription and surface adhesion.

SkaH directly interacts with LovK and SpdS *in vitro* and *in vivo*

To further explore the molecular basis of the genetic interactions I observed between TCS genes, I tested whether the corresponding TCS proteins physically interact. I first performed a bacterial two-hybrid (B2H) assay in a heterologous *E. coli* system (Karimova et al., 1998), fusing the LovK, SkaH or SpdS kinases to the C-terminus of the T25 and T18c split domain fragments of adenylate cyclase. The transmembrane domain was not included in the SpdS construct in these experiments, which contains only the

cytoplasmic kinase domain (hereafter named SpdS*). Sensor histidine kinases generally form stable homodimers (Tomomori et al., 1999;Marina et al., 2005;Jacob-Dubuisson et al., 2018). Co-expression of the same kinase fused to the T25 and T18c fragments resulted in blue colonies on X-gal medium (Fig 2.13A), indicative of a homomeric histidine kinase interaction that reconstitutes the split adenylate cyclase. Co-expression of the T25-SkaH fusion with either a T18c-LovK or T18c-SpdS* fusion also resulted in blue colonies (Fig 2.13A) indicative of heteromeric kinase interactions. B2H reporter strains co-expressing T25-LovK and T18c-SpdS* constructs yielded white colonies (Fig 2.13A). In addition, I tested whether the conserved phosphorylation sites of SkaH, LovK and SpdS were required for the heteromeric kinase interactions I observed in the B2H assay (Fig 2.14). Specifically, I assayed a number of combinations where I mutated the conserved phosphorylation site to an alanine in either one or both of the kinase fusions. All combinations of SkaH alleles co-expressed with either allele of LovK or SpdS* resulted in a positive interaction (Fig 2.14). These data provide evidence that SkaH can directly interact with both LovK and SpdS to form a heteromeric kinase complex, and that these interactions do not require the conserved protein phosphorylation sites. I further conclude that LovK and SpdS do not interact in a B2H assay.

SkaH is a hybrid histidine kinase, so I next tested whether the heteromeric B2H interactions I observed required the REC domain of SkaH. I fused a short version of SkaH that lacked the REC domain ($SkaH^{\Delta REC}$) to the T25 and T18c fragments of adenylate cyclase. Co-expression of $SkaH^{\Delta REC}$ fusions to both T25 and T18c resulted in blue colonies when grown on agar plates containing X-gal (Fig 2.13A); co-expression of

T25-SkaH$^{\Delta REC}$ and T18c-SkaH also yielded blue colonies (Fig 2.13A) providing evidence that the homomeric interaction between SkaH monomers does not require the REC domain. A B2H reporter strain co-expressing T25-SkaH$^{\Delta REC}$ and T18c-LovK yielded white colonies (Fig 2.13A). Co-expressing T25-SkaH$^{\Delta REC}$ and T18c-SpdS* yielded light blue colonies, indicating the interaction was qualitatively weaker than the observed B2H interaction between full-length SkaH and SpdS* (Fig 2.13A). These results provide evidence that the REC domain of SkaH is necessary for the heteromeric interaction with LovK observed in a B2H assay. The SkaH REC domain is not strictly required, but apparently contributes to the B2H interaction I observe between SkaH and SpdS*.

To test if the SkaH-LovK and SkaH-SpdS interactions observed in the heterologous B2H system occur *in vivo*, I tested whether LovK and/or SpdS co-purify with SkaH when affinity purified from *C. crescentus* lysate. Specifically, I ectopically expressed different combinations of histidine kinase fusions in *C. crescentus*: a maltose binding protein (MBP) N-terminally fused to SkaH (MBP-SkaH), a HA epitope tag N-terminally fused to LovK (HA-LovK), and a 3xFLAG tag C-terminally fused to full length SpdS (SpdS-3xFLAG). After co-expression in *C. crescentus*, cells were lysed and MBP-SkaH was immobilized on an amylose affinity resin in an effort to pull down potential heteromeric kinase complexes. Co-purifying proteins were detected by Western blot. When MBP-SkaH was co-expressed with either HA-LovK or SpdS-3xFLAG, both fusions co-eluted with MBP-SkaH after a stringent wash and elution with maltose (Fig 2.13B and 2.13C). As a negative control, I co-expressed MBP alone with either HA-LovK or SpdS-3xFLAG. Neither HA-LovK nor SpdS-3xFLAG co-eluted with MBP (Fig 2.13B and 2.13C),

indicating that the interactions observed were not due to protein interaction with MBP or non-specific interactions with the resin. From these data, I conclude that SkaH physically interacts with both LovK and SpdS in *C. crescentus*.

Discussion

Adherence of a *C. crescentus* cell to a surface via its holdfast is effectively permanent. As such, it is not surprising that holdfast synthesis is a highly regulated process. *hfiA* is a potent inhibitor of holdfast synthesis, and its transcription is controlled by a suite of cell cycle and environmental regulatory proteins (Fiebig et al., 2014;Berne et al., 2018;Lori et al., 2018). In this chapter, I sought to understand the molecular mechanism by which the TCS, LovK-LovR, controls *hfiA* transcription. A forward genetic selection led to the discovery of a network of TCS proteins that coordinately function with LovK to regulate *hfiA* transcription. Specifically, I have shown that the hybrid histidine kinase, SkaH, forms heteromeric complexes with the LovK kinase and the SpdS kinase *in vivo*. These data provide evidence for a multi-component sensory system that has the capacity to detect, and perhaps integrate, multiple signals to regulate holdfast development.

Signaling through the LovK-LovR system

C. crescentus lovK-lovR overexpression modulates cell adhesion by influencing *hfiA* transcription (Fiebig et al., 2014), but it also indirectly represses dozens of genes in the General Stress Response (GSR) regulon (Foreman et al., 2012) through an

independent molecular mechanism (Fiebig et al., 2014). The mechanism by which the SDRR LovR functions with LovK to regulate *hfiA* and the GSR remains unclear, but previous genetic data have suggested that its function may be to simply dephosphorylate LovK (Purcell et al., 2007;Foreman et al., 2012). This model is based on the result that overexpression of either *lovK* or *lovR* alone has no effect on cell-to-cell adhesion (Purcell et al., 2007) or on repression of GSR transcription (Foreman et al., 2012), while coordinate overexpression represses GSR transcription (Foreman et al., 2012), enhances cell-to-cell (Purcell et al., 2007) and surface adhesion (Fiebig et al., 2014), and represses *hfiA* transcription (Fiebig et al., 2014). These data support a model in which LovK is active in its unphosphorylated state. Similarly, my analysis of *hfiA* transcription and surface attachment in a strain that solely expresses the non-phosphorylatable *lovK*(H180A) allele supports such a model (Fig 2.3).

What, then, occurs at the molecular level that explains these phenotypes? Unphosphorylated-LovK may function as an allosteric regulator of other kinases or kinase complexes that control SpdR phosphorylation. LovK may also dephosphorylate a receiver protein or receiver domain required for SpdR phosphorylation. There is evidence that LovK can function as a dephosphorylase in the case of the SDRR MrrA, which was recently proposed to serve as a signaling hub that controls GSR transcription. MrrA also enhances *hfiA* expression in a manner that requires *lovK*, but not *lovR* (Lori et al., 2018). These data suggest MrrA is an additional TCS protein that functions upstream of LovK in the network of TCS proteins that modulate *hfiA* transcription. I hypothesize that non-phosphorylatable LovK(H180A) leads to high levels of phosphorylated SpdR, which

results in repression of *hfiA* transcription. Given the fact that SkaH is required for *hfiA* repression by *lovK*(H180A), I do not favor a model in which LovK directly phosphorylates or dephosphorylates SpdR.

Interplay between HisKA and HWE/HisKA_2 families of histidine kinases

SkaH and LovK belong to the HWE/HisKA_2 histidine kinase family, which is defined by unique primary structure features in the dimerization, histidine phosphotransfer (DHp), and catalytic (CA) domains (Karniol and Vierstra, 2004;Herrou et al., 2017). HWE/HisKA_2 proteins are reported to function as part of complex signal transduction systems, including the GSR in *Alphaproteobacteria* (Kaczmarczyk et al., 2014;Lori et al., 2018) and cyst development in *Rhodospirillum centenum* (He et al., 2013;He et al., 2015), and have been reported in unusual oligomeric forms including monomer (Rivera-Cancel et al., 2014) and hexamer (Wojnowska et al., 2013). Though SpdS is a member of the more common HisKA family of HKs, its homolog in *Rhodobacter capsulatus* (RegB) can form tetramers mediated by an intermolecular disulfide bond that forms through a conserved cysteine (Swem et al., 2003). It is not known if *C. crescentus* SpdS forms tetramers, but I predict that it may since it contains the conserved cysteine (C248) present in *R. capsulatus* RegB. It is therefore possible that HWE/HisKA_2 kinases (SkaH and LovK) and a HisKA kinase (SpdS) form unusual, high-order oligomers in the cell, and that changes in the oligomeric states of these proteins affect their function as regulators of *hfiA* and GSR transcription. Though I have presented clear evidence for interactions between SkaH, LovK, and SpdS, the structural basis of these interactions remains

undefined. It is important to consider the possibility that the heteromeric HK interactions I observe under my assay conditions do not reflect the full extent of kinase interactions in the cell. Specifically, it is possible that the LovK-SkaH-SpdS complex is part of an even higher order complex of regulatory proteins, as the case for the ≈2 MDa stressosome complex that controls the σ^B-dependent general stress response (GSR) in *Bacillus subtilis* (Chen et al., 2003;Kim et al., 2004b).

Finally, my data provide evidence for an emerging model in which HisKA and HWE/HisKA_2-family HKs can function together to regulate gene expression. A previous study showed that the SDRR, MrrA, can be phosphorylated and/or dephosphorylated by kinases from both of these families of HKs in *C. crescentus* (Lori et al., 2018). Furthermore, deletion of all the HWE/HisKA_2 HKs in *Sphingomonas melonis* Fr1 does not completely ablate GSR transcription (Kaczmarczyk et al., 2014), suggesting that one or more HisKA kinases also play a role in regulating this system. This chapter provides additional evidence for complex signal transduction systems involving both HisKA and HWE/HisKA_2 histidine kinase families.

A consortium of sensors regulate *hfiA* to control cell adhesion

Crystal violet staining data reveal a modest but reproducible defect in surface adhesion when *skaH* or *spdR* are deleted in a wild-type background (Fig 2.8D and Fig 2.10). This defect is more pronounced as culture density increases (Fig 2.8D). I note that small differences in surface adhesion are challenging to measure. Nevertheless, my results are consistent with a recent genome-scale analysis of adhesion to a cellulosic

surface where mutants with disruptions in *skaH*, *spdS* and *spdR* exhibited reduced cell attachment to cheesecloth in complex medium (Fig 2.15) (Hershey et al., 2019). This independent genetic experiment supports a model in which these gene products function together as positive regulators of surface attachment across distinct media conditions.

The specific environmental or intracellular signals that regulate activity of the sensory systems identified in my selection are not known. However, SpdS-SpdR homologs have been studied extensively and shown to sense the cellular redox state by three mechanisms: *1)* binding of oxidized quinone (Swem et al., 2006;Wu and Bauer, 2010), *2)* cysteine sulfenic acid modification (Wu et al., 2013a) or *3)* disulfide bond formation (Swem et al., 2003) at a conserved reactive cysteine. *C. crescentus* SpdS transfers a phosphoryl group to SpdR *in vitro* (da Silva et al., 2010), but the role of SpdS-SpdR phosphorylation *in vivo* remains unclear. In other *Alphaproteobacteria*, SpdS homologs have been reported to dephosphorylate their cognate regulators (Bird et al., 1999;Emmerich et al., 1999;Comolli et al., 2002;Potter et al., 2002), and thus I cannot exclude the possibility that SpdS may primarily function to dephosphorylate SpdR in *C. crescentus* cells. In fact, I observed that upon entering stationary phase, a Δ*spdS* mutant has dramatically enhanced surface adhesion (Fig 2.8D), similar to the hyper-adhesive phosphomimetic *spdR*(D64E) mutant. Furthermore, I present evidence that the hyper-adhesive phenotype of the Δ*spdS* mutant depends on *spdR* (Fig 2.11B and Fig 2.11C). This stationary phase ∆*spdS* phenotype suggests a model in which stationary phase signal(s) activate SpdS as a SpdR~P dephosphorylase. This model remains to be tested.

C. crescentus LovK has the capacity to function as both a redox sensor (Purcell et al., 2010) and a photosensor (Purcell et al., 2007). The genetic interactions I uncovered between LovK and SpdS-SpdR may reflect a fundamental role for LovK in redox sensing *in vivo*. *C. crescentus* is reported to undergo a dynamic change in the intracellular redox state during the cell cycle (Narayanan et al., 2015). It would be interesting to know if *hfiA* transcriptional regulation by the SkaH, LovK and SpdS sensor kinases is influenced by intracellular redox. The ability of SkaH to function as a sensor has not been characterized, though a database search using the Phyre2 profile-profile alignment algorithm (Kelley et al., 2015) provides evidence for two PAS-like sensor domains (Henry and Crosson, 2011) N-terminal to the HWE kinase domain. The multi-protein regulatory system I report here could potentially sense a variety of environmental signals (e.g. redox potential, light, small molecules) to control *hfiA* expression.

Chapter III: The response regulator SpdR regulates *hfiA* transcription and holdfast development through two downstream transcription factors

Some of the work presented in this chapter has been adapted from the manuscript: Regulation of bacterial surface attachment by a network of sensory transduction proteins by the authors Leila M. Reyes Ruiz, Aretha Fiebig and Sean Crosson. This manuscript was published in PLoS Genetics in May 2019 (DOI: 10.1371/journal.pgen.1008022). RtrA and RtrB transcriptomic data presented in this chapter are unpublished.

Introduction

In the previous chapter, I described a network of TCS proteins that regulate *hfiA* transcription and holdfast development. A question that remained open was to identify a direct regulator of *hfiA* transcription that functioned downstream from the sensor histidine kinases that I described in Chapter II. The RR SpdR contains a DNA-binding domain and was a clear candidate to function as a direct regulator of *hfiA* transcription. Experiments described here in Chapter III were aimed at defining the role of SpdR in *hfiA* transcriptional regulation. Using genetics, genomics and molecular biology approaches, I discovered that SpdR indirectly regulates *hfiA* transcription. Next, using a candidate gene approach, I identified two XRE-family transcription factors, *rtrA* and *rtrB,* that function downstream of SpdR to directly regulate *hfiA* transcription. I further conducted RNA-seq experiments to define the regulons of RtrA and RtrB. These transcriptomic data show that the RtrA

and RtrB regulons both include *hfiA* and are highly overlapped, suggesting that they serve redundant regulatory functions in the cell.

Results

spdR regulates *hfiA* transcription indirectly

A major goal of my genetic selection was to identify direct regulators of *hfiA* transcription downstream of LovK. The DNA-binding response regulator, SpdR, emerged as a clear candidate from my data. I utilized genetic, molecular and genomic approaches to test the role of SpdR as a regulator of *hfiA* transcription. As described in Chapter II, alanine substitution of the conserved phosphorylation site in the SpdR REC domain (D64A), phenocopied the Δ*spdR* deletion mutation in a *lovK*(H180A) background (Fig 2.8A). Moreover, a classic phosphomimetic substitution of aspartate to glutamate (D64E) in a wild-type background decreased activity from the P_{hfiA}-*lacZ* transcriptional reporter, and increased surface adhesion and holdfast development (Fig 3.1A). These data provide genetic evidence that SpdR can function to repress *hfiA* transcription and modulate holdfast production and cell surface adhesion in a phosphorylation dependent manner.

To test whether SpdR is a direct regulator of *hfiA*, I performed an electrophoretic mobility shift assay (EMSA) using a fluorescently labeled *hfiA* promoter probe and His_6-SpdR(D64E) recombinant protein. Increasing concentrations of His_6-SpdR(D64E) did not retard gel mobility of the *hfiA* probe (Fig 3.1B). As a positive control, I performed an EMSA using a fluorescently labeled probe corresponding to the promoter of *cspD*, a gene

reported to be directly regulated by SpdR (da Silva et al., 2010). Migration of the labeled *cspD* promoter probe was retarded by increasing concentrations of His_6-SpdR(D64E) and this probe displayed the same double shift as previously published (Fig 3.1B) (da Silva et al., 2010). These results provide evidence that SpdR does not bind the *hfiA* promoter, and therefore support a model in which SpdR indirectly regulates *hfiA* transcription.

RNA sequencing analysis defines the SpdR regulon

Given the evidence that SpdR regulates *hfiA* transcription indirectly, I experimentally defined the transcriptional regulon of SpdR with the goal of uncovering direct *hfiA* regulators that are expressed downstream of SpdR. I measured steady-state transcript levels in a strain missing *spdR* (Δ*spdR*) and a strain expressing the active phosphomimetic *spdR*(D64E) allele by RNA sequencing (RNA-seq). The majority of genes with differential transcriptional profiles are more highly expressed in cells expressing the constitutively active *spdR*(D64E) allele (Fig 3.2A and Table 3.1). Importantly, *hfiA* is among the small set of transcripts that are downregulated in the *spdR*(D64E) strain relative to the Δ*spdR* strain (5.1-fold, $p < e^{-16}$) (Fig 3.2A and Table 3.1). A consensus binding motif for SpdR has been defined in *C. crescentus* (da Silva et al., 2010;da Silva et al., 2016) and for SpdR homologs in other species (Emmerich et al., 2000;Swem et al., 2001;Laguri et al., 2003). To predict direct targets of SpdR, I utilized the MEME motif discovery platform (Bailey et al., 2009) to identify shared sequence motifs in the promoter regions (-150 to +50) of all the genes with at least 2-fold expression differences between *spdR*(D64E) and Δ*spdR* strains. This approach revealed a predicted

SpdR binding site (Fig 3.2B) in the promoter regions of 82 genes (Table 3.1), suggesting that these genes are directly regulated by SpdR. A scan of the region from -500 bp to +100 bp around the start *hfiA* codon failed to identify a SpdR binding sequence. These data provide additional evidence that SpdR is not a direct repressor of *hfiA* transcription.

An approach to identify direct *hfiA* transcriptional regulators downstream of SpdR

I hypothesized that transcription factor(s) regulated downstream of SpdR were responsible for repression of *hfiA* transcription. I took a candidate approach and examined all of the differentially expressed genes (>2-fold, P < 0.01) in my RNA-Seq dataset that contained a predicted DNA-binding domain. The six predicted transcription factors in this set were activated by SpdR (Table 3.1). Thus, I reasoned that the repressive effect of SpdR on *hfiA* transcription was the result of activation of one or more transcription factors that directly repress *hfiA* (Fig 3.3A). I individually overexpressed all candidate transcription factors and measured *hfiA* promoter activity and surface attachment. Two *X*enobiotic *R*esponse *E*lement (XRE)-family transcription factors, *CC_3164* (*CCNA_03267*) and *CC_2330* (*CCNA_02415*), displayed the overexpression phenotypes expected for an *hfiA* repressor: a decrease in transcription from the *hfiA* promoter (Fig 3.3B), an increase in holdfast development (Fig 3.3C), and an increase in surface adhesion (Fig 3.3D), compared to wild-type. Importantly, the promoter regions of *rtrA* and *rtrB* each contain a predicted SpdR binding site (Table 3.1), providing evidence that transcription of these genes is directly regulated by SpdR. Since *CC_3164* and *CC_2330*

are in the SpdR regulon and regulate *hfiA* promoter activity, I named these genes *rtrA* and *rtrB,* for *R*eg*BA* *t*ranscriptional *r*egulator *A* and *B*, respectively.

Overexpression of each transcription factor alone provides evidence that *rtrB* is a less potent *hfiA* repressor than *rtrA*. This is congruent with my observation that *rtrB* overexpression does not influence holdfast development to the same extent as *rtrA* overexpression (Fig 3.3B and Fig 3.3C). Coordinate overexpression of *rtrA* and *rtrB* did not decrease *hfiA* expression or increase holdfast development more than *lovK*(H180A) or a strain overexpressing *rtrA* alone (Fig 3.3B and Fig 3.3C). Strains overexpressing either one or both *rtrA* and *rtrB* had the same bulk surface adhesion phenotype as *lovK*(H180A) (Fig 3.3D). These data indicate that *rtrA* and *rtrB* regulate *hfiA* transcription, holdfast development, and surface adhesion.

Finally, I measured the effect of deleting *rtrA* and *rtrB* on surface adhesion in wild-type and in *spdR*(D64E) backgrounds (Fig 3.4). In wild-type cells, deletion of *rtrB* but not *rtrA* resulted in a modest decrease in surface adhesion; the Δ*rtrA*Δ*rtrB* adhesion phenotype was similar to Δ*rtrB* (Fig 3.4). I observed similar results in a *spdR*(D64E) background. I conclude that *rtrB* has a larger effect on surface adhesion than *rtrA* under these assay conditions.

RtrA and RtrB are direct regulators of *hfiA* transcription

Having demonstrated that *rtrA* and *rtrB* repress *hfiA* expression and enhance holdfast development, I next tested whether RtrA and RtrB proteins directly bind the *hfiA* promoter. Increasing concentrations of either protein retarded mobility of a labeled *hfiA*

promoter probe relative to the probe alone in an EMSA assay (Fig 3.5A and Fig 3.5B). To assess the specificity of this interaction, I added 10-fold excess of unlabeled specific probe or unlabeled non-specific probe. The unlabeled specific probe competed with the labeled probe for binding to RtrA and RtrB, while unlabeled non-specific probe had little effect on binding of RtrA or RtrB to the *hfiA* promoter region (Fig 3.5A and Fig 3.5B). These data support a model in which RtrA and RtrB directly repress *hfiA* transcription.

RNA sequencing analysis identifies the RtrA and RtrB regulons

My data support a model in which both RtrA and RtrB directly regulate *hfiA* transcription, so I sought to define the regulon of each of these proteins to better understand their functions as transcription factors. I performed RNA sequencing experiments to measure steady-state transcript levels and compared the transcriptomes of strains overexpressing *rtrA, rtrB,* or both to an empty vector (EV) control strain. The majority of differentially expressed genes have reduced expression relative to the EV control strain, indicating that these transcription factors mainly function as repressors (Table 3.2). *hfiA* expression is low when these transcription factors are overexpressed, validating my genetic results (Table 3.2). I sought to identify binding motifs for RtrA and RtrB. However, subjecting the RtrA and RtrB transcriptiomic data to the MEME Suite Motif Discovery (Bailey et al., 2009) server did not uncover significant binding motifs.

Genes in the RtrA and RtrB regulon are listed in Table 3.2 (fold-change > 1.5 and < -1.5 compared to EV control, P-value < 0.001). The RtrA regulon (251 genes) is larger than the RtrB regulon (90 genes) (Figure 3.6 and Table 3.2). 73 of the 90 genes in the

RtrB regulon overlap with RtrA (Figure 3.6); of these 73 genes, 71 are significantly regulated in a strain overexpressing both *rtrA* and *rtrB*. Together, these data indicate that these two transcription factors regulate a similar set of genes. Nonetheless, the majority of genes in the RtrA regulon are specific to RtrA. And while the RtrB regulon largely overlaps RtrA, the fold change in expression is much less for RtrB-regulated genes (Figure 3.7). In other words, RtrB appears to be a weaker repressor than RtrA. However, I cannot exclude the possibility that steady-state levels of RtrB are lower than RtrA in my overexpression strains due to differences in protein stability.

Select genes that are highly repressed by both RtrA and RtrB include the recently identified contact-dependent inhibition by glycine zipper cluster genes (*cdzA-D* and *cdzI*). The Cdz bacteriocin system is highly expressed in stationary phase and enables *C. crescentus* cells that express these genes to kill neighboring cells by surface-dependent contact (Garcia-Bayona et al., 2017). In addition, RtrA and RtrB strongly repressed *CCNA_03826*, a gene predicted to encode a pyocin large subunit family protein. Thus, an additional bacteriocin system may be regulated by these transcription factors. Interestingly, RtrB represses transcription of two genes encoding hypothetical proteins (*CCNA_03967* and *CCNA_03969*) approximately -400 fold relative to the vector control strain. Furthermore, a number of genes encoding putative endonucleases are repressed by both RtrA and RtrB.

Discussion

I identified two transcription factors, *rtrA* and *rtrB*, that function downstream of SpdR to directly repress *hfiA* transcription and control surface adhesion. *rtrA* was previously identified as part of the stationary phase SpdR regulon in *C. crescentus* (da Silva et al., 2016), while *rtrB* was not. The transcriptomic experiments I've described were conducted on cells in logarithmic phase expressing a phosphomimetic SpdR allele. Studies of RegA-dependent transcription in *Rhodobacter capsulatus* have shown that the composition of the RegA regulon depends on the growth medium (Schindel and Bauer, 2016). Similar conditional regulation by SpdR may exist in *C. crescentus*, and thus *rtrA* and *rtrB* may be activated and control *hfiA* transcription under distinct sets of environmental conditions. Possible conditional regulation of *rtrA* and *rtrB* in *C. crescentus* is reflected in my overexpression and genetic epistasis data that show distinct holdfast and surface adhesion phenotypes for the individual *rtrA* and *rtrB* overexpression and deletion strains.

Finally, I performed RNA-seq experiments with the goal to globally define the regulons of *rtrA* and *rtrB.* Overexpression of these transcription factors decrease transcript levels of the genes encoding the Cdz bacteriocin system. The *cdz* genes are induced during stationary phase and the system has been proposed to provide a mechanism to acquire nutrients when they become scarce (Garcia-Bayona et al., 2017). Given that *spdS, spdR* and *rtrA* are transcriptionally induced in stationary phase, how can we reconcile that RtrA and RtrB repress transcription of the *cdz* genes that are also induced in stationary phase? One hypothesis is that stationary phase is dynamic and expression of *cdz* genes in the population is differentially regulated. In other words, RtrA

and RtrB might be important to repress transcription of the *cdz* operon when the killing of neighboring cells has released the necessary nutrients and it is time to turn off the system. If this hypothesis is true, it would be interesting to know if a molecule released by the lysed cells leads to the activation of *rtrA* and *rtrB* expression. Finally, induction of genes involved in contact-dependent killing might also be a mechanism employed while living as a surface-attached community, since living in a biofilm can be beneficial but it also promotes crowding and competition for resources.

Chapter IV: Conclusions and future directions

Two-component signaling systems (TCS) typically function as insulated linear pathways, but emerging data provide evidence that complex cellular decision circuits use interconnected networks of TCS proteins. In this thesis, I sought to understand the mechanism by which the TCS LovK/LovR regulates the transition between motile and surface-attached lifestyles in *Caulobacter crescentus.* Specifically, I focused in understanding the mechanism by which the potent inhibitor of holdfast synthesis, *hfiA*, is transcriptionally regulated. It was previously known that LovK/LovR, represses transcription of *hfiA*, allowing for the development of the holdfast. However, the response regulator, LovR, lacks a DNA-binding output domain and thus transcriptional regulation of *hfiA* by this TCS must be indirect. During the course of my thesis research, I have used genetic, genomic and biochemical approaches to uncover evidence for an unconventional multi-TCS network that coordinately regulates *hfiA* transcription (Figure 4.1). The work presented in this dissertation provides further evidence that *hfiA* is a critical regulator of holdfast development and that its expression is tightly regulated to control cell attachment and dispersal in different environments. Follow up work to this thesis could develop this adhesion control system in *C. crescentus* as a model to investigate how multi-component TCS signaling networks control developmental decisions in bacteria.

The development of a forward genetic approach to identify regulatory proteins

To understand how LovK/LovR regulates transcription of *hfiA*, I conducted a forward genetic selection to identify mutants in which repression of *hfiA* transcription by

LovK/LovR was disrupted. In this screen, the *hfiA* promoter (P_{hfiA}) controlled the expression of chloramphenicol acetyl transferase (*CAT*), and growth in the presence of chloramphenicol was used to select for spontaneous mutants with increased P_{hfiA} activity (Figure 2.4). An advantage of this approach is that I was able to optimize the screen to identify a chloramphenicol concentration with the highest dynamic range in growth rate between wild-type and *lovK*(H180A) strains carrying this reporter plasmid, indicating a highest difference in the levels of *hfiA* transcription. This allowed me to increase the probability of identifying mutants of interest. However, my data showed that a strong limitation of this genetic selection is the high propensity to enrich for mutations that lead to upregulation of genes important for antibiotic resistance (i.e. mutations in *tipR*) (Figure 2.7). I overcame this limitation by ectopically expressing a second copy of *tipR* under its native promoter. Nevertheless, the fact that I did not identify SpdR, RtrA and RtrB in this selection leads me to conclude that I did not reach saturation. A recently described transposon-based approach is an alternative to the genetic selection I performed (Wetmore et al., 2015). Using this approach, *C. crescentus lovK*(H180A) was mutagenized with a pool of barcoded transposons and each unique barcode is assigned to a position in the genome. By growing the *lovK*(H180A) strain in medium containing cheesecloth to titrate away the mutants that are still hyper-adhesive, I would enrich for less adhesive mutants in the supernatant and determine which genes are suppressors of the *lovK*(H180A) hyper-adhesive phenotype. I would expect to identify genes required for holdfast biosynthesis in addition to the genes of interest. I recently performed this experiment and preliminary analysis of the data indicates that all the genes identified in

this thesis work (*skaH, spdS, spdR, rtrA* and *rtrB*) are suppressors of the *lovK*(H180A) hyper-adhesive allele (Fig 4.2). I am in the process of further analyzing the data in order to determine if there is one or more unidentified genes that are required for the *lovK*(H180A) dependent increase in surface adhesion.

A suite of histidine kinases and response regulators regulate *hfiA* transcription and holdfast development

I followed up the genetic selection with directed deletion of the identified genes and investigated their roles as regulators of *hfiA* transcription. I concluded that *skaH, spdR* and *spdS* are required for the LovK/LovR-dependent repression of *hfiA* transcription, and increase in holdfast development and in surface adhesion (Figure 2.8). Furthermore, although HKs are typically obligate homodimers, bacterial-two hybrid and co-purification assays support a model in which these HKs form heteromeric interactions; the HHK SkaH strongly interacts with both LovK and SpdS, while LovK and SpdS do not interact (Figure 2.13). An important open question that arose from this work is the functional significance of the heteromeric interactions I observed. Different models to explain the role of the heteromeric interactions in the signaling pathway identified can be envisioned. One plausible role is to prevent homodimer formation of one or both of the HKs involved to inhibit autophosphorylation activity. The HKs RetS and GacS in *P. aeruginosa* form heteromeric interactions and it was proposed that this interaction prevented GacS from forming homodimers (Goodman et al., 2009). However, it was recently discovered that RetS does not prevent GacS homodimer formation but instead

inhibits GacS auto-kinase activity by a mechanism that involves heteromeric interactions through the DHp domains (Mancl et al., 2019). The exact oligomeric state of each HK and of the heteromeric complexes is unknown. LovK and SkaH belong to the HWE/HisKA_2 family of HKs and there is precedence for atypical non-homodimeric complexes in this family (Wojnowska et al., 2013;Rivera-Cancel et al., 2014;Rinaldi et al., 2016). In addition, SpdS homologues have been shown to form tetramers by disulfide bond formation through a conserved cysteine (Swem et al., 2003). Therefore, it would be possible that *1)* each HK alone can adopt a non-homodimeric oligomeric state and *2)* the heteromeric interactions I observed experimentally could be manifested in heterodimers, dimers of dimers or other oligomeric state combinations in vivo. Finally, heteromeric kinase interactions could prevent phosphoryl transfer to a REC domain in either SkaH, LovR, SpdR or a yet to be identified RR. In the future, purification of the TCS proteins I have identified followed with *in vitro* experiments to assess oligomeric states will provide a better understanding of the role of the heteromeric interactions in HK regulation.

Another open question that arose from my work is the flow of phosphoryl-groups between the identified HKs and RRs. I presented evidence that phospho-SpdR is a repressor of *hfiA* transcription. I hypothesize that LovK(H180A) leads to hyper phosphorylation of SpdR. In an ideal scenario, to study phosphoryl-transfer in this pathway, all the TCS proteins would be purified and HKs would exhibit autophosphorylation activity in vitro. Lori et al identified a SDRR, MrrA, with the ability to transfer phosphoryl-groups to LovK, suggesting that LovK can act as a phosphotransferase in their experimental conditions (Lori et al., 2018). Thus, it is possible

to use the experimental system of Lori et al to phosphorylate LovK and further test phosphoryl-group transfer to the different REC domains in the pathway I identified. It could also be tested if MrrA is able to phosphotransfer to SkaH as well. By co-incubating combinations of HKs and RRs one could probe whether any of the molecular players I've identified have a role in inhibiting or promoting auto-kinase activity of an HK or a role in inhibiting phosphoryl-group transfer.

Identifying the stimuli perceived by HKs is a major challenge when studying TCS regulation. Even when particular conditions are known to activate the HK, the specific molecule that binds to the HK is unknown. The HKs studied in this thesis contain a LOV domain (LovK), two PAS domains (SkaH) and redox sensing modules (SpdS). It is known that LovK responds to blue light through its LOV domain and this regulation is influenced by the redox state of its environment (Purcell et al., 2007;Purcell et al., 2010). Therefore, it is possible that the physiological role of LovK is that of a redox sensor. A role for redox would not be unexpected as SpdS homologues have been shown to sense the redox state of the cell. Though I did not directly test the redox sensing capacity of SpdS, I provide evidence that SpdS is important for holdfast development regulation during stationary phase (Figure 2.8 and Figure 2.11). Therefore, it is plausible that changes in oxygen levels upon entering stationary phase are sensed by SpdS. Recent work from the Crosson group provides evidence that the holdfast facilitates partitioning of cells to the air-liquid interface in aqueous media, where oxygen levels are elevated (Fiebig, 2019a). It would be interesting to test whetjher any of the TCS presented in this dissertation have a role in this behavior. Finally, nothing is known about the sensory capacities of SkaH but

it contains two PAS domains, suggesting that at least it has the potential of sensing two different stimuli. Identifying the specific signals sensed by these HKs is required to better understand the environmental conditions in which *C. crescentus* modulate its ability to attach to a surface.

Two levels of transcriptional regulation modulate *hfiA* transcription

A mutant in which the conserved aspartyl phosphorylation site of *spdR* is changed to alanine (D64A) phenocopies a Δ*spdR* strain (Figure 2.8). On the other hand, replacing the aspartyl phosphorylation site with a glutamate (D64E), which can serve as a phosphomimetic mutation, represses *hfiA* transcription (Figure 3.1). These data support a model in which the DNA-binding RR SpdR functions downstream of LovK/LovR as a transcriptional repressor of *hfiA.* However, with genomics, molecular biology and forward genetics I concluded that SpdR is not a direct regulator of *hfiA* transcription, but instead, it activates transcription of two transcription factors (RtrA and RtrB) which bind and regulate the *hfiA* promoter.

RtrA and RtrB could be activated by SpdR in different environmental conditions. Studies of the SpdR homologue, RegA, in *Rhodobacter capsulatus* have shown that the composition of the RegA regulon depends on the growth medium (Schindel and Bauer, 2016). Different intracellular or extracellular stimuli could lead to different levels of phosphorylated SpdR. One can envision a scenario in which different levels of SpdR~P lead to differences in the SpdR regulon, depending on the binding affinities of SpdR to the conserved promoter binding sequences. The different levels of SpdR-dependent

activation of gene expression could lead to expression of RtrA and RtrB during different growth conditions. Transcriptomic analysis where SpdR is gradually overexpressed and comparison of the conserved binding motifs could be used to test this hypothesis.

Finally, I conducted RNA-sequencing experiments to identify the regulon of RtrA and RtrB as an approach to better understand the global role of these transcription factors. I conclude that the regulon of RtrB highly overlaps with the RtrA regulon. However, I cannot identify a clear and specific role for RtrA and RtrB based on the transcriptomic data alone. Nevertheless, the fact that both RtrA and RtrB regulate genes of contact-dependent bacteriocin systems might indicate a role for these transcription factors in the scavenging of nutrients when *C. crescentus* encounters crowded niches. Finally, subjecting the promoter regions of the genes regulated by RtrA and RtrB to MEME Suite Motif Discovery (Bailey et al., 2009) server did not uncover significant binding motifs. Therefore, the DNA-binding motifs occupied by RtrA and RtrB in the promoter of *hfiA* and of other genes are still unknown. ChIP-sequencing together with DNA-footprinting experiments would provide knowledge of the genes directly regulated by RtrA and RtrB and the specific motifs they bind to.

Chapter V: Materials and Methods

Strain growth conditions. Escherichia coli strains (listed in Table 5.2) were cultivated in Lysogeny Broth (LB) or LB agar plates (1.5% agar) at 37°C. Antibiotics were added to the following final concentrations as required: kanamycin 50 µg/ml, chloramphenicol 20 µg/ml, ampicillin 100 ug/ml, tetracycline 12 µg/ml, gentamicin 15 µg/ml and spectinomycin/streptomycin 50 µg/ml / 30 µg/ml.

Caulobacter crescentus strains (listed in Table 5.1) were cultivated on peptone-yeast extract (PYE) agar (0.2% peptone, 0.1% yeast extract, 1 mM $MgSO_4$, 0.5 mM $CaCl_2$, 1.5% agar) at 30°C (Ely, 1991). Liquid cultures of *C. crescentus* were cultivated at 30°C in PYE (0.2% peptone, 0.1% yeast extract, 1 mM $MgSO_4$, 0.5 mM $CaCl_2$) or M2 defined medium supplemented with xylose (0.15%) as the carbon source (M2X) (Ely, 1991). For expression of a gene from the *vanA* promoter, a final concentration of 0.5 mM of vanillate was added to M2X (M2VX). Antibiotics were added to the following final concentrations as required for liquid cultures: tetracycline 1 µg/ml. For solid medium growth, antibiotics were added as follows: kanamycin 25 µg/ml, chloramphenicol 1 µg/ml, tetracycline 2 µg/ml, gentamicin 5 µg/ml and spectinomycin/streptomycin 50 µg/ml / 5 µg/ml.

Plasmid construction. C. crescentus DNA was amplified from colonies using KOD Xtreme hot-start polymerase (EMD Biosciences/Novagen). PCR reactions were supplemented with 5% dimethyl sulfoxide (DMSO). All plasmids were cloned in *E. coli* Top10 (Invitrogen, Carlsbad, CA). The sequences of all cloned products were confirmed in the target plasmids.

Gene deletions and point mutations were generated by spliced overlapping PCR using external primers with specific restriction enzyme sites. PCR products were digested and ligated into the pNPTS138 plasmid (Ried and Collmer, 1987) at the necessary restriction enzyme sites. Plasmids for ectopic gene expression were generated by ligating a digested PCR product of the gene of interest into *colE1*-based integrating plasmids with either a vanillate (P_{van}) or xylose (P_{xyl}) inducible promoter (Thanbichler et al., 2007). Plasmids are listed in Table 5.2.

Strain construction. Chromosomal allele replacements were made using standard two-step recombination using *sacB* for counterselection. Briefly, pNPTS138-derived plasmids were transformed into *C. crescentus* by electroporation. Primary integrants were selected by growth on PYE Kan plates, inoculated into liquid PYE and grown overnight without selection. Cultures were plated on PYE agar supplemented with 3% sucrose to select for clones that lost the plasmid in a second recombination event. Chromosomal allele replacements were confirmed by PCR amplification of the gene of interest from kanamycin sensitive clones. pRKlac290 transcriptional reporter plasmid P_{hfiA}-*lacZ* was conjugated into *C. crescentus* by triparental mating (Ely, 1991). *C. crescentus* was then grown on PYE plates supplemented with tetracycline to select for cells carrying the plasmid and nalidixic acid to counterselect against *E. coli*. Integrating expression plasmids were transformed into *C. crescentus* by electroporation. Strains carrying the plasmid were selected on PYE agar supplemented with a plasmid-specific antibiotic.

Protein purification for EMSA. For heterologous expression, pET-based plasmids were transformed into Rosetta(DE3)pLysS (Novagen). Strains were inoculated into 10 ml of LB liquid medium supplemented with Kan 50 µg/ml and grown overnight shaking at 37°C. The overnight culture was used to inoculate a flask with 250 ml of fresh LB liquid medium supplemented with Kan 50 µg/ml and grown shaking to OD_{600nm} of 0.6 - 0.8 at 37°C. Then, a final concentration of 1 mM isopropyl β-D-1-thiogalactopyranoside (GoldBio) was used to induce protein expression for 4 hrs under the same conditions. The cultures were harvested by centrifugation at 11,000 × g for 7 min at 4˚C, resuspended in 15 ml of resuspension buffer (10 mM Imidazole, 125 mM NaCl, 25 mM Tris pH 7.6) and saved at -20°C until needed. When samples were thawed, DNAseI and PMSF were added to final concentrations of 250 µg/ml and 800 µM, respectively. The cells were then disrupted by one passage in a microfluidizer (Microfluidics LV1) and the resulting lysate was clarified by centrifugation at 39,000 × g for 15 min at 4°C. Purification of His_6-RtrB, His_6-SUMO-RtrA and His_6-SpdR(D64E) was performed by nickel affinity chromatography (nitrilotriacetic acid resin; GE Healthcare). After binding of the clarified lysate to the column, the sample was washed with 30-50 ml of resuspension buffer followed by 10 ml of wash buffer (75 mM Imidazole, 125 mM NaCl, 25 mM Tris pH 7.6). Protein was eluted with 1-2 ml of elution buffer (500 mM Imidazole, 125 mM NaCl, 25 mM Tris pH 7.6). Sample purity was assessed by resolving the purified protein by 12% SDS-PAGE and the gels were stained with Coomassie blue.

For purification of RtrA, ULP1 protease was added to the eluted protein to cleave the His_6SUMO tag and this sample was dialyzed overnight at 4°C in 1 L of buffer

containing 25 mM Tris pH 7.6, 125 mM $NaCl_2$, 50% glycerol. The cleaved protein was further purified by nickel affinity chromatography, where the flow through containing the cleaved untagged RtrA was collected for use in this study. For His_6-RtrB purification, the eluted protein was dialyzed overnight at 4°C in 1 L of buffer containing 25 mM Tris pH 7.6, 200 mM KCl, 10 mM $MgCl_2$, 15mM EDTA pH 8.0 and 50% glycerol. For purification of His_6-SpdR(D64E), the eluted protein was dialyzed overnight at 4°C in 1 L of buffer containing 25 mM Tris pH 7.6, 125 mM $NaCl_2$ and 50% glycerol. Aliquots of the purified proteins were saved at -80°C.

Genetic selection. The goal of this selection was to isolate spontaneous mutants in which the *lovK*(H180A) allele could no longer repress the *hfiA* promoter (P_{hfiA}). We used a *lovK*(H180A) strain carrying an integrated plasmid in which the promoter of *hfiA* was fused to *cat* (P_{hfiA}-*cat*). The *cat* gene encodes chloramphenicol acetyltransferase, which confers resistance to the antibiotic chloramphenicol in a dose dependent manner. We used growth rate in the presence of chloramphenicol as a proxy for P_{hfiA}-*cat* activity and enriched for spontaneous mutants that grew faster than the parental *lovK*(H180A) strain in the presence of chloramphenicol. We identified optimal selection conditions by comparing growth rate of wild-type and *lovK*(H180A) strains carrying the P_{hfiA}-*cat* plasmid in a range of chloramphenicol concentrations. The highest dynamic range between growth rates in the wild-type and *lovK*(H180A) backgrounds was at 1 - 2 µg/ml chloramphenicol.

For each independent selection, one ml of an overnight culture of the parental strain was used to inoculate flasks containing 500 ml of M2X medium supplemented with a final concentration of 1 - 2 µg/ml of chloramphenicol. Flasks were incubated shaking at 30˚C until they reached OD_{660nm} of 0.6 - 0.9. At this point, 100 - 500 µl of the cultures was used to inoculate 500 ml of fresh M2X medium. Cultures were allowed to grow and this passaging process was performed until the diluted cultures grew fast enough to reach OD_{660nm} of 0.6 - 0.9 in a single overnight incubation. Upon initial inoculation, cultures took 3 – 5 days to reach the desired OD_{660nm} but after 5 - 7 serial passages cultures only took about 18 hrs to reach OD_{660nm} of 0.6 - 0.9. This indicated that the population was more resistant to chloramphenicol. Individual mutant clones were isolated from the final (fast-growing) culture by plating serial dilutions on PYE agar supplemented with gentamicin 5 µg/ml, which selected for cells carrying the P_{hfiA}-*cat* plasmid. Several isolated single colonies were picked for further analysis.

Several secondary screens were performed on the isolated mutants. First, we sequenced the *hfiA* promoter from the P_{hfiA}-*cat* plasmid in order to distinguish clones with intact selection systems from those with *cis* mutations in the *hfiA* promoter. We then confirmed faster growth rate in the presence of chloramphenicol. Then, to distinguish mutants that result in de-repression of P_{hfiA} from those that confer chloramphenicol resistance through other mechanisms, we evaluated bulk surface attachment of the isolated mutants by crystal violet stain or ring assay (see below) as a proxy for activity of the native *hfiA* promoter. Finally, mutants that displayed faster growth rate and decreased surface attachment compared to the *lovK*(H180A) parental strain were transformed with

the P_{hfiA}-*lacZ* transcriptional fusion plasmid. β-galactosidase activity was measured to confirm de-repression of the *hfiA* promoter in the spontaneous mutants compared to the *lovK*(H180A) strain. Based on the results obtained during the secondary screens, several isolates were selected and the mutations in these strains were mapped by whole genome sequencing.

Whole genome sequencing. We isolated genomic DNA from the *C. crescentus* parent and the mutant strains arising from our chloramphenicol selection using a standard guanidinium thiocyanate extraction and isopropanol/ethanol precipitation. The DNA was randomly sheared and libraries were prepared for whole genome shotgun sequencing using an Illumina HighSeq 2500 (50-bp single end reads). Whole-genome sequence data from the parent and mutant strains were assembled to the *C. crescentus* NA1000 genome (genbank accession CP001340), and polymorphisms between the parental strain and suppressor strains were identified using Geneious v 11 (Kearse et al., 2012), see Table S1. The sequenced genome libraries yielded an average of 11.4 million reads, resulting in global average depth of coverage of 141x.

Holdfast quantification and imaging. Overnight cultures grown in M2X were diluted to an approximate OD_{660nm} of 0.00002 in 2 ml of M2X. When the cultures reached 0.05 - 0.1 OD_{660nm}, 5 µl of 1 mg/ml Wheat Germ Agglutinin, Alexa Fluor 594 Conjugate (ThermoFisher) was added to the culture. After incubation for 10 minutes, cells were collected by centrifugation for 2 minutes and resuspended in 10 - 25 µl of M2X. Cells were

imaged with a Leica DM5000 microscope. Fluorescent WGA staining was visualized using an RFP fluorescent filter (Chroma set 41043).

Crystal violet stain surface attachment assay. Colonies from PYE agar plates were inoculated into 2 ml of M2X broth and grown overnight at 30°C shaking at 200 rpm. Overnight liquid cultures were diluted to a final OD_{660nm} of 0.00625 into wells of 24-well polystyrene plates containing 1 ml of M2X or M2VX. Plates were grown shaking at 155rpm at 30°C for varying times periods as indicated in the results. Plates were removed from the incubator and surface attached cells were measured with crystal violet stain. Briefly, the cultures were discarded and the wells were thoroughly washed with water. Then, 1.5 ml of 0.01% crystal violet in water was added to each well and the plates were incubated with shaking at 155 rpm for 5 min. The crystal violet solution was discarded and the wells thoroughly washed with water again. To dissolve the stain, 1.5 ml of 100% EtOH was added to each well and the plates incubated again with shaking for 5 min. Crystal violet stain extracted in each well was quantified by measuring absorbance at 575nm.

B-galactosidase assay. Strains were inoculated in M2X medium from colonies on PYE-agar plates and grown shaking overnight at 30°C. Overnight cultures were then diluted back into fresh M2X to an OD_{660nm} of 0.00025 – 0.00075 and grown in a shaking incubator at 30°C until they reached OD_{660nm} 0.05 - 0.15 (15 - 20 hrs). At this point, β-galactosidase activity was measured as previously described (Fiebig et al., 2014).

Bacterial two-hybrid. We utilized a system previously described (Karimova et al., 1998). We co-transformed plasmids bearing fusions to either the T18c or T25 domains of adenylate cyclase into the adenylate cyclase null strain, BTH101, by electroporation. An aliquot of the outgrowth was plated on LB agar supplemented with Amp 100 µg/ml, Kan 50 µg/ml, X-gal 80 µg/ml and IPTG 0.5 mM and grown at 30°C for 24 hrs. Strains were inoculated into LB liquid medium supplemented with Amp 100 µg/ml, Kan 50 µg/ml and IPTG 0.5 mM and grown shaking overnight at 30°C. Overnight cultures were diluted to OD_{600nm} of 0.05 and 5 µl of each culture were spotted into LB agar plates supplemented with Amp 100 µg/ml, Kan 50 µg/ml, X-gal 80 µg/ml and IPTG 0.5 mM. The color of each spot was evaluated after 36 hours of growth at 30°C.

Protein pulldown from cell lysate. C. crescentus colonies from PYE agar plates were inoculated into 2 ml of M2X and grown shaking overnight at 30°C. These cultures were diluted into 40 ml of M2X and grown until they reached OD_{660nm} of 0.4 - 0.6. The cultures were then induced with 0.5 mM vanillate for 3 hrs. The cells were harvested by centrifugation at 11,000 × g for 15 min at 4°C, resuspended in 6 ml of standard buffer (25 mM Tris pH 7.6, 125 mM NaCl) and frozen at -20°C. When thawed, cells were disrupted by one passage in a microfluidizer (Microfluidics LV1), PMSF was added to a final concentration of 800 uM and the lysates were cleared by centrifugation at 21,000 × g for 5 min at 4°C. The cleared lysate was loaded into a column of approximately 200 µl amylose resin (NEB) prequilibrated in buffer, and allowed to bind. The resin was washed with 30-50 ml standard buffer and eluted with 150 µl standard buffer supplemented with

40 mM maltose. All the fractions collected for analysis were mixed 1:1 with SDS loading buffer, boiled for 2 min at 95°C and saved at -20°C for further Western Blot analysis.

Western blot of pulldown fractions. 15 µl of each pulldown fraction sample was loaded onto a Mini-PROTEAN TGX Precast 4-20% Gradient Gel (Bio-Rad). Samples were resolved at 35 mA constant current in SDS running buffer (0.3% Tris, 18.8% Glycine, 0.1% SDS). Proteins in the gel were transferred to an Immobilon®-P PVDF Membrane using a Mini Trans-Blot® Cell after preincubation in Western transfer buffer (0.3% Tris, 18.8% Glycine, 20% methanol). Transfer was carried out at 100 volts for 1 hr at 4°C in Western transfer buffer. The membrane was then blocked in 5% powdered milk in Tris-buffered Saline Tween (TBST: 137 mM NaCl, 2.3 mM KCl, 20 mM Tris pH 7.4, 0.1% Tween 20) shaking for 1 hr at 4°C. Incubation with the primary antibody, FLAG monoclonal antibody (clone FG4R) or HA-Tag monoclonal antibody, was carried out shaking overnight in 5% powdered milk TBST at 4°C. Membrane was then washed 3 times in TBST for a total of 30 min at room temperature. Incubation with Goat anti-Mouse IgG (H+L)-HRP secondary antibody was at room temperature for 1 h in TBST. Finally, the membrane was washed 3 times in TBST for a total of 30 min at room temperature. Chemiluminescence was performed using the SuperSignal™ West Femto Maximum Sensitivity Substrate (Pierce) and was imaged using a ChemiDoc MP imaging system (Bio-Rad). Chemiluminescence was measured using the ChemSens program with an exposure time of 30 - 60 sec.

Electrophoretic Mobility Shift Assays. Promoter regions of interest were amplified by PCR using a set of appropriate primers in which the reverse primer was fluorescently labeled with AlexaFluor 488 (Integrated DNA Technology, DTT). The *hfiA* promoter probe was 275 bp (primers F:TGGTGGTCCTGATCCTCCTG and R: CACTGACAACATCCTGTCCG) and the *cspD* promoter probe was 199 bp (primers F: CTAGGGACTGCCATCTTCGG and R: TCGTAACCAGACATCCCACC). Specific cold specific competitor probe represents the same region as the corresponding labeled probe amplified with unlabeled primers with the same sequence. The cold non-specific competitor was a 220 bp PCR product amplified from the pNPTS138 plasmid using primers F: GTAAAACGACGGCCAG and R: CAGGAAACAGCTATGAC.

For reactions in which the binding of His_6-SpdR(D64E) was tested, DNA binding reactions were performed in 20 µl reaction volumes containing binding buffer (20mM HEPES pH 7.8, 4 mM $MgCl_2$, 37 mM KCl, 1 mM $CaCl_2$, 0.75 mM dithiothreitol, 8% glycerol and 75 µg/ml bovine serum albumin), 8 ng of fluorescently labeled DNA probe and increasing concentrations of His_6-SpdR(D64E) (0, 50, 125, 250, 375 nM for incubation with the *hfiA* promoter probe and 0, 25, 50, 125, 250 nM for incubation with the *cspD* promoter probe). Reactions were incubated at room temperature for 30 min in the dark. Then, 12.5 µl of each reaction was loaded on a fresh 6% native acrylamide gel pre-run for at least 30 min in 1X Tris Acetate EDTA buffer (TAE: 40 mM Tris, 20 mM acetic acid and 1 mM EDTA) in the dark at 4˚C for 1 hr at 70 volts followed by 1.5 hr at 90 volts. The gels were imaged using the BioRad Chemidoc MP imaging system with a 60 - 120 sec exposure and the manufacturer's settings for AlexaFluor 488 detection.

For reactions where the binding of RtrA was tested, DNA binding reactions were performed in 20 µl reaction volumes containing binding buffer (10 mM Tris pH7.6, 2 mM $MgCl_2$, 50 mM KCl, 2 mM $CaCl_2$, 0.2 mM dithiothreitol, 5% glycerol, 1 mM EDTA pH 8.0 and 1 mg/ml bovine serum albumin), 8 ng of fluorescently labeled DNA probe and increasing concentrations of RtrA (0, 100, 200, 300, 400 nM). For competition experiments, a 10-fold excess of cold competitor probe was added to the reaction. Reactions were incubated at room temperature for 30 min in the dark. 12.5 µl of each reaction was loaded on a fresh 6% native acrylamide gel pre-run for at least 30 min in 1X Tris Borate EDTA (TBE) in the dark at 4˚C for 1 hr at 100 volts. The gels were imaged using the BioRad Chemidoc MP imaging system with a 30 sec exposure and the manufacturer's settings for AlexaFluor 488 detection.

For reactions where the binding of RtrB was tested, DNA binding reactions were performed in 20 µl reaction volumes containing binding buffer (20mM HEPES pH 7.8, 4 mM $MgCl_2$, 37 mM KCl, 1 mM $CaCl_2$, 0.75 mM dithiothreitol, 8% glycerol and 75 µg/ml bovine serum albumin), 7 ng of fluorescently labeled DNA probe and increasing concentrations of RtrB (0, 10, 25, 50, 75, 100, 250 nM). For competition experiments, a 10-fold excess of cold competitor probe was added to the reaction. Reactions were incubated at room temperature for 30 min in the dark. 12.5 µl of each reaction was loaded on a fresh 6% native acrylamide gel pre-run for at least 30 min in 1X TAE in the dark at 4˚C for 30 min at 80 volts, and subsequently 1 hr at 100 volts. The gels were imaged using the BioRad Chemidoc MP imaging system with a 30 sec exposure and the manufacturer's settings for AlexaFluor 488 detection.

RNA preparation, sequencing and analysis. Strains grown on PYE agar plates were inoculated into 5 ml of M2X medium in 3 biological replicates for each strain, and grown overnight on a rolling incubator at 30°C. Cultures were diluted to OD_{660nm} of 0.008 in 8 ml of fresh M2X and grown in the same manner until they reached OD_{660nm} of 0.30 – 0.35. At this point, 6 ml of each replicate were collected by centrifugation at 15,000 × g for 1 min. This cell pellet was immediately resuspended in 1 ml of TRIzol and stored at -80˚ C. The samples were heated for 10 min at 65°C. Then, after addition of 200 µl of chloroform samples were vortexed and incubated at room temperature for 5 min. Phases were separated by centrifugation at 15,000 × g for 15 min at 4°C. The upper aqueous phase was transferred to a new tube and 0.7 volume of cold 100% isopropanol was added. Samples were then stored at -80°C overnight. The overnight precipitation was centrifuged at 15,000 × for 30 min at 4°C and the resulting nucleic acid pellet was washed twice with cold 70% ethanol. The pellet was centrifuged again at 15,000 × for 5 min at 4°C, ethanol was removed and the pellet was allowed to dry. The nucleic acid pellet was resuspended with nuclease-free water. The samples were treated with Turbo DNase (Ambion, Life Technologies) and clean-up was performed with RNeasy Mini Kit (Qiagen).

Stranded cDNA libraries were prepared and sequenced by the Functional Genomics Core at the University of Chicago. Briefly, RNA samples were treated with the Ribo-zero Kit (Illumina) for rRNA removal and library preparation proceeded with Illumina ScriptSeq RNA-Seq Library Preparation Kit. All the libraries were sequenced with an Illumina HiSeq4000 instrument. Analysis of RNA-seq data was performed with CLC

Genomics Workbench (Qiagen). Reads were mapped to the *C. crescentus* NA1000 genome (Genome accession number CP001340). We used the RNA-seq Analysis Tool to analyze the differential expression between the different strains.

RNA-seq data availability. The raw RNA-seq data for each *spdR*(D64E) and Δ*spdR* sample are available in the NCBI Gene Expression Omnibus (GEO) under accession number GSE125783.

SpdR binding motif discovery. We used the MEME Suite (Bailey et al., 2009) to identify potential SpdR binding motifs in promoter regions of genes with more than 2-fold differences in expression between the *spdR*(D64E) and Δ*spdR* strains. The promoter sequences (-150 to +50, based on translation start site annotations in GenBank accession CP001340) of genes with > +2-fold change (up-regulated in *spdR*(D64E) relative to Δ*spdR*) were submitted to MEME to identify motifs. A motif with high similarity to that previously reported for *C. crescentus* SpdR (da Silva et al., 2010;da Silva et al., 2016) and for SpdR homologs (Emmerich et al., 2000;Swem et al., 2001;Laguri et al., 2003) was found in 65 of 119 promoter sequences, each with a p-value < e^{-4} (listed in Table 3.1; see Fig 3.2). Using the FIMO Motif Scanning algorithm, we applied the SpdR motif matrix from MEME to scan the promoter sequences (-150 to +50) of genes with < -2-fold change (down-regulated by *spdR*(D64E) strain relative to Δ*spdR*) for potential SpdR motif sequences. Predicted SpdR binding motifs with a p-value cutoff < e^{-4} were present in the

promoter sequence of 17 out of 48 of the down-regulated genes, and are listed in Table 3.1.

Bibliography

Abel, S., Bucher, T., Nicollier, M., Hug, I., Kaever, V., Abel Zur Wiesch, P., and Jenal, U. (2013). Bi-modal distribution of the second messenger c-di-GMP controls cell fate and asymmetry during the caulobacter cell cycle. *PLoS Genet* 9**,** e1003744.

Aguilar, C., Vlamakis, H., Guzman, A., Losick, R., and Kolter, R. (2010). KinD is a checkpoint protein linking spore formation to extracellular-matrix production in Bacillus subtilis biofilms. *MBio* 1.

Allen, H.L. (1971). Primary Productivity, Chemo-organotrophy, and Nutritional Interactions of Epiphytic Algae and Bacteria on Macrophytes in the Littoral of a Lake. *Ecological Monographs* 41**,** 97-127.

Ardissone, S., and Viollier, P.H. (2015). Interplay between flagellation and cell cycle control in Caulobacter. *Curr Opin Microbiol* 28**,** 83-92.

Ashenberg, O., Rozen-Gagnon, K., Laub, M.T., and Keating, A.E. (2011). Determinants of homodimerization specificity in histidine kinases. *J Mol Biol* 413**,** 222-235.

Azam, F., and Malfatti, F. (2007). Microbial structuring of marine ecosystems. *Nat Rev Microbiol* 5**,** 782-791.

Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W., and Noble, W.S. (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res* 37**,** W202-208.

Benson, A.K., and Haldenwang, W.G. (1993). Bacillus subtilis sigma B is regulated by a binding protein (RsbW) that blocks its association with core RNA polymerase. *Proc Natl Acad Sci U S A* 90**,** 2330-2334.

Benson, A.K., Wu, J., and Newton, A. (1994). The role of FlbD in regulation of flagellar gene transcription in Caulobacter crescentus. *Res Microbiol* 145**,** 420-430.

Berne, C., Ellison, C.K., Agarwal, R., Severin, G.B., Fiebig, A., Morton, R.I., 3rd, Waters, C.M., and Brun, Y.V. (2018). Feedback regulation of Caulobacter crescentus

holdfast synthesis by flagellum assembly via the holdfast inhibitor HfiA. *Mol Microbiol* 110**,** 219-238.

Berne, C., Ma, X., Licata, N.A., Neves, B.R., Setayeshgar, S., Brun, Y.V., and Dragnea, B. (2013). Physiochemical properties of Caulobacter crescentus holdfast: a localized bacterial adhesive. *J Phys Chem B* 117**,** 10492-10503.

Bhate, M.P., Molnar, K.S., Goulian, M., and Degrado, W.F. (2015). Signal transduction in histidine kinases: insights from new structures. *Structure* 23**,** 981-994.

Bielnicki, J., Devedjiev, Y., Derewenda, U., Dauter, Z., Joachimiak, A., and Derewenda, Z.S. (2006). *B. subtilis* ykuD protein at 2.0 A resolution: insights into the structure and function of a novel, ubiquitous family of bacterial enzymes. *Proteins* 62**,** 144-151.

Biondi, E.G., Reisinger, S.J., Skerker, J.M., Arif, M., Perchuk, B.S., Ryan, K.R., and Laub, M.T. (2006). Regulation of the bacterial cell cycle by an integrated genetic circuit. *Nature* 444**,** 899-904.

Bird, T.H., Du, S., and Bauer, C.E. (1999). Autophosphorylation, phosphotransfer, and DNA-binding properties of the RegB/RegA two-component regulatory system in Rhodobacter capsulatus. *J Biol Chem* 274**,** 16343-16348.

Boyd, C.D., and O'toole, G.A. (2012). Second messenger regulation of biofilm formation: breakthroughs in understanding c-di-GMP effector systems. *Annu Rev Cell Dev Biol* 28**,** 439-462.

Broder, U.N., Jaeger, T., and Jenal, U. (2016). LadS is a calcium-responsive kinase that induces acute-to-chronic virulence switch in Pseudomonas aeruginosa. *Nat Microbiol* 2**,** 16184.

Brunberg, A.K., Nilsson, E., and Blomqvist, P. (2002). Characteristics of oligotrophic hardwater lakes in a postglacial land-rise area in mid-Sweden. *Freshwater Biology* 47**,** 1451-1462.

Burbulys, D., Trach, K.A., and Hoch, J.A. (1991). Initiation of sporulation in B. subtilis is controlled by a multicomponent phosphorelay. *Cell* 64**,** 545-552.

Cao, M., and Goodrich-Blair, H. (2017). Ready or Not: Microbial Adaptive Responses in Dynamic Symbiosis Environments. *J Bacteriol* 199**,** pii: e00883-00816.

Capra, E.J., and Laub, M.T. (2012). Evolution of two-component signal transduction systems. *Annu Rev Microbiol* 66**,** 325-347.

Capra, E.J., Perchuk, B.S., Skerker, J.M., and Laub, M.T. (2012). Adaptive mutations that prevent crosstalk enable the expansion of paralogous signaling protein families. *Cell* 150**,** 222-232.

Carlson, R.E. (1977). Trophic State Index for Lakes. *Limnology and Oceanography* 22**,** 361-369.

Carmany, D.O., Hollingsworth, K., and Mccleary, W.R. (2003). Genetic and biochemical studies of phosphatase activity of PhoR. *J Bacteriol* 185**,** 1112-1115.

Chambonnier, G., Roux, L., Redelberger, D., Fadel, F., Filloux, A., Sivaneson, M., De Bentzmann, S., and Bordi, C. (2016). The Hybrid Histidine Kinase LadS Forms a Multicomponent Signal Transduction System with the GacS/GacA Two-Component System in Pseudomonas aeruginosa. *PLoS Genet* 12**,** e1006032.

Chamnongpol, S., Cromie, M., and Groisman, E.A. (2003). Mg2+ sensing by the Mg2+ sensor PhoQ of Salmonella enterica. *J Mol Biol* 325**,** 795-807.

Chao, M.C., Abel, S., Davis, B.M., and Waldor, M.K. (2016). The design and analysis of transposon insertion sequencing experiments. *Nat Rev Microbiol* 14**,** 119-128.

Chen, C.C., Lewis, R.J., Harris, R., Yudkin, M.D., and Delumeau, O. (2003). A supramolecular complex in the environmental stress signalling pathway of Bacillus subtilis. *Mol Microbiol* 49**,** 1657-1669.

Cheung, J., Bingman, C.A., Reyngold, M., Hendrickson, W.A., and Waldburger, C.D. (2008). Crystal structure of a functional dimer of the PhoQ sensor domain. *J Biol Chem* 283**,** 13762-13770.

Cheung, J., and Hendrickson, W.A. (2009). Structural analysis of ligand stimulation of the histidine kinase NarX. *Structure* 17**,** 190-201.

Cheung, J., and Hendrickson, W.A. (2010). Sensor domains of two-component regulatory systems. *Curr Opin Microbiol* 13**,** 116-123.

Childers, W.S., Xu, Q., Mann, T.H., Mathews, Ii, Blair, J.A., Deacon, A.M., and Shapiro, L. (2014). Cell fate regulation governed by a repurposed bacterial histidine kinase. *PLoS Biol* 12**,** e1001979.

Christen, B., Abeliuk, E., Collier, J.M., Kalogeraki, V.S., Passarelli, B., Coller, J.A., Fero, M.J., Mcadams, H.H., and Shapiro, L. (2011). The essential genome of a bacterium. *Mol Syst Biol* 7**,** 528.

Collier, J. (2016). Cell cycle control in Alphaproteobacteria. *Curr Opin Microbiol* 30**,** 107-113.

Comolli, J.C., Carl, A.J., Hall, C., and Donohue, T. (2002). Transcriptional activation of the Rhodobacter sphaeroides cytochrome c(2) gene P2 promoter by the response regulator PrrA. *J Bacteriol* 184**,** 390-399.

Costerton, J.W., Lewandowski, Z., Caldwell, D.E., Korber, D.R., and Lappin-Scott, H.M. (1995). Microbial biofilms. *Annu Rev Microbiol* 49**,** 711-745.

Curtis, P.D., and Brun, Y.V. (2014). Identification of essential alphaproteobacterial genes reveals operational variability in conserved developmental and cell cycle systems. *Mol Microbiol* 93**,** 713-735.

Da Silva, C.A., Balhesteros, H., Mazzon, R.R., and Marques, M.V. (2010). SpdR, a response regulator required for stationary-phase induction of Caulobacter crescentus cspD. *J Bacteriol* 192**,** 5991-6000.

Da Silva, C.A., Lourenco, R.F., Mazzon, R.R., Ribeiro, R.A., and Marques, M.V. (2016). Transcriptomic analysis of the stationary phase response regulator SpdR in Caulobacter crescentus. *BMC Microbiol* 16**,** 66.

Davis, M.C., Kesthely, C.A., Franklin, E.A., and Maclellan, S.R. (2017). The essential activities of the bacterial sigma factor. *Can J Microbiol* 63**,** 89-99.

De Been, M., Tempelaars, M.H., Van Schaik, W., Moezelaar, R., Siezen, R.J., and Abee, T. (2010). A novel hybrid kinase is essential for regulating the sigma(B)-mediated stress response of Bacillus cereus. *Environ Microbiol* 12**,** 730-745.

Degnen, S.T., and Newton, A. (1972). Chromosome replication during development in Caulobacter crescentus. *J Mol Biol* 64**,** 671-680.

Dejesus, M.A., Ambadipudi, C., Baker, R., Sassetti, C., and Ioerger, T.R. (2015). TRANSIT--A Software Tool for Himar1 TnSeq Analysis. *PLoS Comput Biol* 11**,** e1004401.

Delumeau, O., Chen, C.C., Murray, J.W., Yudkin, M.D., and Lewis, R.J. (2006). High-molecular-weight complexes of RsbR and paralogues in the environmental signaling pathway of Bacillus subtilis. *J Bacteriol* 188**,** 7885-7892.

Djordjevic, S., Goudreau, P.N., Xu, Q., Stock, A.M., and West, A.H. (1998). Structural basis for methylesterase CheB regulation by a phosphorylation-activated domain. *Proc Natl Acad Sci U S A* 95**,** 1381-1386.

Domian, I.J., Quon, K.C., and Shapiro, L. (1997). Cell type-specific phosphorylation and proteolysis of a transcriptional regulator controls the G1-to-S transition in a bacterial cell cycle. *Cell* 90**,** 415-424.

Domian, I.J., Reisenauer, A., and Shapiro, L. (1999). Feedback control of a master bacterial cell-cycle regulator. *Proc Natl Acad Sci U S A* 96**,** 6648-6653.

Drever, J.I. (1997). *The Geochemistry of Natural Waters: Surface and Groundwater Environments.* Englewood Cliffs, NJ: Prentice Hall.

Dutta, R., and Inouye, M. (1996). Reverse phosphotransfer from OmpR to EnvZ in a kinase-/phosphatase+ mutant of EnvZ (EnvZ.N347D), a bifunctional signal transducer of Escherichia coli. *J Biol Chem* 271**,** 1424-1429.

Ellison, C.K., Kan, J., Dillard, R.S., Kysela, D.T., Ducret, A., Berne, C., Hampton, C.M., Ke, Z., Wright, E.R., Biais, N., Dalia, A.B., and Brun, Y.V. (2017). Obstruction of pilus retraction stimulates bacterial surface sensing. *Science* 358**,** 535-538.

Elsen, S., Swem, L.R., Swem, D.L., and Bauer, C.E. (2004). RegB/RegA, a highly conserved redox-responding global two-component regulatory system. *Microbiol Mol Biol Rev* 68**,** 263-279.

Ely, B. (1991). Genetics of Caulobacter crescentus. *Methods Enzymol* 204**,** 372-384.

Ely, B., Croft, R.H., and Gerardot, C.J. (1984). Genetic mapping of genes required for motility in Caulobacter crescentus. *Genetics* 108**,** 523-532.

Emmerich, R., Panglungtshang, K., Strehler, P., Hennecke, H., and Fischer, H.M. (1999). Phosphorylation, dephosphorylation and DNA-binding of the Bradyrhizobium japonicum RegSR two-component regulatory proteins. *Eur J Biochem* 263**,** 455-463.

Emmerich, R., Strehler, P., Hennecke, H., and Fischer, H.M. (2000). An imperfect inverted repeat is critical for DNA binding of the response regulator RegR of Bradyrhizobium japonicum. *Nucleic Acids Res* 28**,** 4166-4171.

Faulds-Pain, A., Birchall, C., Aldridge, C., Smith, W.D., Grimaldi, G., Nakamura, S., Miyata, T., Gray, J., Li, G., Tang, J.X., Namba, K., Minamino, T., and Aldridge, P.D. (2011). Flagellin redundancy in Caulobacter crescentus and its implications for flagellar filament assembly. *J Bacteriol* 193**,** 2695-2707.

Fiebig, A. (2019a). The role of Caulobacter cell surface structures in colonization of the air-liquid interface. *J Bacteriol*.

Fiebig, A., Herrou, J., Fumeaux, C., Radhakrishnan, S.K., Viollier, P.H., and Crosson, S. (2014). A cell cycle and nutritional checkpoint controlling bacterial surface adhesion. *PLoS Genet* 10**,** e1004101.

Fiebig, A., Herrou, J., Willett, J., and Crosson, S. (2015). General Stress Signaling in the Alphaproteobacteria. *Annu Rev Genet* 49**,** 603-625.

Fiebig, A., Varesio, L.M., Navarreto, X.A., and Crosson, S. (2019b). Regulation of the Erythrobacter litoralis DSM 8509 general stress response by visible light. . *Molecular Microbiology. In Press*.

Flemming, H.C., Wingender, J., Szewzyk, U., Steinberg, P., Rice, S.A., and Kjelleberg, S. (2016). Biofilms: an emergent form of bacterial life. *Nat Rev Microbiol* 14**,** 563-575.

Foreman, R., Fiebig, A., and Crosson, S. (2012). The LovK-LovR two-component system is a regulator of the general stress pathway in Caulobacter crescentus. *J Bacteriol* 194**,** 3038-3049.

Francez-Charlot, A., Frunzke, J., Reichen, C., Ebneter, J.Z., Gourion, B., and Vorholt, J.A. (2009). Sigma factor mimicry involved in regulation of general stress response. *Proc Natl Acad Sci U S A* 106**,** 3467-3472.

Francez-Charlot, A., Kaczmarczyk, A., Fischer, H.M., and Vorholt, J.A. (2015). The general stress response in Alphaproteobacteria. *Trends Microbiol* 23**,** 164-171.

Francis, V.I., Stevenson, E.C., and Porter, S.L. (2017). Two-component systems required for virulence in Pseudomonas aeruginosa. *FEMS Microbiol Lett* 364.

Francis, V.I., Waters, E.M., Finton-James, S.E., Gori, A., Kadioglu, A., Brown, A.R., and Porter, S.L. (2018). Multiple communication mechanisms between sensor kinases are crucial for virulence in Pseudomonas aeruginosa. *Nat Commun* 9**,** 2219.

Fujita, M., and Losick, R. (2005). Evidence that entry into sporulation in Bacillus subtilis is governed by a gradual increase in the level and activity of the master regulator Spo0A. *Genes Dev* 19**,** 2236-2244.

Galperin, M.Y., Makarova, K.S., Wolf, Y.I., and Koonin, E.V. (2015). Expanded microbial genome coverage and improved protein family annotation in the COG database. *Nucleic Acids Res* 43**,** D261-269.

Gao, R., Mack, T.R., and Stock, A.M. (2007). Bacterial response regulators: versatile regulatory strategies from common domains. *Trends Biochem Sci* 32**,** 225-234.

Garcia-Bayona, L., Guo, M.S., and Laub, M.T. (2017). Contact-dependent killing by Caulobacter crescentus via cell surface-associated, glycine zipper proteins. *Elife* 6.

Gober, J.W., and Shapiro, L. (1992). A developmentally regulated Caulobacter flagellar promoter is activated by 3' enhancer and IHF binding elements. *Mol Biol Cell* 3**,** 913-926.

Goodman, A.L., Merighi, M., Hyodo, M., Ventre, I., Filloux, A., and Lory, S. (2009). Direct interaction between sensor kinase proteins mediates acute and chronic disease phenotypes in a bacterial pathogen. *Genes Dev* 23**,** 249-259.

Gourion, B., Sulser, S., Frunzke, J., Francez-Charlot, A., Stiefel, P., Pessi, G., Vorholt, J.A., and Fischer, H.M. (2009). The PhyR-sigma(EcfG) signalling cascade is involved in stress response and symbiotic efficiency in Bradyrhizobium japonicum. *Mol Microbiol* 73**,** 291-305.

Hardy, G.G., Allen, R.C., Toh, E., Long, M., Brown, P.J., Cole-Tobian, J.L., and Brun, Y.V. (2010). A localized multimeric anchor attaches the Caulobacter holdfast to the cell pole. *Mol Microbiol* 76**,** 409-427.

He, K., Dragnea, V., and Bauer, C.E. (2015). Adenylate Charge Regulates Sensor Kinase CheS3 To Control Cyst Formation in Rhodospirillum centenum. *MBio* 6**,** e00546-00515.

He, K., Marden, J.N., Quardokus, E.M., and Bauer, C.E. (2013). Phosphate flow between hybrid histidine kinases CheA(3) and CheS(3) controls Rhodospirillum centenum cyst formation. *PLoS Genet* 9**,** e1004002.

Hecht, G.B., Lane, T., Ohta, N., Sommer, J.M., and Newton, A. (1995). An essential single domain response regulator required for normal cell division and differentiation in Caulobacter crescentus. *EMBO J* 14**,** 3915-3924.

Hecker, M., Pane-Farre, J., and Volker, U. (2007). SigB-dependent general stress response in Bacillus subtilis and related gram-positive bacteria. *Annu Rev Microbiol* 61**,** 215-236.

Helmann, J.D., Wu, M.F., Kobel, P.A., Gamo, F.J., Wilson, M., Morshedi, M.M., Navre, M., and Paddon, C. (2001). Global transcriptional response of Bacillus subtilis to heat shock. *J Bacteriol* 183**,** 7318-7328.

Henry, J.T., and Crosson, S. (2011). Ligand-binding PAS domains in a genomic, cellular, and structural context. *Annu Rev Microbiol* 65**,** 261-286.

Hentchel, K.L., Reyes Ruiz, L.M., Curtis, P.D., Fiebig, A., Coleman, M.L., and Crosson, S. (2019). Genome-scale fitness profile of Caulobacter crescentus grown in natural freshwater. *ISME J* 13**,** 523-536.

Herrou, J., and Crosson, S. (2011). Function, structure and mechanism of bacterial photosensory LOV proteins. *Nat Rev Microbiol* 9**,** 713-723.

Herrou, J., Crosson, S., and Fiebig, A. (2017). Structure and function of HWE/HisKA2-family sensor histidine kinases. *Curr Opin Microbiol* 36**,** 47-54.

Herrou, J., Foreman, R., Fiebig, A., and Crosson, S. (2010). A structural model of anti-anti-sigma inhibition by a two-component receiver domain: the PhyR stress response regulator. *Mol Microbiol* 78**,** 290-304.

Hershey, D.M., Fiebig, A., and Crosson, S. (2019). A genome-wide analysis of adhesion in Caulobacter crescentus identifies new regulatory and biosynthetic components for holdfast assembly. *mBio* 12**,** pii: e02273-02218.

Hershey, D.M., Porfírio, S., Jaehrig, B., Heiss, C., Azadi, P., Fiebig, A., and Crosson, S. (2019). Composition of the holdfast polysaccharide from Caulobacter crescentus. *Biorxiv*.

Hoch, J.A. (2017). A Life in Bacillus subtilis Signal Transduction. *Annu Rev Microbiol* 71**,** 1-19.

Hoffman, M.D., Zucker, L.I., Brown, P.J., Kysela, D.T., Brun, Y.V., and Jacobson, S.C. (2015). Timescales and Frequencies of Reversible and Irreversible Adhesion Events of Single Bacterial Cells. *Anal Chem* 87**,** 12032-12039.

Hottes, A.K., Meewan, M., Yang, D., Arana, N., Romero, P., Mcadams, H.H., and Stephens, C. (2004). Transcriptional profiling of *Caulobacter crescentus* during growth on complex and minimal media. *J Bacteriol* 186**,** 1448-1461.

Hsing, W., and Silhavy, T.J. (1997). Function of conserved histidine-243 in phosphatase activity of EnvZ, the sensor for porin osmoregulation in Escherichia coli. *J Bacteriol* 179**,** 3729-3735.

Hsu, J.L., Chen, H.C., Peng, H.L., and Chang, H.Y. (2008). Characterization of the histidine-containing phosphotransfer protein B-mediated multistep phosphorelay system in Pseudomonas aeruginosa PAO1. *J Biol Chem* 283**,** 9933-9944.

Hug, I., Deshpande, S., Sprecher, K.S., Pfohl, T., and Jenal, U. (2017). Second messenger-mediated tactile response by a bacterial rotary motor. *Science* 358**,** 531-534.

Hutchison, E.A., Miller, D.A., and Angert, E.R. (2014). Sporulation in Bacteria: Beyond the Standard Model. *Microbiol Spectr* 2.

Iniesta, A.A., Mcgrath, P.T., Reisenauer, A., Mcadams, H.H., and Shapiro, L. (2006). A phospho-signaling pathway controls the localization and activity of a protease complex critical for bacterial cell cycle progression. *Proc Natl Acad Sci U S A* 103**,** 10935-10940.

Jacob-Dubuisson, F., Mechaly, A., Betton, J.M., and Antoine, R. (2018). Structural insights into the signalling mechanisms of two-component systems. *Nat Rev Microbiol* 16**,** 585-593.

Jacobs, C., Hung, D., and Shapiro, L. (2001). Dynamic localization of a cytoplasmic signal transduction response regulator controls morphogenesis during the Caulobacter cell cycle. *Proc Natl Acad Sci U S A* 98**,** 4095-4100.

Jenal, U., and Fuchs, T. (1998). An essential protease involved in bacterial cell-cycle control. *EMBO J* 17**,** 5658-5669.

Jiang, M., Shao, W., Perego, M., and Hoch, J.A. (2000a). Multiple histidine kinases regulate entry into stationary phase and sporulation in Bacillus subtilis. *Mol Microbiol* 38**,** 535-542.

Jiang, P., Atkinson, M.R., Srisawat, C., Sun, Q., and Ninfa, A.J. (2000b). Functional dissection of the dimerization and enzymatic activities of Escherichia coli nitrogen regulator II and their regulation by the PII protein. *Biochemistry* 39**,** 13433-13449.

Johnson, R.C., and Ely, B. (1979). Analysis of nonmotile mutants of the dimorphic bacterium Caulobacter crescentus. *J Bacteriol* 137**,** 627-634.

Kaczmarczyk, A., Campagne, S., Danza, F., Metzger, L.C., Vorholt, J.A., and Francez-Charlot, A. (2011). Role of Sphingomonas sp. strain Fr1 PhyR-NepR-sigmaEcfG cascade in general stress response and identification of a negative regulator of PhyR. *J Bacteriol* 193**,** 6629-6638.

Kaczmarczyk, A., Hochstrasser, R., Vorholt, J.A., and Francez-Charlot, A. (2014). Complex two-component signaling regulates the general stress response in Alphaproteobacteria. *Proc Natl Acad Sci U S A* 111**,** E5196-5204.

Karimova, G., Pidoux, J., Ullmann, A., and Ladant, D. (1998). A bacterial two-hybrid system based on a reconstituted signal transduction pathway. *Proc Natl Acad Sci U S A* 95**,** 5752-5756.

Karniol, B., and Vierstra, R.D. (2004). The HWE histidine kinases, a new family of bacterial two-component sensor kinases with potentially diverse roles in environmental signaling. *J Bacteriol* 186**,** 445-453.

Kato, M., Mizuno, T., Shimizu, T., and Hakoshima, T. (1997). Insights into multistep phosphorelay from the crystal structure of the C-terminal HPt domain of ArcB. *Cell* 88**,** 717-723.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., and Drummond, A. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28**,** 1647-1649.

Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., and Sternberg, M.J.E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols* 10**,** 845.

Kent, W.J. (2002). BLAT--the BLAST-like alignment tool. *Genome Res* 12**,** 656-664.

Kim, H.S., Willett, J.W., Jain-Gupta, N., Fiebig, A., and Crosson, S. (2014). The Brucella abortus virulence regulator, LovhK, is a sensor kinase in the general stress response signalling pathway. *Mol Microbiol* 94**,** 913-925.

Kim, T.J., Gaidenko, T.A., and Price, C.W. (2004a). In vivo phosphorylation of partner switching regulators correlates with stress transmission in the environmental signaling pathway of Bacillus subtilis. *J Bacteriol* 186**,** 6124-6132.

Kim, T.J., Gaidenko, T.A., and Price, C.W. (2004b). A multicomponent protein complex mediates environmental stress signaling in Bacillus subtilis. *J Mol Biol* 341**,** 135-150.

Kirkpatrick, C.L., and Viollier, P.H. (2014). Synthetic interaction between the TipN polarity factor and an AcrAB-family efflux pump implicates cell polarity in bacterial drug resistance. *Chem Biol* 21**,** 657-665.

Kong, W., Chen, L., Zhao, J., Shen, T., Surette, M.G., Shen, L., and Duan, K. (2013). Hybrid sensor kinase PA1611 in Pseudomonas aeruginosa regulates transitions between acute and chronic infection through direct interaction with RetS. *Mol Microbiol* 88**,** 784-797.

Kumar, P., Kaushik, A., Lloyd, E.P., Li, S.G., Mattoo, R., Ammerman, N.C., Bell, D.T., Perryman, A.L., Zandi, T.A., Ekins, S., Ginell, S.L., Townsend, C.A., Freundlich, J.S., and Lamichhane, G. (2017). Non-classical transpeptidases yield insight into new antibacterials. *Nat Chem Biol* 13**,** 54-61.

Kwon, Y.M., Ricke, S.C., and Mandal, R.K. (2016). Transposon sequencing: methods and expanding applications. *Appl Microbiol Biotechnol* 100**,** 31-43.

Laguri, C., Phillips-Jones, M.K., and Williamson, M.P. (2003). Solution structure and DNA binding of the effector domain from the global regulator PrrA (RegA) from Rhodobacter sphaeroides: insights into DNA binding specificity. *Nucleic Acids Res* 31**,** 6778-6787.

Lam, H., Matroule, J.Y., and Jacobs-Wagner, C. (2003). The asymmetric spatial distribution of bacterial signal transduction proteins coordinates cell cycle events. *Dev Cell* 5**,** 149-159.

Langridge, G.C., Phan, M.D., Turner, D.J., Perkins, T.T., Parts, L., Haase, J., Charles, I., Maskell, D.J., Peters, S.E., Dougan, G., Wain, J., Parkhill, J., and Turner, A.K. (2009). Simultaneous assay of every Salmonella Typhi gene using one million transposon mutants. *Genome Res* 19**,** 2308-2316.

Lapouge, K., Schubert, M., Allain, F.H., and Haas, D. (2008). Gac/Rsm signal transduction pathway of gamma-proteobacteria: from RNA recognition to regulation of social behaviour. *Mol Microbiol* 67**,** 241-253.

Lau, J., Hernandez-Alicea, L., Vass, R.H., and Chien, P. (2015). A Phosphosignaling Adaptor Primes the AAA+ Protease ClpXP to Drive Cell Cycle-Regulated Proteolysis. *Mol Cell* 59**,** 104-116.

Laub, M.T., Chen, S.L., Shapiro, L., and Mcadams, H.H. (2002). Genes directly controlled by CtrA, a master regulator of the Caulobacter cell cycle. *Proc Natl Acad Sci U S A* 99**,** 4632-4637.

Laub, M.T., and Goulian, M. (2007). Specificity in two-component signal transduction pathways. *Annu Rev Genet* 41**,** 121-145.

Lee, K., and Yoon, S.S. (2017). Pseudomonas aeruginosa Biofilm, a Programmed Bacterial Life for Fitness. *J Microbiol Biotechnol* 27**,** 1053-1064.

Lee, P.O., Mclellan, S.L., Graham, L.E., and Young, E.B. (2015). Invasive dreissenid mussels and benthic algae in Lake Michigan: characterizing effects on sediment bacterial communities. *FEMS Microbiol Ecol* 91**,** 1-12.

Leroux, M., Kirkpatrick, R.L., Montauti, E.I., Tran, B.Q., Peterson, S.B., Harding, B.N., Whitney, J.C., Russell, A.B., Traxler, B., Goo, Y.A., Goodlett, D.R., Wiggins, P.A., and Mougous, J.D. (2015). Kin cell lysis is a danger signal that activates antibacterial pathways of Pseudomonas aeruginosa. *Elife* 4.

Levi, A., and Jenal, U. (2006). Holdfast formation in motile swarmer cells optimizes surface attachment during Caulobacter crescentus development. *J Bacteriol* 188**,** 5315-5318.

Li, G., Brown, P.J., Tang, J.X., Xu, J., Quardokus, E.M., Fuqua, C., and Brun, Y.V. (2012). Surface contact stimulates the just-in-time deployment of bacterial adhesins. *Mol Microbiol* 83**,** 41-51.

Li, G., Smith, C.S., Brun, Y.V., and Tang, J.X. (2005). The elastic properties of the caulobacter crescentus adhesive holdfast are dependent on oligomers of N-acetylglucosamine. *J Bacteriol* 187**,** 257-265.

Lin, C.T., Huang, Y.J., Chu, P.H., Hsu, J.L., Huang, C.H., and Peng, H.L. (2006). Identification of an HptB-mediated multi-step phosphorelay in Pseudomonas aeruginosa PAO1. *Res Microbiol* 157**,** 169-175.

Loeb, G.I., and Neihof, R.A. (1975). Marine Conditional Films. *Adv Chem***,** 319-335.

Loeffler, F. (1890). Weitere Untersuchungen uber die Beizung und Farbung der Geisseln bei den Bakterien. *Centralbl Bakteriol Parasitenkd* 7**,** 625-639.

Lopez, D., Fischbach, M.A., Chu, F., Losick, R., and Kolter, R. (2009). Structurally diverse natural products that cause potassium leakage trigger multicellularity in Bacillus subtilis. *Proc Natl Acad Sci U S A* 106**,** 280-285.

Lori, C., Kaczmarczyk, A., De Jong, I., and Jenal, U. (2018). A Single-Domain Response Regulator Functions as an Integrating Hub To Coordinate General Stress Response and Development in Alphaproteobacteria. *MBio* 9.

Lori, C., Ozaki, S., Steiner, S., Bohm, R., Abel, S., Dubey, B.N., Schirmer, T., Hiller, S., and Jenal, U. (2015). Cyclic di-GMP acts as a cell cycle oscillator to drive chromosome replication. *Nature* 523**,** 236-239.

Mancl, J.M., Ray, W.K., Helm, R.F., and Schubot, F.D. (2019). Helix Cracking Regulates the Critical Interaction between RetS and GacS in Pseudomonas aeruginosa. *Structure*.

Mann, T.H., and Shapiro, L. (2018). Integration of cell cycle signals by multi-PAS domain kinases. *Proc Natl Acad Sci U S A* 115**,** E7166-E7173.

Marina, A., Waldburger, C.D., and Hendrickson, W.A. (2005). Structure of the entire cytoplasmic portion of a sensor histidine-kinase protein. *EMBO J* 24**,** 4247-4259.

Marks, M.E., Castro-Rojas, C.M., Teiling, C., Du, L., Kapatral, V., Walunas, T.L., and Crosson, S. (2010). The genetic basis of laboratory adaptation in Caulobacter crescentus. *J Bacteriol* 192**,** 3678-3688.

Marshall, K.C. (1996). "Adhesion as a strategy for access to nutrients," in *Bacterial Adhesion: Molecular and Ecological Diversity,* ed. M. Fletcher. (New York: Wiley-Liss, Inc.), 59-87.

Martinez-Garcia, E., Nikel, P.I., Chavarria, M., and De Lorenzo, V. (2014). The metabolic cost of flagellar motion in Pseudomonas putida KT2440. *Environ Microbiol* 16**,** 291-303.

Mascher, T., Helmann, J.D., and Unden, G. (2006). Stimulus perception in bacterial signal-transducing histidine kinases. *Microbiol Mol Biol Rev* 70**,** 910-938.

Mcmillan, L., and Stout, R. (1977). Occurrence of *Sphaerotilus*, *Caulobacter*, and *Gallionella* in Raw and Treated Water. *J Am Water Works Assoc* 69**,** 171-173.

Merker, R.I., and Smit, J. (1988). Characterization of the adhesive holdfast of marine and freshwater caulobacters. *Appl Environ Microbiol* 54**,** 2078-2085.

Metsalu, T., and Vilo, J. (2015). ClustVis: a web tool for visualizing clustering of multivariate data using Principal Component Analysis and heatmap. *Nucleic Acids Res* 43**,** W566-570.

Mike, L.A., Choby, J.E., Brinkman, P.R., Olive, L.Q., Dutter, B.F., Ivan, S.J., Gibbs, C.M., Sulikowski, G.A., Stauff, D.L., and Skaar, E.P. (2014). Two-component system cross-regulation integrates Bacillus anthracis response to heme and cell envelope stress. *PLoS Pathog* 10**,** e1004044.

Mira, A., Ochman, H., and Moran, N.A. (2001). Deletional bias and the evolution of bacterial genomes. *Trends Genet* 17**,** 589-596.

Moore, C.M., Nakano, M.M., Wang, T., Ye, R.W., and Helmann, J.D. (2004). Response of Bacillus subtilis to nitric oxide and the nitrosating agent sodium nitroprusside. *J Bacteriol* 186**,** 4655-4664.

Narayanan, S., Janakiraman, B., Kumar, L., and Radhakrishnan, S.K. (2015). A cell cycle-controlled redox switch regulates the topoisomerase IV activity. *Genes Dev* 29**,** 1175-1187.

Nesper, J., Hug, I., Kato, S., Hee, C.S., Habazettl, J.M., Manfredi, P., Grzesiek, S., Schirmer, T., Emonet, T., and Jenal, U. (2017). Cyclic di-GMP differentially tunes a bacterial flagellar motor through a novel class of CheY-like regulators. *Elife* 6.

Norsworthy, A.N., and Visick, K.L. (2015). Signaling between two interacting sensor kinases promotes biofilms and colonization by a bacterial symbiont. *Mol Microbiol* 96**,** 233-248.

Ong, C.J., Wong, M.L., and Smit, J. (1990). Attachment of the adhesive holdfast organelle to the cellular stalk of Caulobacter crescentus. *J Bacteriol* 172**,** 1448-1456.

Petersohn, A., Brigulla, M., Haas, S., Hoheisel, J.D., Volker, U., and Hecker, M. (2001). Global analysis of the general stress response of Bacillus subtilis. *J Bacteriol* 183**,** 5617-5631.

Pitcher, D.G., Saunders, N.A., and Owen, R.J. (1989). Rapid extraction of bacterial genomic DNA with guanidium thiocyanate. *Lett Appl Microbiol* 8**,** 151-156.

Poindexter, J.S. (1964). Biological Properties and Classification of the Caulobacter Group. *Bacteriol Rev* 28**,** 231-295.

Poindexter, J.S. (2006). "Dimorphic Prosthecate Bacteria: The Genera Caulobacter, Asticcacaulis, Hyphomicrobium, Pedomicrobium, Hyphomonas and Thiodendron," in *The Prokaryotes: Volume 5: Proteobacteria: Alpha and Beta Subclasses,* eds. M. Dworkin, S. Falkow, E. Rosenberg, K.-H. Schleifer & E. Stackebrandt. (New York, NY: Springer New York), 72-90.

Porter, S.L., and Armitage, J.P. (2002). Phosphotransfer in Rhodobacter sphaeroides chemotaxis. *J Mol Biol* 324**,** 35-45.

Potter, C.A., Ward, A., Laguri, C., Williamson, M.P., Henderson, P.J., and Phillips-Jones, M.K. (2002). Expression, purification and characterisation of full-length histidine protein kinase RegB from Rhodobacter sphaeroides. *J Mol Biol* 320**,** 201-213.

Price, C.W., Fawcett, P., Ceremonie, H., Su, N., Murphy, C.K., and Youngman, P. (2001). Genome-wide analysis of the general stress response in Bacillus subtilis. *Mol Microbiol* 41**,** 757-774.

Price, M.N., Wetmore, K.M., Waters, R.J., Callaghan, M., Ray, J., Liu, H., Kuehl, J.V., Melnyk, R.A., Lamson, J.S., Suh, Y., Carlson, H.K., Esquivel, Z., Sadeeshkumar, H., Chakraborty, R., Zane, G.M., Rubin, B.E., Wall, J.D., Visel, A., Bristow, J., Blow, M.J., Arkin, A.P., and Deutschbauer, A.M. (2018). Mutant phenotypes for thousands of bacterial genes of unknown function. *Nature* 557**,** 503-509.

Purcell, E.B., Mcdonald, C.A., Palfey, B.A., and Crosson, S. (2010). An analysis of the solution structure and signaling mechanism of LovK, a sensor histidine kinase integrating light and redox signals. *Biochemistry* 49**,** 6761-6770.

Purcell, E.B., Siegal-Gaskins, D., Rawling, D.C., Fiebig, A., and Crosson, S. (2007). A photosensory two-component system regulates bacterial cell attachment. *Proc Natl Acad Sci U S A* 104**,** 18241-18246.

Quon, K.C., Yang, B., Domian, I.J., Shapiro, L., and Marczynski, G.T. (1998). Negative control of bacterial DNA replication by a cell cycle regulatory protein that binds at the chromosome origin. *Proc Natl Acad Sci U S A* 95**,** 120-125.

Reyes Ruiz, L.M., Fiebig, A., and Crosson, S. (2019). Regulation of bacterial surface attachment by a network of sensory transduction proteins. *PLoS Genet* 15**,** e1008022.

Ried, J.L., and Collmer, A. (1987). An nptI-sacB-sacR cartridge for constructing directed, unmarked mutations in gram-negative bacteria by marker exchange-eviction mutagenesis. *Gene* 57**,** 239-246.

Rinaldi, J., Arrar, M., Sycz, G., Cerutti, M.L., Berguer, P.M., Paris, G., Estrin, D.A., Marti, M.A., Klinke, S., and Goldbaum, F.A. (2016). Structural Insights into the HWE Histidine Kinase Family: The Brucella Blue Light-Activated Histidine Kinase Domain. *J Mol Biol* 428**,** 1165-1179.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 43**,** e47.

Rivera-Cancel, G., Ko, W.H., Tomchick, D.R., Correa, F., and Gardner, K.H. (2014). Full-length structure of a monomeric histidine kinase reveals basis for sensory regulation. *Proc Natl Acad Sci U S A* 111**,** 17839-17844.

Romling, U., Gomelsky, M., and Galperin, M.Y. (2005). C-di-GMP: the dawning of a novel bacterial signalling system. *Mol Microbiol* 57**,** 629-639.

Schindel, H.S., and Bauer, C.E. (2016). The RegA regulon exhibits variability in response to altered growth conditions and differs markedly between Rhodobacter species. *Microb Genom* 2**,** e000081.

Schneider, R.P., and Leis, A. (2003). Conditioning Films in Aquatic Environments. *In Encyclopedia of Environmental Microbiology* G. Bitton (Ed.).

Schrader, J.M., Zhou, B., Li, G.W., Lasker, K., Childers, W.S., Williams, B., Long, T., Crosson, S., Mcadams, H.H., Weissman, J.S., and Shapiro, L. (2014). The coding and noncoding architecture of the Caulobacter crescentus genome. *PLoS Genet* 10**,** e1004463.

Skerker, J.M., Perchuk, B.S., Siryaporn, A., Lubin, E.A., Ashenberg, O., Goulian, M., and Laub, M.T. (2008). Rewiring the specificity of two-component signal transduction systems. *Cell* 133**,** 1043-1054.

Skerker, J.M., Prasol, M.S., Perchuk, B.S., Biondi, E.G., and Laub, M.T. (2005). Two-component signal transduction pathways regulating growth and cell cycle progression in a bacterium: a system-level analysis. *PLoS Biol* 3**,** e334.

Smith, C.S., Hinz, A., Bodenmiller, D., Larson, D.E., and Brun, Y.V. (2003). Identification of genes required for synthesis of the adhesive holdfast in Caulobacter crescentus. *J Bacteriol* 185**,** 1432-1442.

Smith, D.R., and Chapman, M.R. (2010). Economical evolution: microbes reduce the synthetic cost of extracellular proteins. *MBio* 1**,** 10.1128/mBio.00131-00110.

Sourjik, V., and Schmitt, R. (1996). Different roles of CheY1 and CheY2 in the chemotaxis of Rhizobium meliloti. *Mol Microbiol* 22**,** 427-436.

Sourjik, V., and Schmitt, R. (1998). Phosphotransfer between CheA, CheY1, and CheY2 in the chemotaxis signal transduction chain of Rhizobium meliloti. *Biochemistry* 37**,** 2327-2335.

Sprecher, K.S., Hug, I., Nesper, J., Potthoff, E., Mahi, M.A., Sangermani, M., Kaever, V., Schwede, T., Vorholt, J., and Jenal, U. (2017). Cohesive Properties of the Caulobacter crescentus Holdfast Adhesin Are Regulated by a Novel c-di-GMP Effector Protein. *MBio* 8.

Stein, B.J., Grant, R.A., Sauer, R.T., and Baker, T.A. (2016). Structural Basis of an N-Degron Adaptor with More Stringent Specificity. *Structure* 24**,** 232-242.

Stephens, C., Christen, B., Fuchs, T., Sundaram, V., Watanabe, K., and Jenal, U. (2007a). Genetic analysis of a novel pathway for D-xylose metabolism in Caulobacter crescentus. *J Bacteriol* 189**,** 2181-2185.

Stephens, C., Christen, B., Watanabe, K., Fuchs, T., and Jenal, U. (2007b). Regulation of D-xylose metabolism in Caulobacter crescentus by a LacI-type repressor. *J Bacteriol* 189**,** 8828-8834.

Stock, A.M., Robinson, V.L., and Goudreau, P.N. (2000). Two-component signal transduction. *Annu Rev Biochem* 69**,** 183-215.

Stocker, R. (2012). Marine microbes see a sea of gradients. *Science* 338**,** 628-633.

Swem, L.R., Elsen, S., Bird, T.H., Swem, D.L., Koch, H.G., Myllykallio, H., Daldal, F., and Bauer, C.E. (2001). The RegB/RegA two-component regulatory system controls synthesis of photosynthesis and respiratory electron transfer components in Rhodobacter capsulatus. *J Mol Biol* 309**,** 121-138.

Swem, L.R., Gong, X., Yu, C.A., and Bauer, C.E. (2006). Identification of a ubiquinone-binding site that affects autophosphorylation of the sensor kinase RegB. *J Biol Chem* 281**,** 6768-6775.

Swem, L.R., Kraft, B.J., Swem, D.L., Setterdahl, A.T., Masuda, S., Knaff, D.B., Zaleski, J.M., and Bauer, C.E. (2003). Signal transduction by the global regulator RegB is mediated by a redox-active cysteine. *EMBO J* 22**,** 4699-4708.

Thanbichler, M., Iniesta, A.A., and Shapiro, L. (2007). A comprehensive set of plasmids for vanillate- and xylose-inducible gene expression in Caulobacter crescentus. *Nucleic Acids Res* 35**,** e137.

Toh, E., Kurtz, H.D., Jr., and Brun, Y.V. (2008). Characterization of the Caulobacter crescentus holdfast polysaccharide biosynthesis pathway reveals significant redundancy in the initiating glycosyltransferase and polymerase steps. *J Bacteriol* 190**,** 7219-7231.

Tomomori, C., Tanaka, T., Dutta, R., Park, H., Saha, S.K., Zhu, Y., Ishima, R., Liu, D., Tong, K.I., Kurokawa, H., Qian, H., Inouye, M., and Ikura, M. (1999). Solution structure of the homodimeric core domain of Escherichia coli histidine kinase EnvZ. *Nat Struct Biol* 6**,** 729-734.

Tsang, P.H., Li, G., Brun, Y.V., Freund, L.B., and Tang, J.X. (2006). Adhesion of single bacterial cells in the micronewton range. *Proc Natl Acad Sci U S A* 103**,** 5764-5768.

Tsokos, C.G., Perchuk, B.S., and Laub, M.T. (2011). A dynamic complex of signaling proteins uses polar localization to regulate cell-fate asymmetry in Caulobacter crescentus. *Dev Cell* 20**,** 329-341.

Van Gestel, J., Vlamakis, H., and Kolter, R. (2015). Division of Labor in Biofilms: the Ecology of Cell Differentiation. *Microbiol Spectr* 3**,** MB-0002-2014.

Van Opijnen, T., and Camilli, A. (2013). Transposon insertion sequencing: a new tool for systems-level analysis of microorganisms. *Nat Rev Microbiol* 11**,** 435-442.

Varughese, K.I., Madhusudan, Zhou, X.Z., Whiteley, J.M., and Hoch, J.A. (1998). Formation of a novel four-helix bundle and molecular recognition sites by dimerization of a response regulator phosphotransferase. *Mol Cell* 2**,** 485-493.

Vijay, K., Brody, M.S., Fredlund, E., and Price, C.W. (2000). A PP2C phosphatase containing a PAS domain is required to convey signals of energy stress to the sigmaB transcription factor of Bacillus subtilis. *Mol Microbiol* 35**,** 180-188.

Voelker, U., Voelker, A., Maul, B., Hecker, M., Dufour, A., and Haldenwang, W.G. (1995). Separate mechanisms activate sigma B of Bacillus subtilis in response to environmental and metabolic stresses. *J Bacteriol* 177**,** 3771-3780.

Vos, M., Wolf, A.B., Jennings, S.J., and Kowalchuk, G.A. (2013). Micro-scale determinants of bacterial diversity in soil. *FEMS Microbiol Rev* 37**,** 936-954.

Wan, Z., Brown, P.J., Elliott, E.N., and Brun, Y.V. (2013). The adhesive and cohesive properties of a bacterial polysaccharide adhesin are modulated by a deacetylase. *Mol Microbiol* 88**,** 486-500.

Wang, L., Grau, R., Perego, M., and Hoch, J.A. (1997). A novel histidine kinase inhibitor regulating development in Bacillus subtilis. *Genes Dev* 11**,** 2569-2579.

Wetmore, K.M., Price, M.N., Waters, R.J., Lamson, J.S., He, J., Hoover, C.A., Blow, M.J., Bristow, J., Butland, G., Arkin, A.P., and Deutschbauer, A. (2015). Rapid quantification of mutant fitness in diverse bacteria by sequencing randomly bar-coded transposons. *MBio* 6**,** e00306-00315.

Wheeler, R.T., and Shapiro, L. (1999). Differential localization of two histidine kinases controlling bacterial cell differentiation. *Mol Cell* 4**,** 683-694.

Wilhelm, R. (2018). Following the terrestrial tracks of Caulobacter - redefining the ecology of a reputed aquatic oligotroph. *ISME J* doi: 10.1038/s41396-018-0257-z.

Willett, J.W., and Crosson, S. (2017). Atypical modes of bacterial histidine kinase signaling. *Mol Microbiol* 103**,** 197-202.

Wojnowska, M., Yan, J., Sivalingam, G.N., Cryar, A., Gor, J., Thalassinos, K., and Djordjevic, S. (2013). Autophosphorylation activity of a soluble hexameric histidine kinase correlates with the shift in protein conformational equilibrium. *Chem Biol* 20**,** 1411-1420.

Wu, J., and Bauer, C.E. (2010). RegB kinase activity is controlled in part by monitoring the ratio of oxidized to reduced ubiquinones in the ubiquinone pool. *MBio* 1.

Wu, J., Cheng, Z., Reddie, K., Carroll, K., Hammad, L.A., Karty, J.A., and Bauer, C.E. (2013a). RegB kinase activity is repressed by oxidative formation of cysteine sulfenic acid. *J Biol Chem* 288**,** 4755-4762.

Wu, J., Ohta, N., and Newton, A. (1998). An essential, multicomponent signal transduction pathway required for cell cycle regulation in Caulobacter. *Proc Natl Acad Sci U S A* 95**,** 1443-1448.

Wu, R., Gu, M., Wilton, R., Babnigg, G., Kim, Y., Pokkuluri, P.R., Szurmant, H., Joachimiak, A., and Schiffer, M. (2013b). Insight into the sporulation phosphorelay: crystal structure of the sensor domain of Bacillus subtilis histidine kinase, KinD. *Protein Sci* 22**,** 564-576.

Xu, H., Dingwall, A., and Shapiro, L. (1989). Negative transcriptional regulation in the Caulobacter flagellar hierarchy. *Proc Natl Acad Sci U S A* 86**,** 6656-6660.

Xu, Q., and West, A.H. (1999). Conservation of structure and function among histidine-containing phosphotransfer (HPt) domains as revealed by the crystal structure of YPD1. *J Mol Biol* 292**,** 1039-1050.

Yang, X., Kang, C.M., Brody, M.S., and Price, C.W. (1996). Opposing pairs of serine protein kinases and phosphatases transmit signals of environmental stress to activate a bacterial transcription factor. *Genes Dev* 10**,** 2265-2275.

Zhang, S., and Haldenwang, W.G. (2005). Contributions of ATP, GTP, and redox state to nutritional stress activation of the Bacillus subtilis sigmaB transcription factor. *J Bacteriol* 187**,** 7554-7560.

Zhou, B., Schrader, J.M., Kalogeraki, V.S., Abeliuk, E., Dinh, C.B., Pham, J.Q., Cui, Z.Z., Dill, D.L., Mcadams, H.H., and Shapiro, L. (2015). The global regulatory architecture of transcription during the Caulobacter cell cycle. *PLoS Genet* 11**,** e1004831.

Zhu, Y., Qin, L., Yoshida, T., and Inouye, M. (2000). Phosphatase activity of histidine kinase EnvZ without kinase catalytic domain. *Proc Natl Acad Sci U S A* 97**,** 7808-7813.

Zschiedrich, C.P., Keidel, V., and Szurmant, H. (2016). Molecular Mechanisms of Two-Component Signal Transduction. *J Mol Biol* 428**,** 3752-3775.

Appendix A: Figures

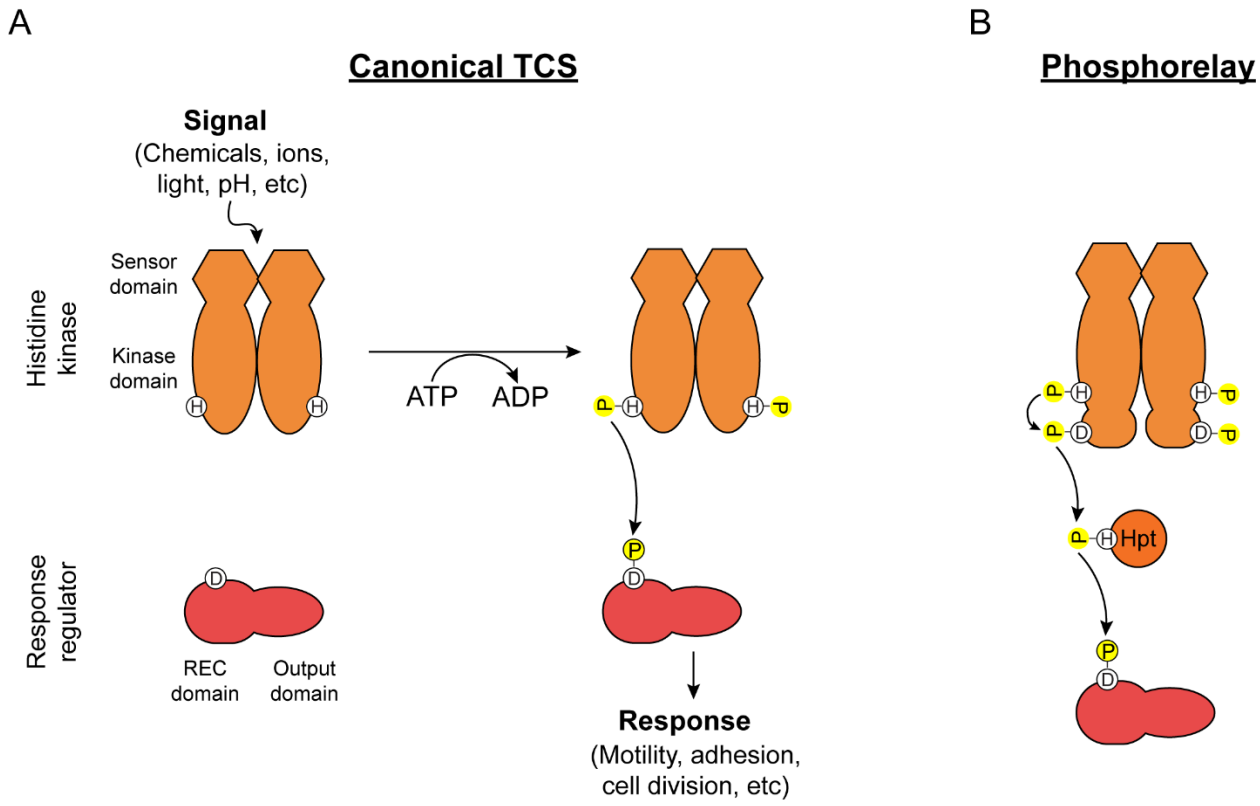


Figure 1.1: Two-component systems are employed by bacteria to sense and respond to stimuli. A) In a canonical two component system (TCS), a histidine kinase autophosphorylates a conserved histidine and the phosphoryl group is subsequently transferred to a conserved aspartate in the receiver (REC) domain of the response regulator. TCS perceive intracellular or extracellular stimuli to cue physiological adaptation of the cell. **B)** Phosphorelays employ hybrid histidine kinases and histidine phosphotransferases (Hpt) to transduce the signal to a terminal response regulator involved in the output response.

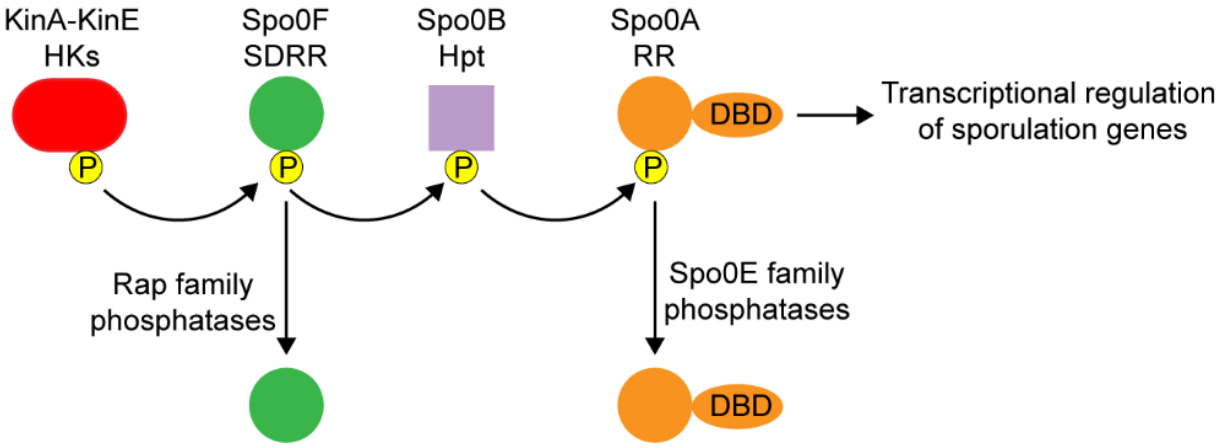


Figure 1.2: Regulation of sporulation in *Bacillus subtilis* by a phosphorelay. The histidine kinases (HK) KinA-KinE phosphorylate the single domain response regulator (SDRR) Spo0F, followed by transfer to the histidine phosphotransferase (Hpt) Spo0B, and a final phosphotransfer to the response regulator (RR) with a DNA-binding domain (DBD), Spo0A. Spo0A regulates transcription of genes involved in sporulation. Negative regulation of the phosphorelay include dephosphorylation of Spo0F by Rap family phosphatases and dephosphorylation of Spo0A by Spo0E family phosphatases. Figure adapted from (Hoch, 2017).

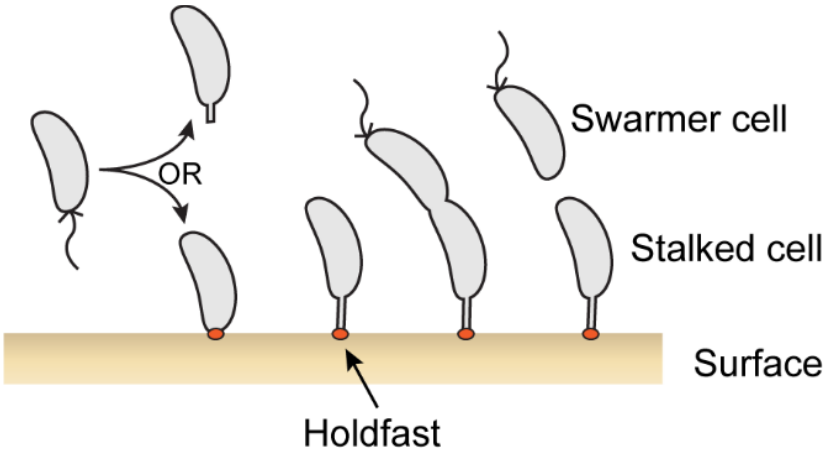


Figure 1.3: *Caulobacter crescentus* cell cycle. The dimorphic bacterium *C. crescentus* builds a polysaccharide-based adhesin known as the holdfast (orange) at one cell pole, which mediates irreversible surface attachment.

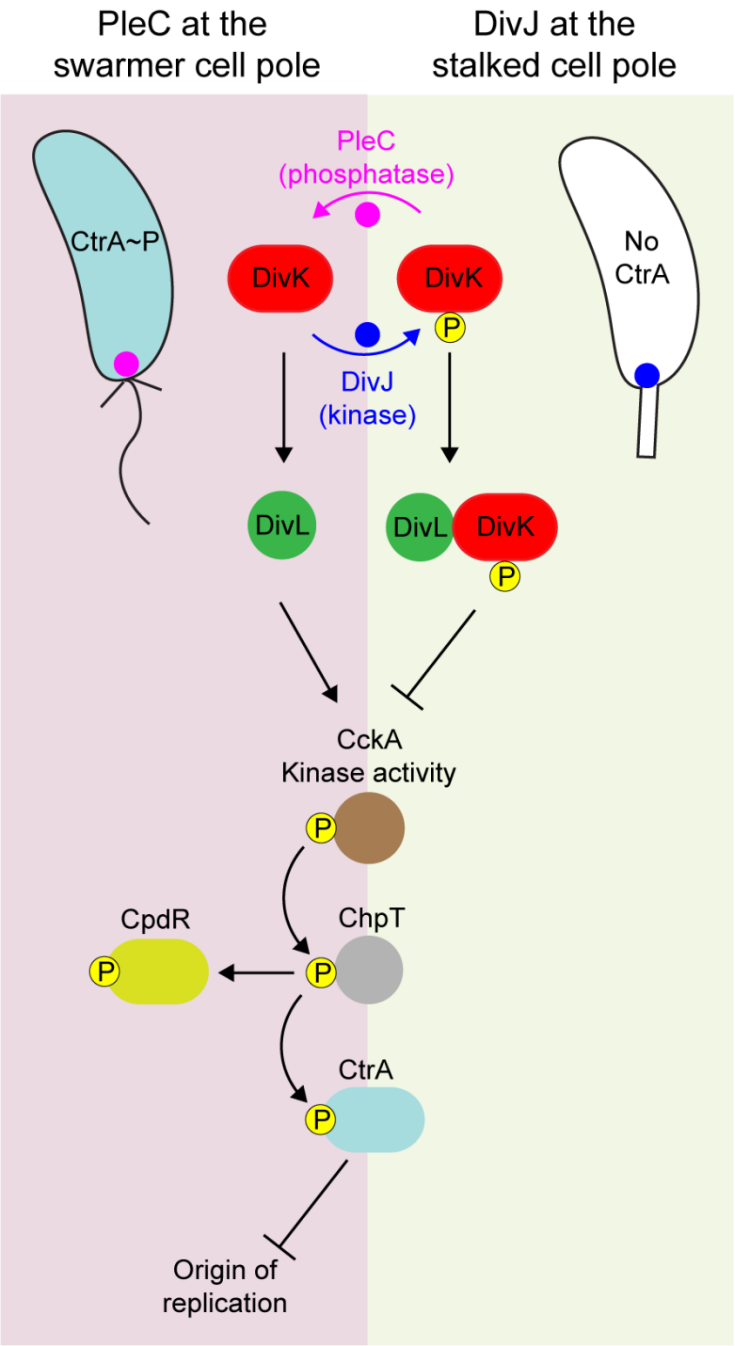


Figure 1.4: Spatial and temporal regulation of the cell cycle in *Caulobacter crescentus* by TCS. *C. crescentus* dimorphic cell cycle consists of a swarmer and a stalked cell. The spatio-temporal regulation that differentiates swarmer and stalked cells is regulated by the master cell-cycle RR CtrA. In the swarmer cell pole, the phosphatase PleC dephosphorylates the RR DivK. Free DivL promotes kinase activity of CckA, which phosphorylates the Hpt ChpT and this phosphoryl group is transferred to RRs CpdR and

Figure 1.4: Spatial and temporal regulation of the cell cycle in *Caulobacter crescentus* by TCS (continued): CtrA. CpdR~P prevents CtrA proteolysis and CtrA~P inhibits the origin of replication. In a stalked cell pole, the kinase DivJ phosphorylates DivK and DivK~P forms a complex with DivL. DivL-DivK~P complex promotes CckA phosphatase activity, which leads to CtrA dephosphorylation and degradation.

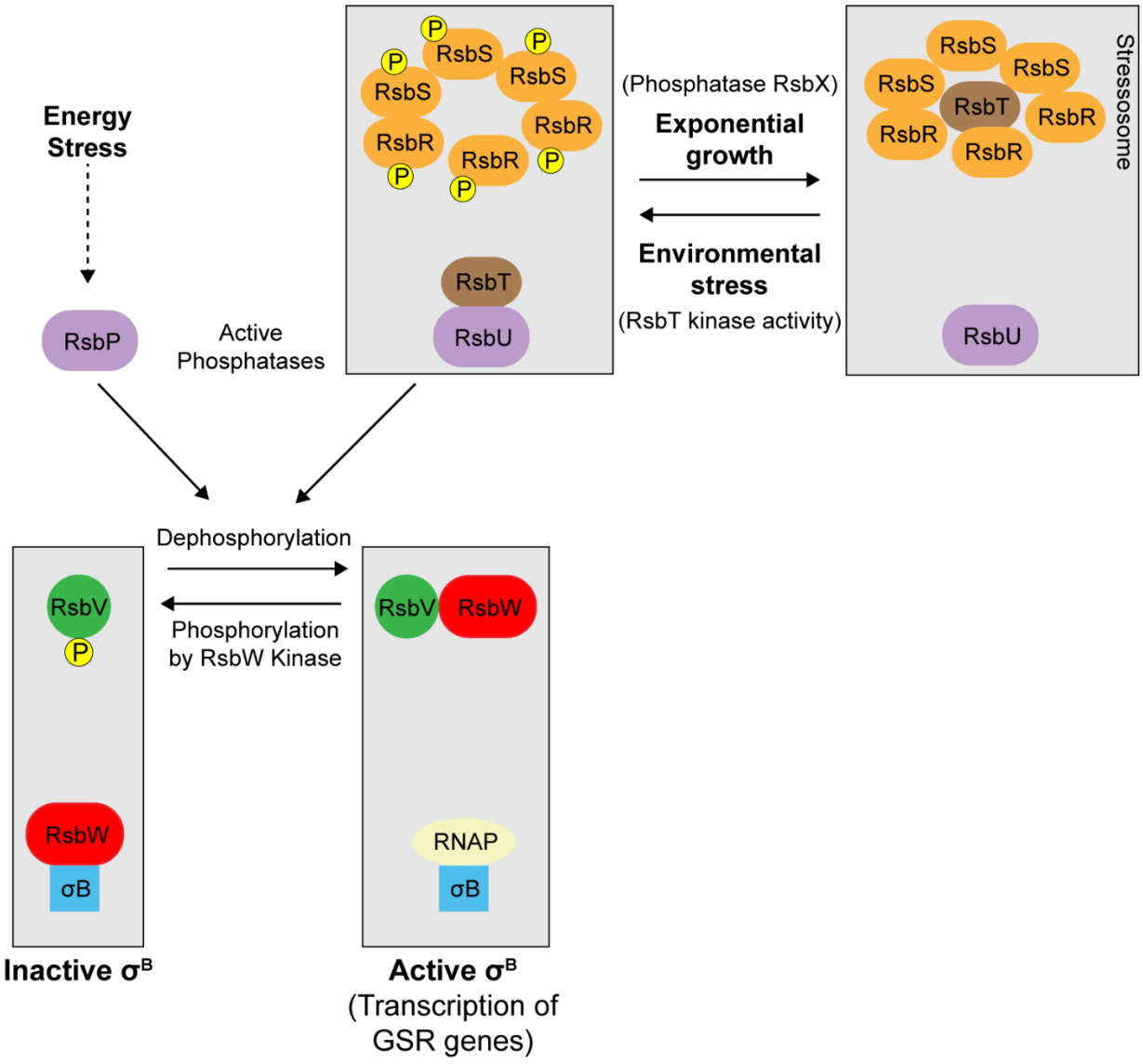


Figure 1.5: Regulation of the General Stress Response (GSR) in *Bacillus subtilis*. σ^B is an alternative sigma factor that regulates transcription of genes involve in stress tolerance and resistance. Energy and environmental stresses positively regulate σ^B activity through a partner switch mechanism that results in unphosphorylated anti-anti- σ^B RsbV and release of σ^B by the anti-σ^B RsbW. Energy stress regulation employs the phosphatase RsbP to dephosphorylate RsbV. In the other hand, environmental stresses employ a high-order complex of kinases known as the stressosome. Under stress conditions, release of RbsT from the stressosome leads to activation of the phosphatase RsbU, which dephosphorylates RsbV. Figure adapted from (Hecker et al., 2007).

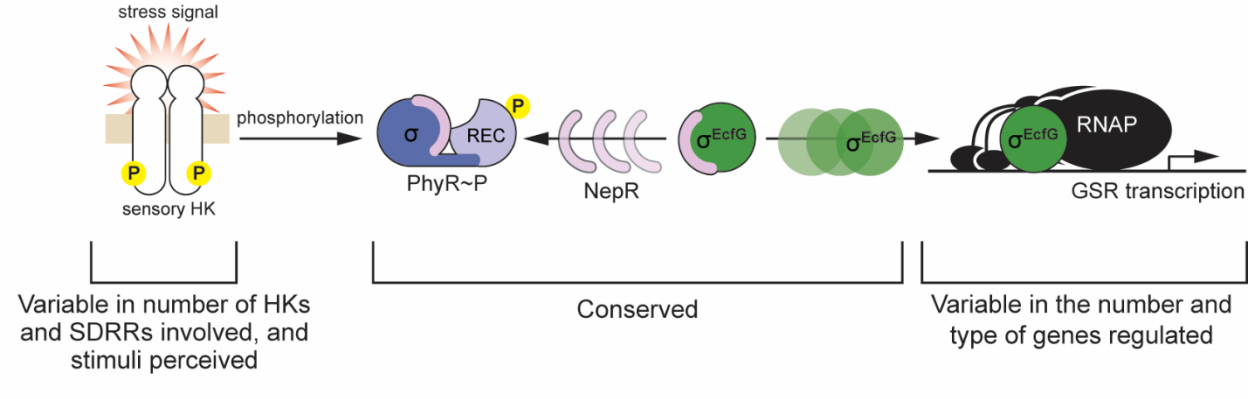


Figure 1.6: Regulation of the General Stresss Response (GSR) in *Alphaproteobacteria*. Upon phosphorylation of the response regulator PhyR, NepR switches partners leaving σ^{EcfG} free to interact with RNA polymerase core. This partner switch mechanism is well conserved throughout *Alphaproteobacteria.* The number of histidine kinases (HKs) and single domain response regulators (SDRRs) that modulate PhyR phosphorylation varies between species. Figure adapted from (Fiebig et al., 2015).

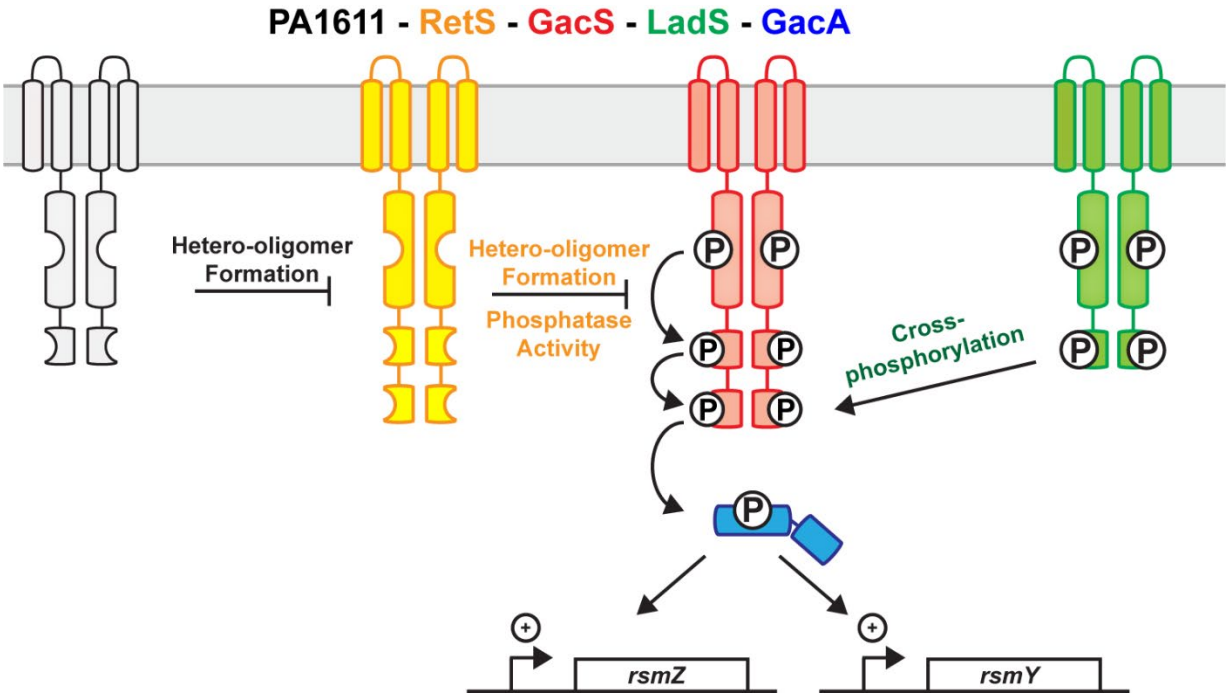


Figure 1.7: Biofilm regulation by atypical heteromeric interactions between histidine kinases in *Pseudomonas aeruginosa*. Phosphorylation of the response regulator GacA is directly or indirectly regulated by four hybrid histidine kinases. GacA is positively regulated by GacS, LadS, and PA1611 by different mechanisms. On the other hand, GacA is negatively regulated by RetS by the formation of hetero-oligomers with GacS and by RetS phosphatase activity towards GacS. Figure adapted from (Willett and Crosson, 2017).

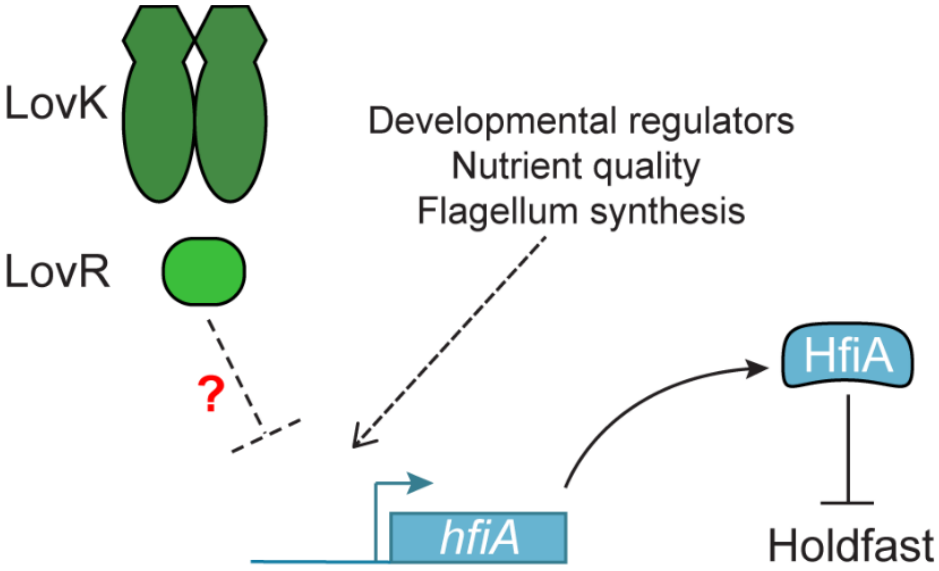


Figure 2.1: Transcriptional regulation of *hfiA*. Transcription of the holdfast inhibitor, *hfiA*, is regulated by developmental factors (Fiebig et al., 2014), flagellum synthesis (Berne et al., 2018), and environmental response systems including the LovK/LovR two-component system (Fiebig et al., 2014). I sought to understand the mechanism of *hfiA* transcriptional regulation by LovK/LovR.

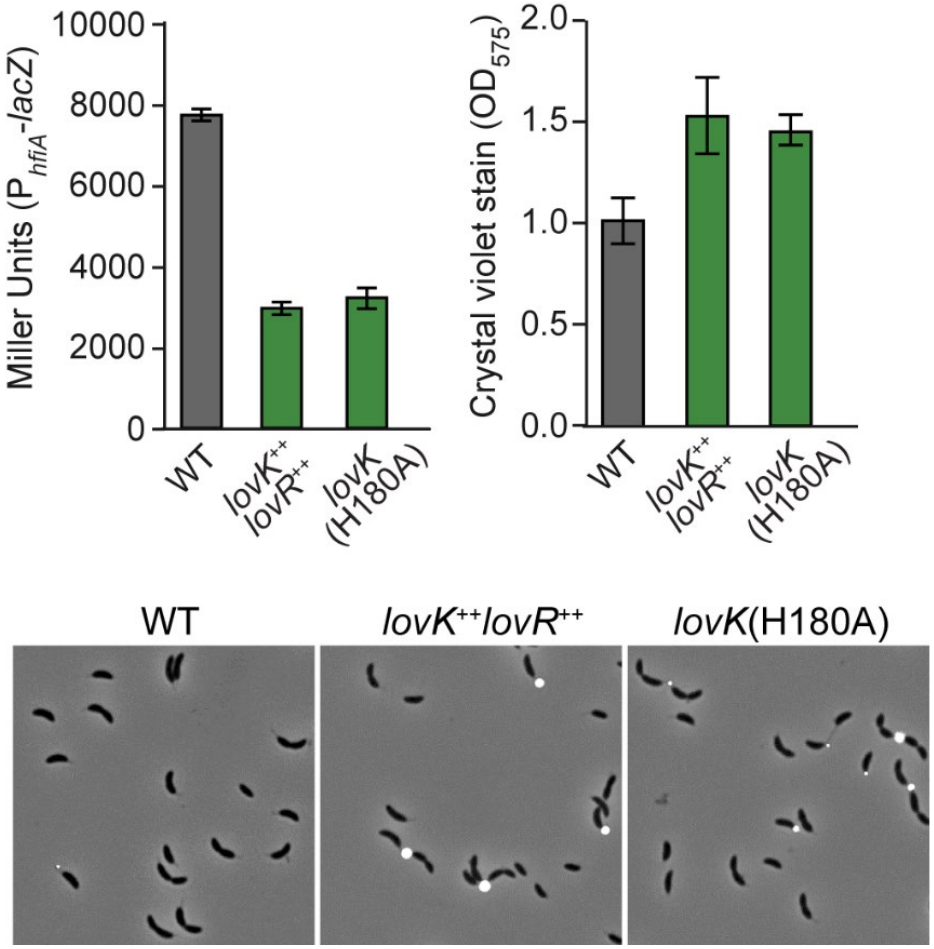


Figure 2.2: *lovK/lovR* regulation of *hfiA* expression and surface adhesion. Characterization of *hfiA* expression, surface attachment and holdfast development in wild-type (WT), *lovK/lovR* overexpression (*lovK*$^{++}$*lovR*$^{++}$) and *lovK*(H180A) strains. Left top: β-galactosidase activity in strains carrying a P_{hfiA}-*lacZ* transcriptional fusion. Right top: Crystal violet staining of surface attached cells after growth in polystyrene wells. Bottom: Representative images of cultures where the holdfast was stained (white) with fluorescent Wheat Germ Agglutinin. Strains were grown in defined M2X medium. Bars represent mean ± s.d.; n=6 for β-galactosidase assay and n=8 for crystal violet stain.

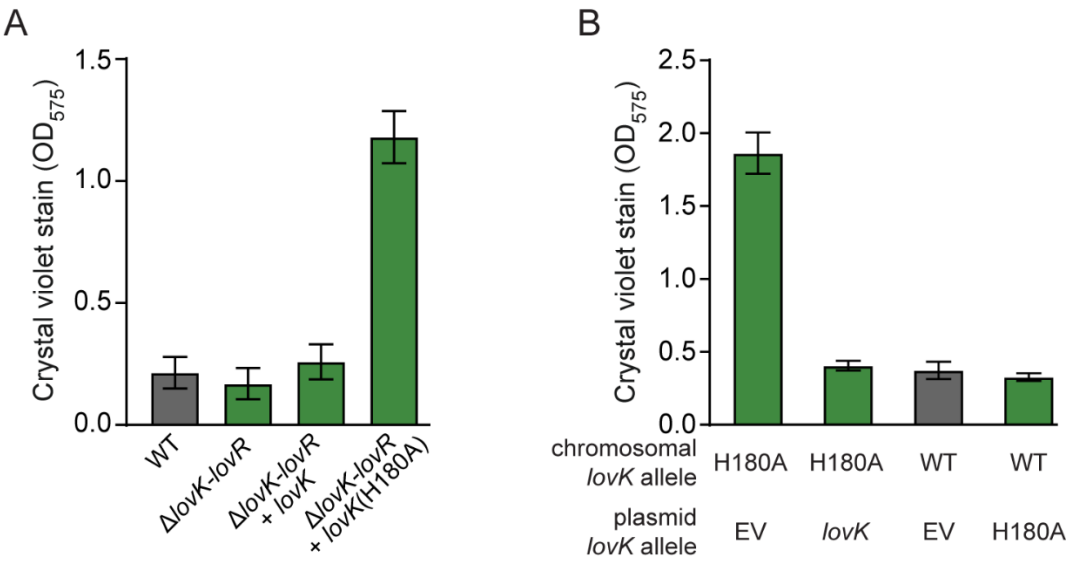
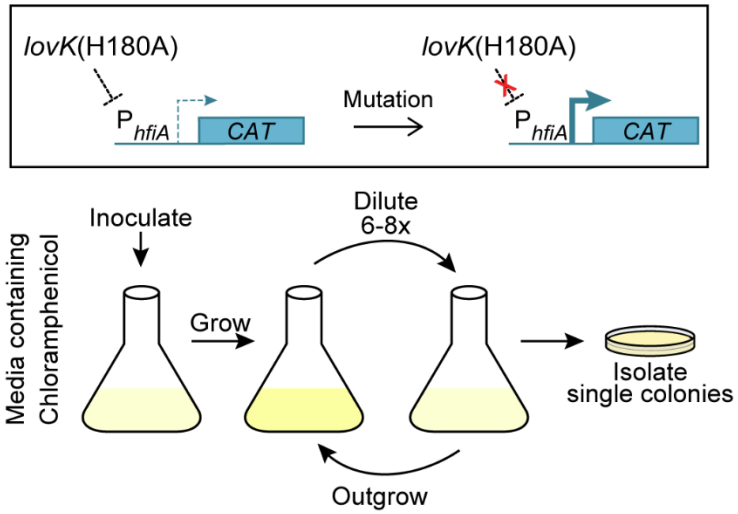


Figure 2.3: Genetic analysis of *lovK* mediated regulation of surface adhesion. A) Strains bearing a deletion of the *lovK-lovR* locus were complemented with a wild-type *lovK* or *lovK*(H180A) allele. Surface attachment to polystyrene plates was measured by crystal violet stain. Cultures were grown in M2X medium. Data are representative of at least three independent experiments. Bars represent mean ± s.d.; n=7. **B)** Surface adhesion (crystal violet stain) measured in different genetic backgrounds. Cultures were grown in M2X medium. Data are representative of at least three independent experiments. Bars represent mean ± s.d.; n=8.

A



B

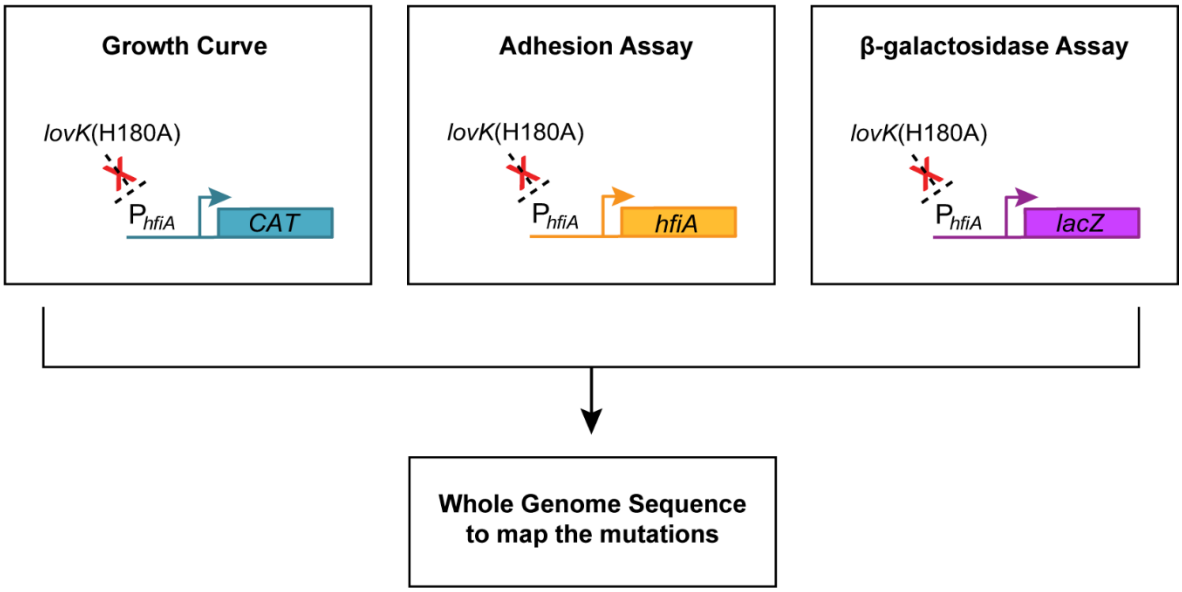


Figure 2.4: Genetic selection designed to identify regulators that function with *lovK*(H180A) to repress *hfiA*. A) A strategy to select for spontaneous mutants that disrupt *lovK*(H180A) repression of the *hfiA* promoter. The P_{hfiA}-*CAT* transcriptional fusion allows for selection of mutants with increased *hfiA* promoter activity. Specifically, cultivation in M2X medium supplemented with chloramphenicol allows for selection of enhanced expression of chloramphenicol acetyltransferase (*CAT*) from P_{hfiA}-*CAT*. **B)** Spontaneous mutants isolated from the genetic screen were individually tested in several secondary assays.

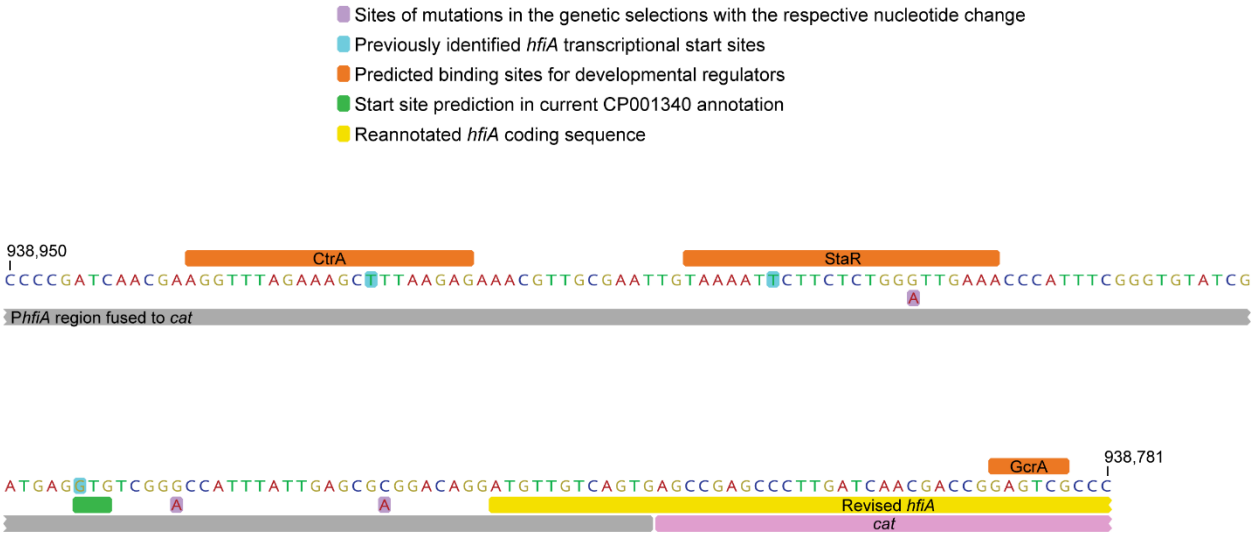


Figure 2.5: Diagram of the *hfiA* promoter region. Predicted DNA-binding regions of several developmental regulators in orange (Fiebig et al., 2014). Predicted start codon corresponding to original *hfiA* annotation in green; the revised *hfiA* reading frame is in yellow (Fiebig et al., 2014). Experimentally identified *hfiA* transcriptional start sites in light blue (Fiebig et al., 2014) (Zhou et al., 2015). The central portion of the P*hfiA-cat* fusion is shown with P_{hfiA} in gray while and *cat* in pink. Sites of spontaneous mutations identified in the genetic selections with the respective nucleotide change in light purple. Genome coordinates are based on the NA1000 genome sequence (CP001340; NC_011916).

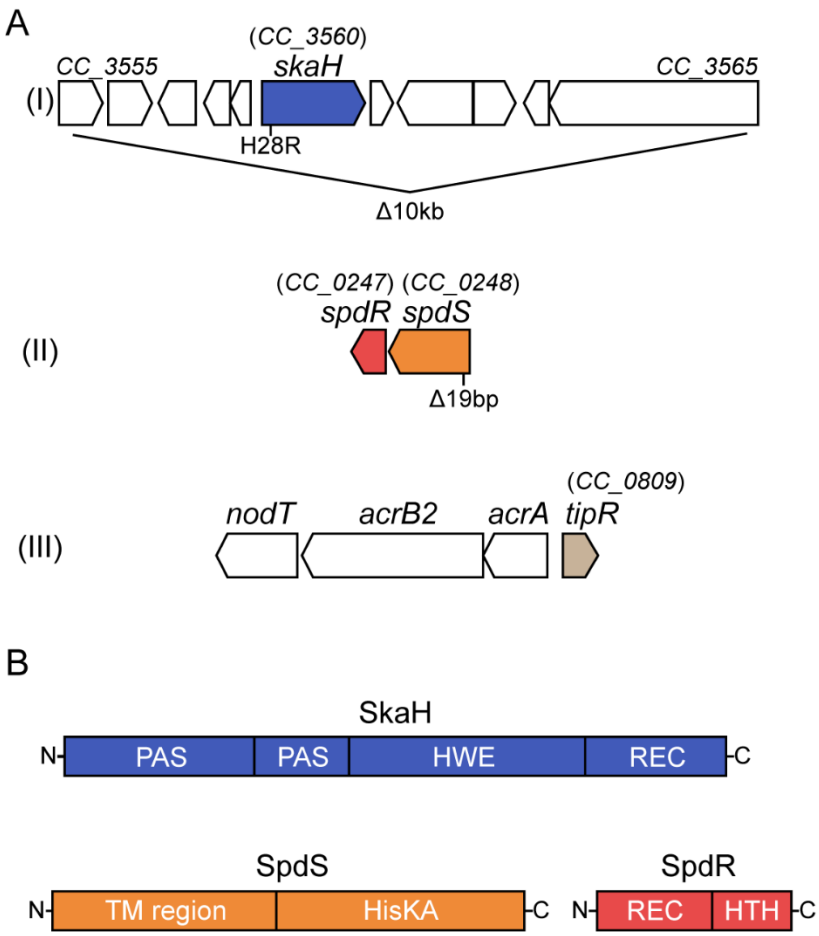


Figure 2.6: Genomic lesions identified in the genetic selection. A) Mutations in three genomic regions identified in the selection with specific lesions marked below the genetic loci. Lesions at the *acr* locus confer increased chloramphenicol resistance irrespective of *lovK* (Fig 2.7); the mutant alleles identified at this locus are not indicated in the diagram but are listed in Table S1. **B)** Predicted domain organization of the TCS proteins SkaH, SpdS and SpdR. PAS – sensory domain, HWE and HisKA – histidine kinase domains, REC – receiver domain, TM – transmembrane, HTH – helix-turn-helix DNA-binding domain.

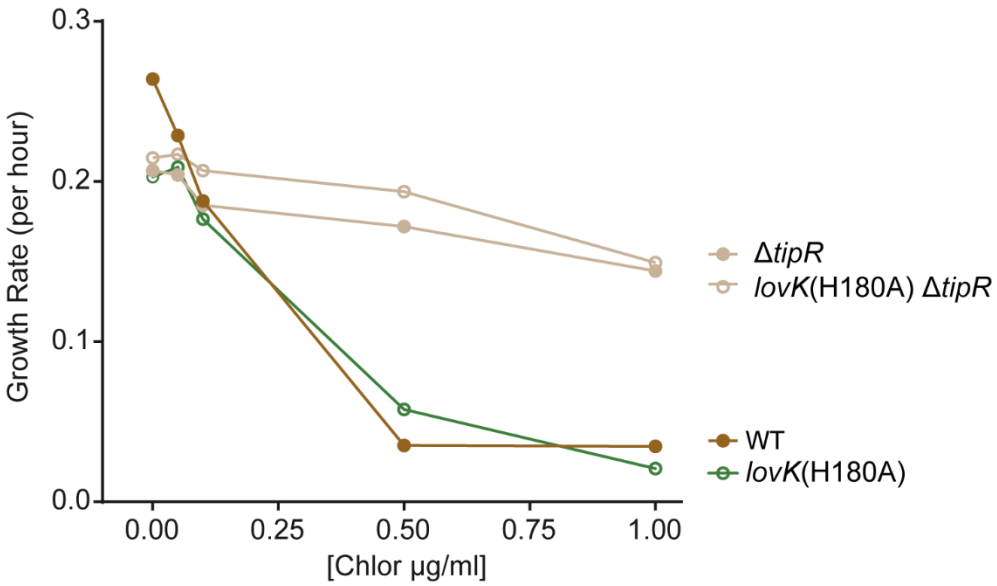


Figure 2.7: Strains carrying a *tipR* in-frame deletion are resistant to chloramphenicol. Growth rate per hour at different chloramphenicol concentrations for wild-type (WT), *lovK*(H180A), Δ*tipR*, and *lovK*(H180A) Δ*tipR* strains. Cultures were grown in M2X medium supplemented to the final chloramphenicol concentrations indicated on the x-axis. One biological replicate per condition per strain.

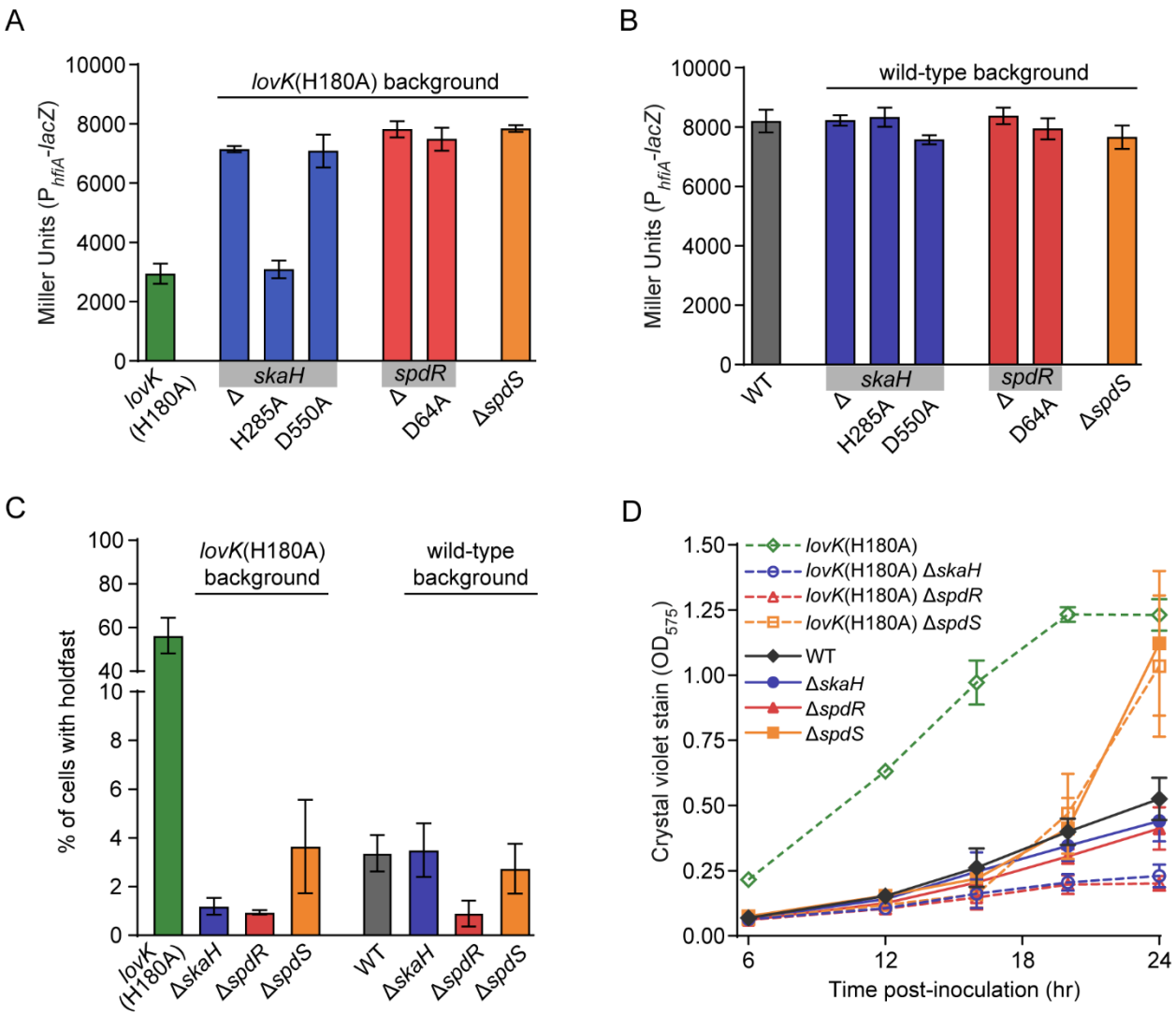


Figure 2.8: TCS genes identified in the selection are necessary for *lovK*(H180A) to repress *hfiA* transcription. A and B) P_{hfiA}-*lacZ* activity in strains bearing in-frame deletions (Δ) or the indicated substitutions of the conserved phosphorylation sites of *skaH, spdS* and *spdR* in strains with a chromosomal *lovK*(H180A) allele (A) and in strains with a wild-type *lovK* allele (B). Bars represent mean ± s.d.; n=5. **C)** Holdfast was stained with fluorescent Wheat Germ Agglutinin in strains bearing deletions in *skaH, spdS* and *spdR* in a *lovK*(H180A) or a wild-type background. Percentage of cells bearing a holdfast was quantified by microscopy. Bars represent mean ± s.d.; n=4. **D)** Attachment of cells to polystyrene plates measured by crystal violet stain. Strains bear in-frame deletions of *skaH, spdS* or *spdR* in a *lovK*(H180A) background and in a wild-type background. Bars represent mean ± s.d.; n=6 for each strain at each timepoint.

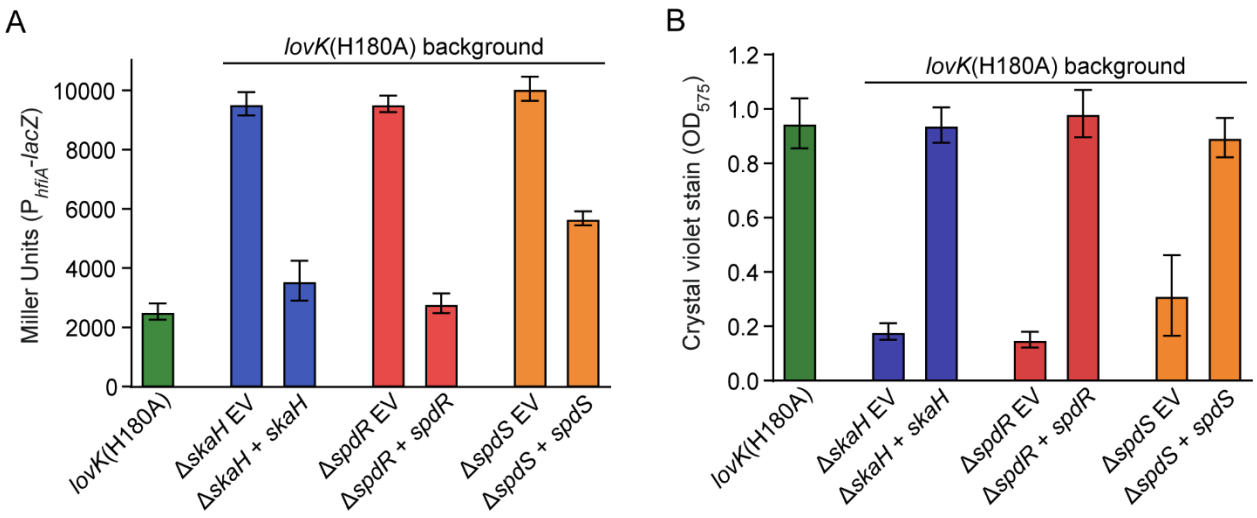


Figure 2.9: Complementation of in-frame deletions in strains bearing the *lovK*(H180A) allele. Strains bearing in-frame deletions of *skaH, spdS* or *spdR* in a *lovK*(H180A) background were transformed with either pMT680 plasmids that express the corresponding disrupted gene from a xylose inducible promoter or pMT680 as an empty vector (EV) control. **A)** β-galactosidase activity from the P_{hfiA}-*lacZ* transcriptional fusion was measured in each strain. Cultures were grown in M2X medium to OD_{660nm} of 0.05 - 0.15. Presented data are representative of at least three independent experiments. Bars represent mean ± s.d.; n=6. **B)** Surface attachment of cells grown in polystyrene plates was measured by crystal violet stain. Cultures were grown in M2X medium for 16 hrs post-inoculation. Data are representative of three independent experiments. Bars represent mean ± s.d.; n=9.

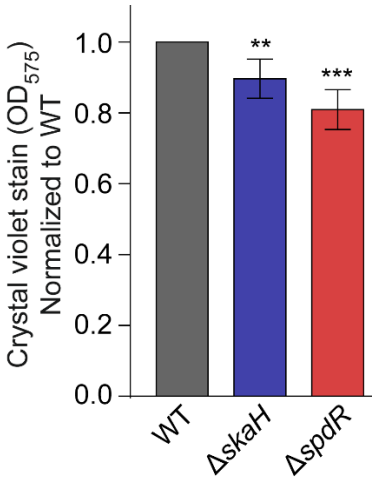


Figure 2.10: Deletion of *skaH* or *spdR* decreases surface attachment. Crystal violet stain from cells growing in polystyrene plates from seven different experiments was collected. The mean Δ*skaH* and Δ*spdR* crystal violet stain for each experiment was normalized to the mean wild-type stain of the respective day. One-way ANOVA followed by Dunnett's multiple comparisons test was performed using GraphPad Prism version 8.0.0. to compare wild-type to the mutant strains. ** $P<0.005$, *** $P<0.0005$

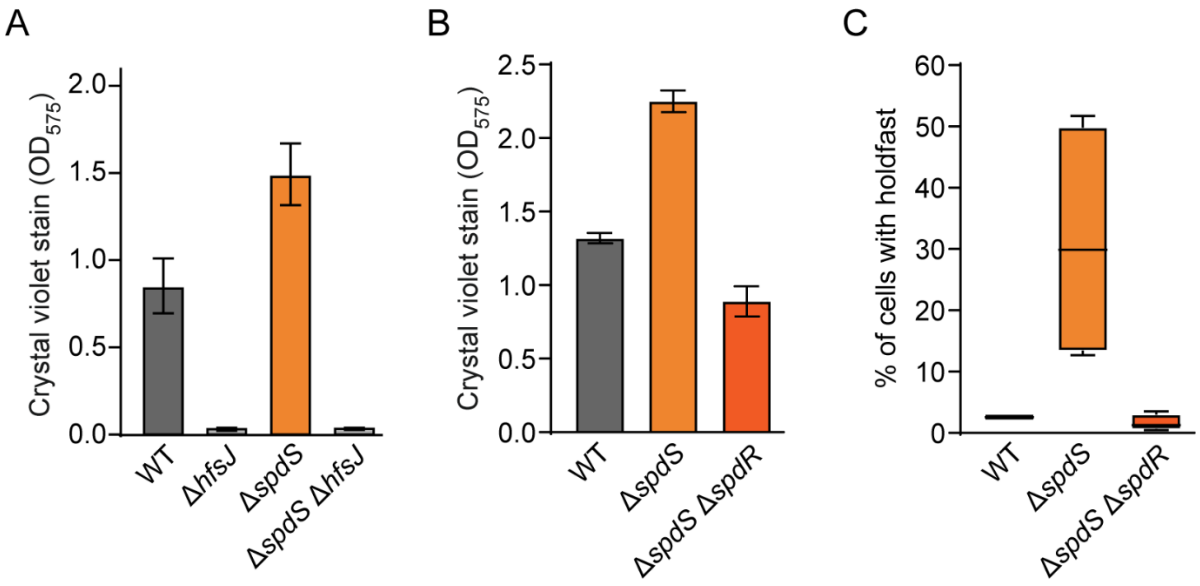


Figure 2.11: Holdfast synthesis and *spdR* are required for the Δ*spdS* stationary phase hyper-attachment phenotype. A) Surface attachment to polystyrene plates was measured by crystal violet stain in wild-type (WT), Δ*spdS*, Δ*hfsJ* and Δ*spdS*Δ*hfsJ* strains. Cultures were grown in M2X medium until stationary phase (24 hrs post-inoculation). Data are representative of at least three independent experiments. Bars represent mean ± s.d.; n=12. **B)** Surface attachment to polystyrene plates was measured by crystal violet stain in wild-type (WT), Δ*spdS*, and Δ*spdS*Δ*spdR* strains. Cultures were grown in M2X medium until stationary phase (24 hrs post-inoculation). Data are representative of at least three independent experiments. Bars represent mean ± s.d.; n=12. **C)** Holdfast was stained with fluorescent Wheat Germ Agglutinin in wild-type (WT), Δ*spdS*, and Δ*spdS*Δ*spdR* strains. Percentage of cells bearing a holdfast was quantified by microscopy. Cultures were grown in polystyrene plates with M2X medium until stationary phase (24 hrs post-inoculation). Bars represent mean ± s.d.; n=4.

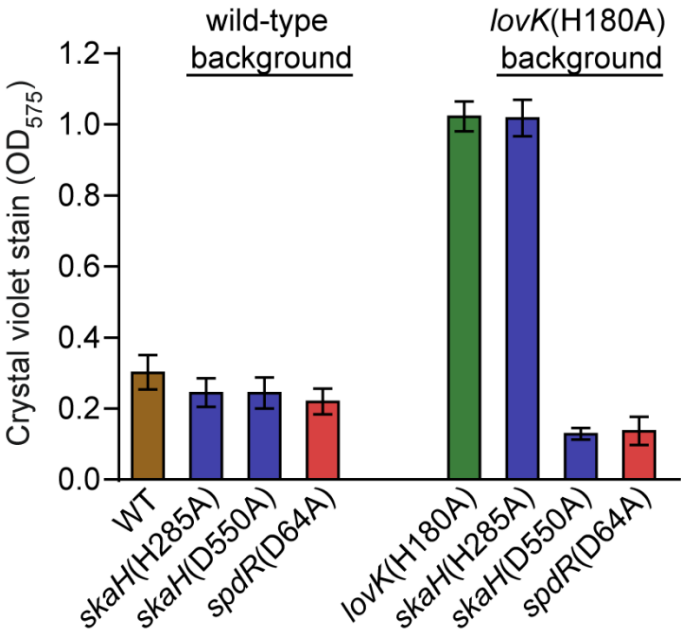


Figure 2.12: Surface attachment of strains bearing mutations in the TCS protein phosphorylation sites. Strains bearing point mutations in the conserved phosphorylation sites of *skaH, spdS* or *spdR* in either a wild-type (WT) or *lovK*(H180A) background were grown in M2X medium in polystyrene plates for 16 hrs post inoculation. Surface attachment was measured by crystal violet stain. Data presented are representative of three independent experiments. Bars represent mean ± s.d.; n=8.

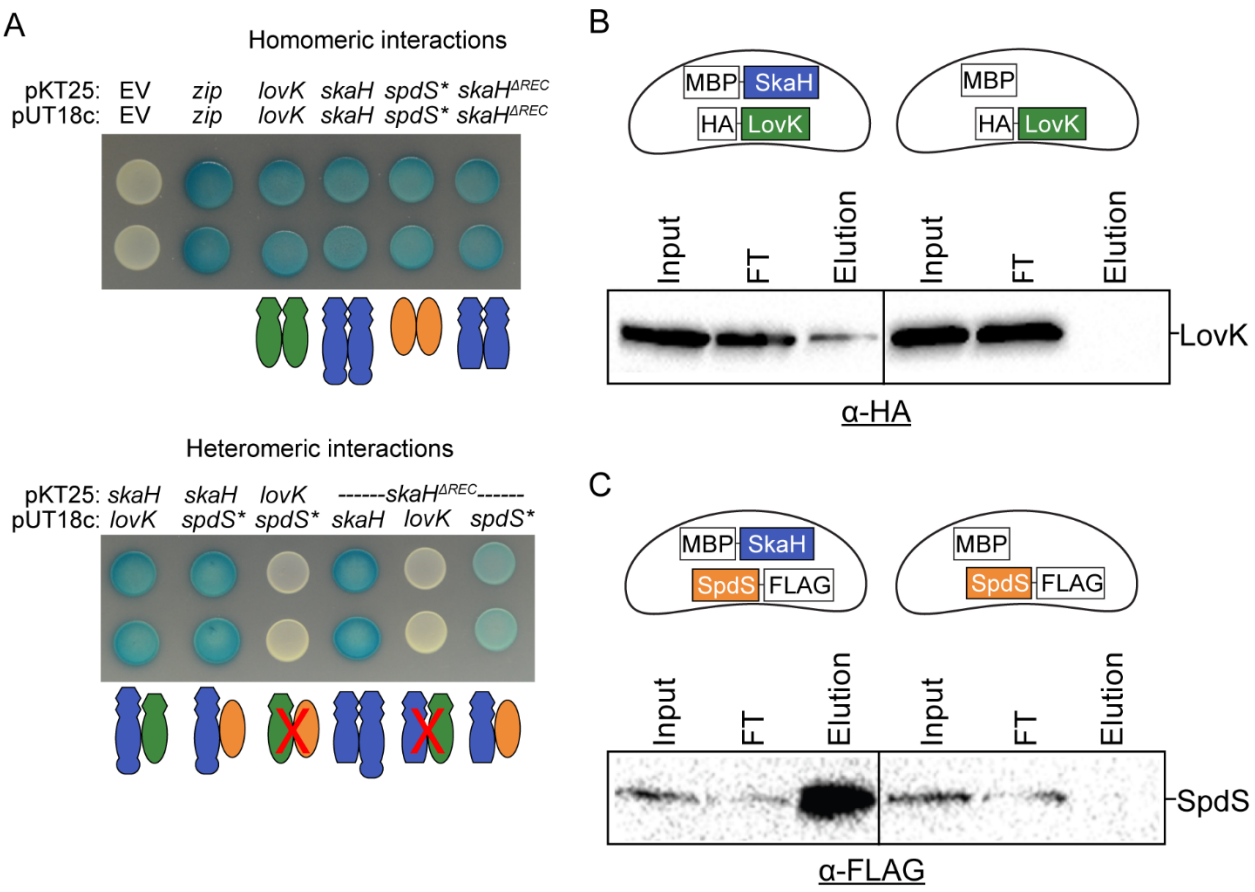


Figure 2.13: SkaH physically interacts with LovK and SpdS *in vitro* and *in vivo*. A) Bacterial two-hybrid (BTH) experiments to assess interactions between histidine kinase fusions to a split adenylate cyclase. Fusions with *spdS* lack the transmembrane domain, notated *spdS**. *skaH*$^{\Delta REC}$ is a truncated version of SkaH that lacks the REC domain. Two biological replicates are shown for each co-expression combination. Protein-protein interaction between fusions reconstitutes the split adenylate cyclase encoded on pKT25 and pUT18c, and results in a blue color on agar plates containing x-gal. Strains expressing fusions that do not interact appear white. Zip = positive control. Histidine kinase protein-protein interaction models are schematized below each BTH assay as 1:1 interactions. However, the actual oligomeric state of detected interactions is not known. **B and C)** Co-purification experiments in which MBP is affinity purified with amylose resin and co-purifying kinases are detected by Western blot. The fusion proteins co-expressed in each experiment are schematized in the cells above each blot. Representative immunoblot against the HA and FLAG epitope tags are shown in B and C respectively. Input = clarified lysate, FT = Flow through.

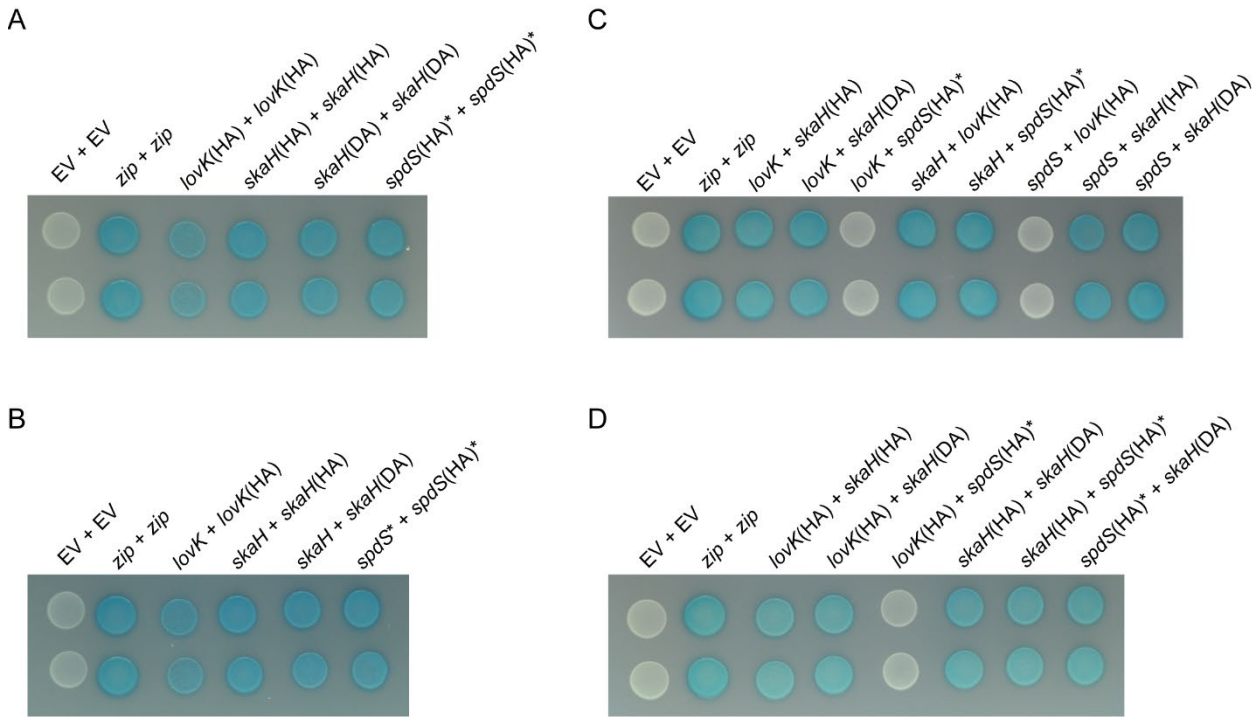


Figure 2.14: SkaH heteromeric interactions with LovK and SpdS do not require the conserved sites of phosphorylation. A-D) Bacterial two-hybrid (BTH) experiments to assess interactions between mutated histidine kinase fusions to a split adenylate cyclase. *lovK*(HA) = *lovK*(H180A), *skaH*(HA) = *skaH*(H285A), *skaH*(DA) = *skaH*(D550A), *spdS*(HA)* = *spdS*(H248A)*. Fusions with *spdS* lack the transmembrane domain, notated *spdS** or *spdS*(HA)*. Zip = positive control. Two biological replicates are shown for each co-expression combination. Protein-protein interaction of fusions reconstitutes the split adenylate cyclase encoded on pKT25 and pUT18c, and results in a blue color on agar plates containing x-gal. Strains expressing fusions that do not interact appear white.

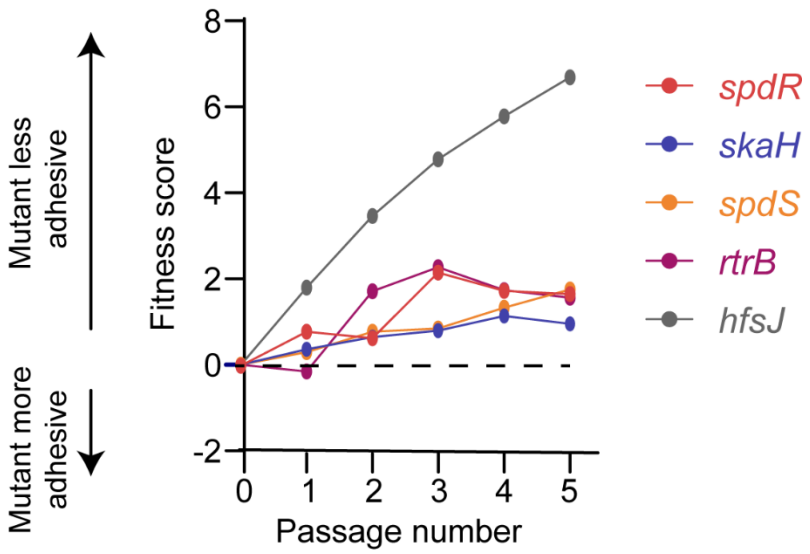


Figure 2.15: Global Tn-seq approach provides an additional line of evidence that genes identified in this study play a role in regulating adhesion in wild-type cells. Extracted fitness profiles of mutants characterized in a genome-wide attachment screen (Hershey et al., 2019). Briefly, a *C. crescentus* transposon mutant library was grown and passaged in complex PYE medium in the presence of cheesecloth to provide surface area for cell attachment. Cells were periodically sampled from the broth over several days. Hyper-adhesive mutants are titrated out of the broth by the cheesecloth and, consequently, have negative fitness scores (in which the fitness score is a $\log_2$ ratio of mutant strain abundance relative to the average strain). Conversely, hypo-adhesive mutants are enriched in the broth resulting in positive fitness scores. Strains bearing transposon insertions in TCS genes identified in our study have positive fitness scores. *hfsJ* mutants, which do not produce holdfast and are non-adhesive, are presented as a reference. *rtrA* was not captured in this analysis; the library contained insufficient Tn-insertions in this gene.

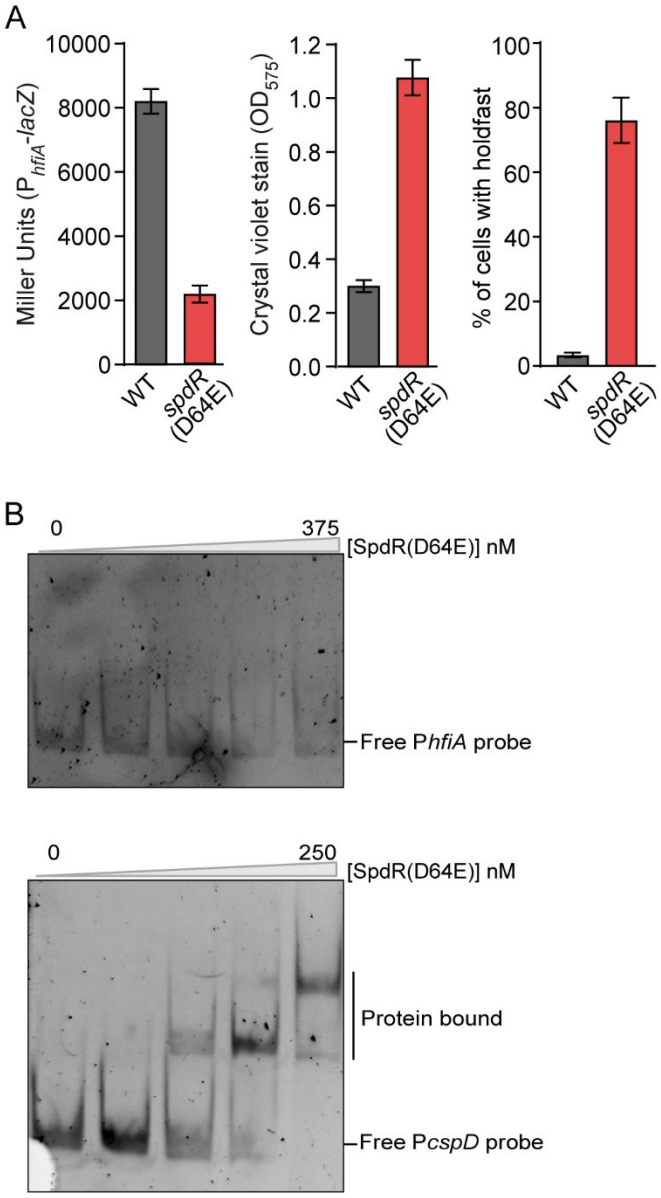


Figure 3.1: SpdR is an indirect repressor of *hfiA* trancription. A) β-galactosidase activity of P_{hfiA}-*lacZ* transcriptional fusion (left), attachment of cells to polystyrene measured by crystal violet stain (center) and percentage of cells bearing a holdfast by fluorescent Wheat Germ Agglutinin staining of the holdfast (right) for wild-type (WT) and *spdR*(D64E) strains. Data represent mean ± s.d.; n=4-8. **B)** Electrophoretic mobility shift assay using recombinant His_6-SpdR(D64E) and a fluorescently labeled *hfiA* promoter probe (top) or *cspD* promoter probe (bottom). His_6-SpdR(D64E) protein concentrations when incubated with *hfiA* probe were 0, 50, 125, 250, 375 nM, and when incubated with *cspD* probe were 0, 25, 50, 125, 250 nM. Experiments were conducted three times.

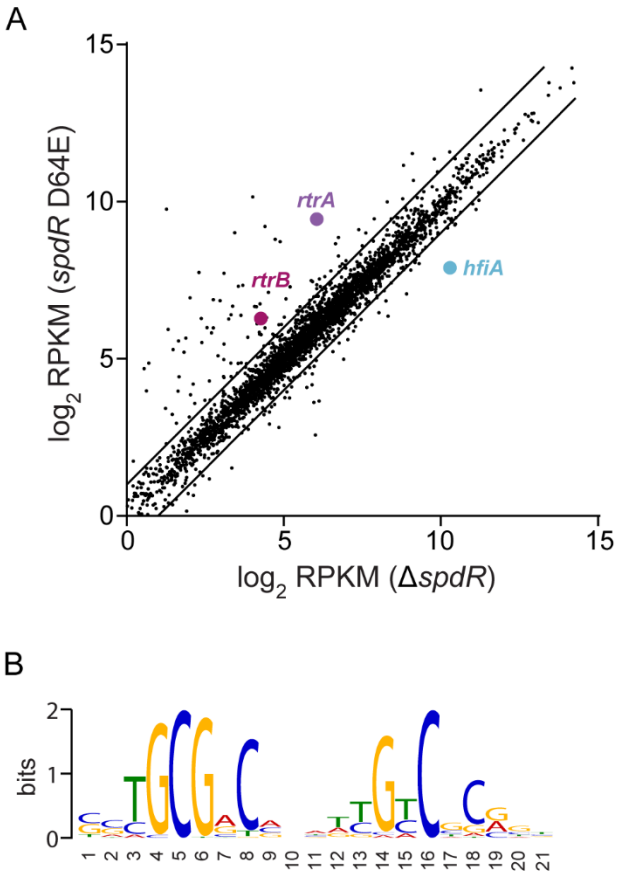


Figure 3.2: Transcriptomic analysis of the SpdR regulon. A) RNA-Seq analysis of transcript levels in *spdR*(D64E) and Δ*spdR* strain. Log_2(mean RPKM) for genes from each strain are plotted against each other where each gene is a dot. Genes of interest (*hfiA, rtrA, rtrB*) are marked. Lines indicate two-fold change cutoffs. See Methods for details. **B)** Promoter motif identified by MEME in genes positively regulated by *spdR*(D64E) compared to Δ*spdR*.

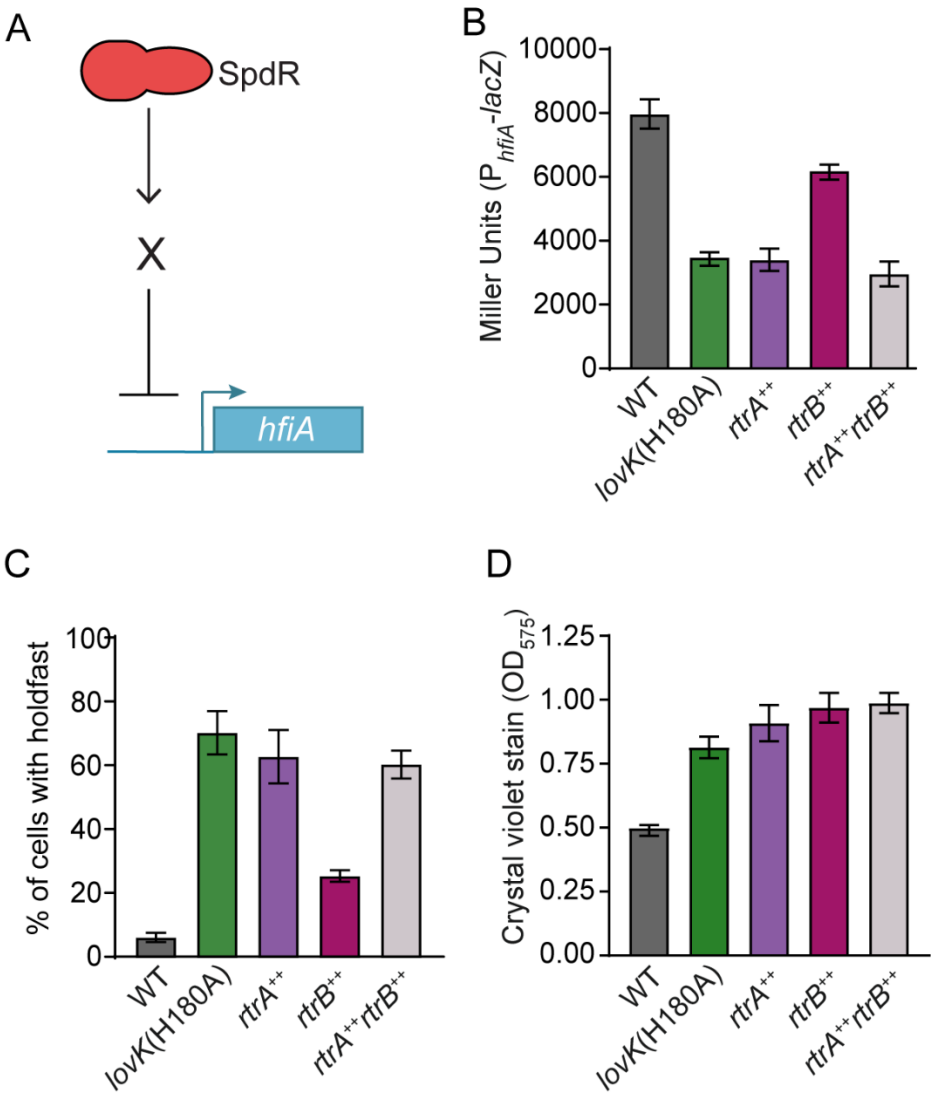


Figure 3.3: *rtrA* (*CC_3164*) and *rtrB* (*CC_2330*) are regulators of *hfiA* transcription and holdfast development. A) I postulated that transcription factor(s) X, downstream of SpdR, repressed *hfiA* transcription. **B-D)** β-galactosidase activity from P_{hfiA}-*lacZ* transcriptional reporter (B), percentage of cells bearing a holdfast (C), and surface adhesion by crystal violet stain (D) in different genetic backgrounds. Wild-type (WT) and *lovK*(H180A) carry empty vectors pMT680 and pMT585 integrated at the xylose locus as controls. *rtrA*$^{++}$ carries pMT680-*rtrA* and pMT585 empty vector. *rtrB*$^{++}$carries pMT585-*rtrB* and pMT680 empty vector. *rtrA*$^{++}$*rtrB*$^{++}$ carries pMT680-*rtrA* and pMT585-*rtrB*. Genes in both plasmids are expressed from a xylose inducible promoter. Data are representative of at least three independent experiments. Data represent mean ± s.d.; n=6-8.

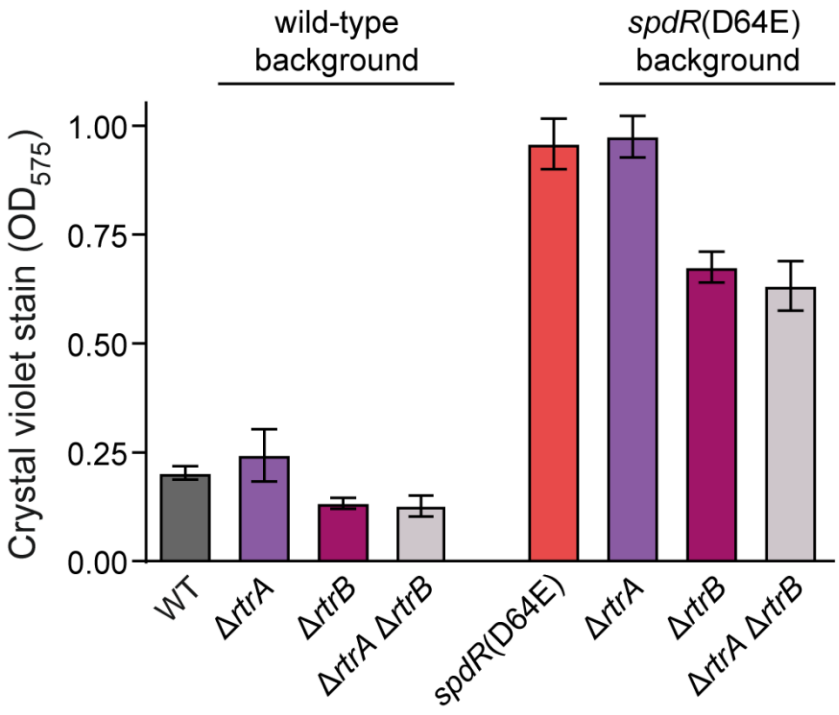


Figure 3.4: *rtrB* (*CC_2330*) is a regulator of surface adhesion under regular growth conditions. Surface attachment of cells grown in polystyrene plates was measured by crystal violet stain in Δ*rtrA*, Δ*rtrB,* and Δ*rtrA*Δ*rtrB* in a wild-type (WT) and *spdR*(D64E) backgrounds. Cultures were grown for 16 hrs post-inoculation in M2X medium. Data presented are representative of at least three independent experiments. Data represent mean ± s.d.; n=9.

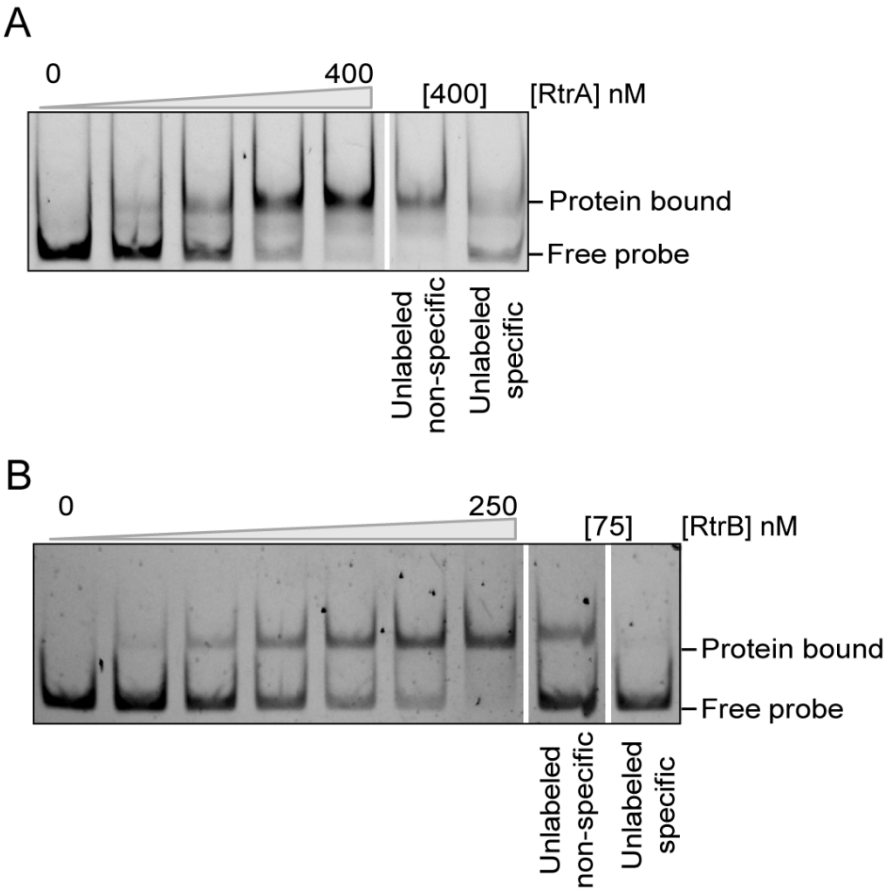


Figure 3.5: *rtrA* and *rtrB* are direct repressors of *hfiA*. A and B) Electrophoretic mobility shift assay of purified RtrA (D) and His_6-RtrB (E) with a fluorescently labeled *hfiA* promoter probe, with increasing concentrations of protein (0, 100, 200, 300, 400 nM for RtrA and 0, 10, 25, 50, 75, 100, 250 nM for His_6-RtrB). Unlabeled specific and non-specific probes are in 10-fold excess of the labeled *hfiA* promoter probe. Blots are representative of at least three independent experiments.

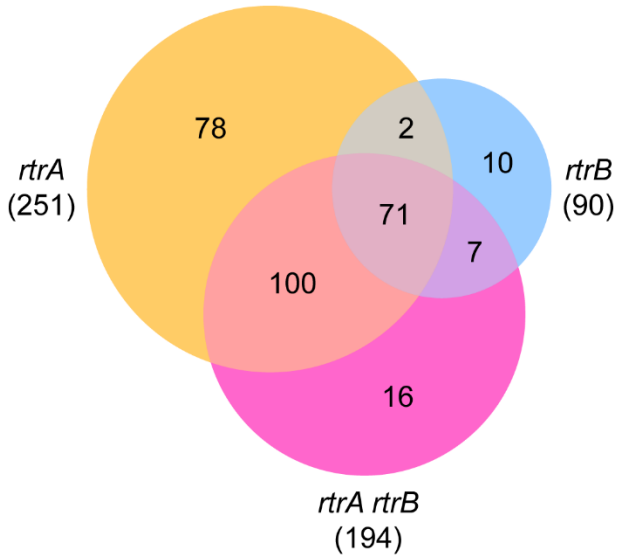


Figure 3.6: Overlap of the RtrA and RtrB regulons. Venn diagram representing the overlap of differentially expressed genes (fold-change > 1.5 and < -1.5 compared to EV control, P-value < 0.001) of between strains overexpressing *rtrA, rtrB* or both *rtrA rtrB.* The full list of genes that met the criteria to be considered differentially expressed for any strain is listed in Table 3.2.

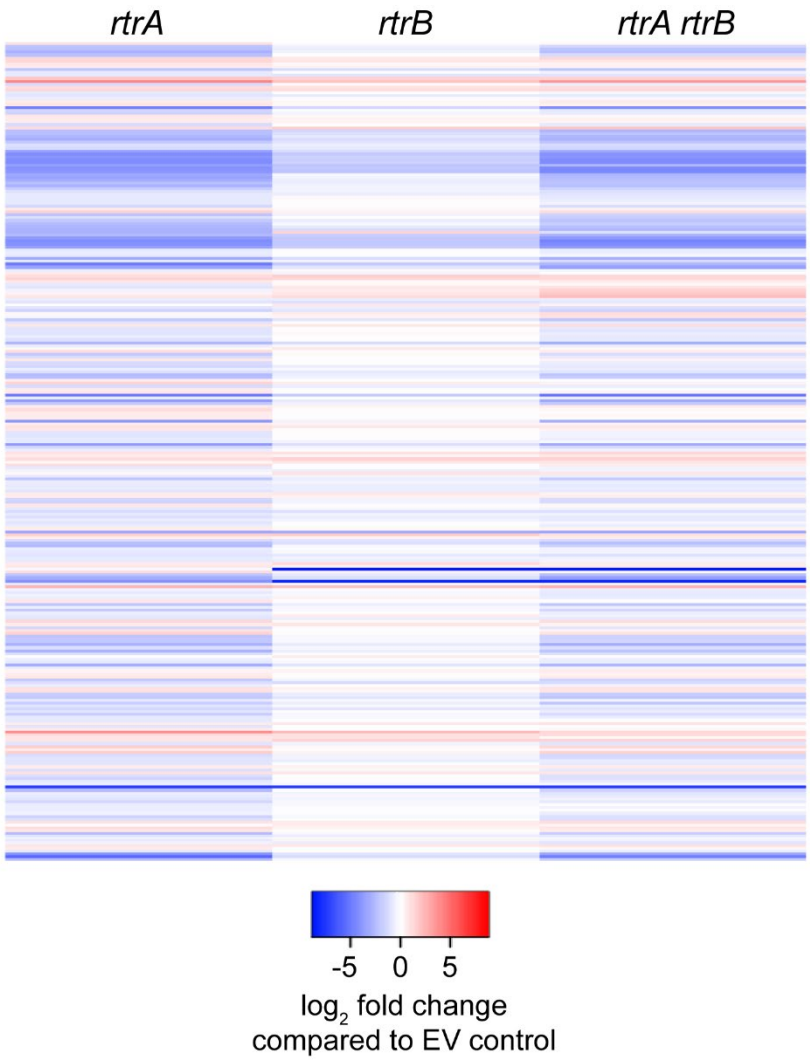


Figure 3.7: The RtrB regulon has lower overall fold changes than the RtrA regulon relative to a vector control strain. Heat map representing the $\log_2$ fold change of each overexpression strain (*rtrA*$^{++}$, *rtrB*$^{++}$, *rtrA*$^{++}$*rtrB*$^{++}$) compared to a wild-type empty vector (EV) control. Genes that show fold-change > 1.5 or < -1.5 relative to EV control, and P-value < 0.001 in at least one overexpression strain are shown. Overall, the fold changes of the *rtrB*$^{++}$ strain are lower than the other strains. Each line represents a gene. The full list of genes in this cluster that met the criteria to be considered differentially expressed for any strain are listed in Table 3.2.

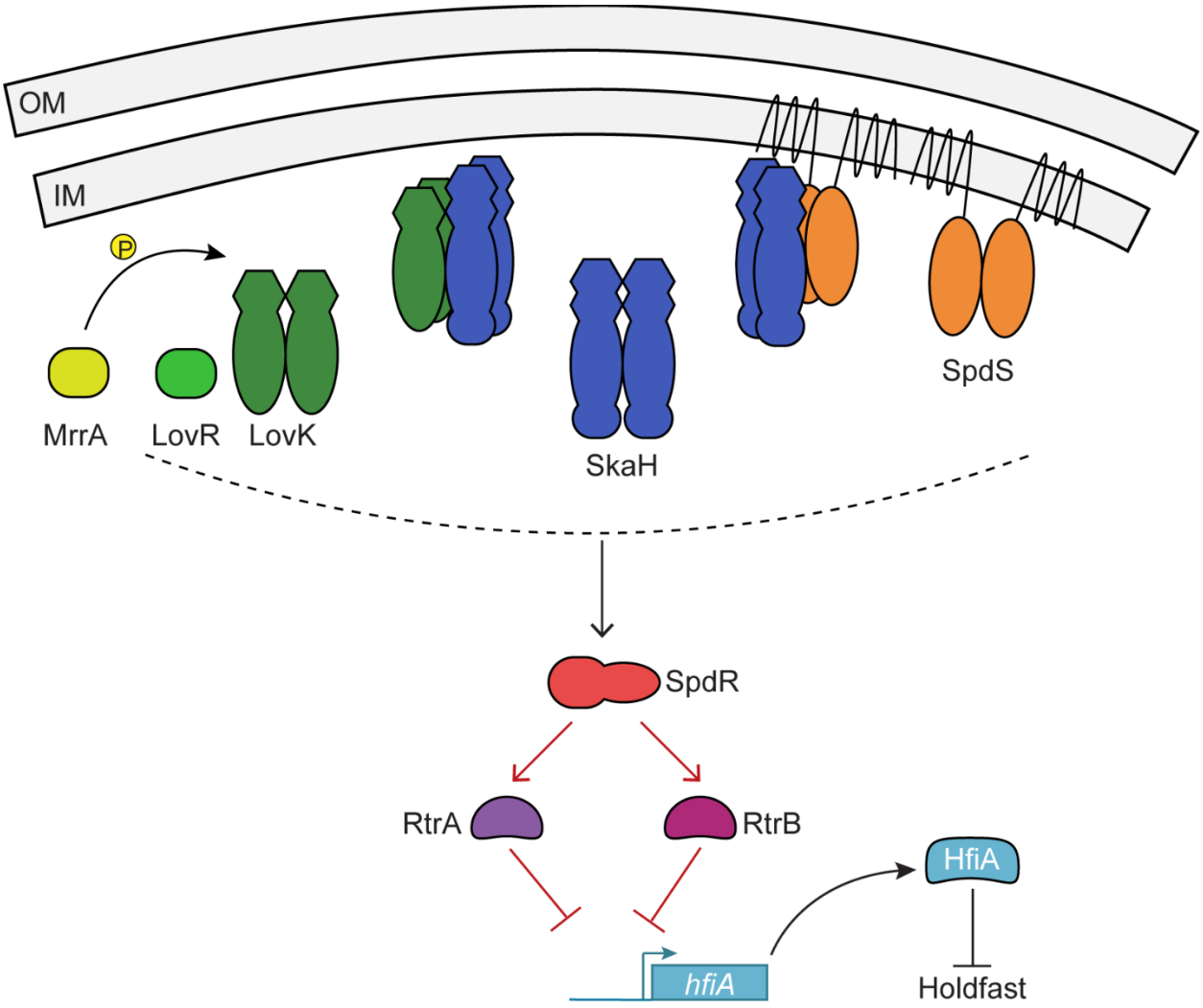


Figure 4.1: Network model of TCS proteins and transcription factors that regulate *hfiA* transcription and surface adhesion. Proposed model in which the hybrid histidine kinase, SkaH, interacts with both LovK and SpdS to mediate *hfiA* transcription. The oligomeric state of these kinase complexes is not known. Interactions and/or phosphoryl transfer between the kinases and response regulators are proposed to collectively regulate SpdR phosphorylation. Phosphorylated SpdR activates transcription of RtrA and RtrB, transcription factors that directly bind to and repress the *hfiA* promoter. The SDRR, MrrA, can transfer a phosphoryl group to LovK and also modulates *hfiA* transcription in a LovK dependent manner (Lori et al., 2018). Red lines indicate direct transcriptional regulation. Dashed line indicates the unknown flow of phosphoryl groups.

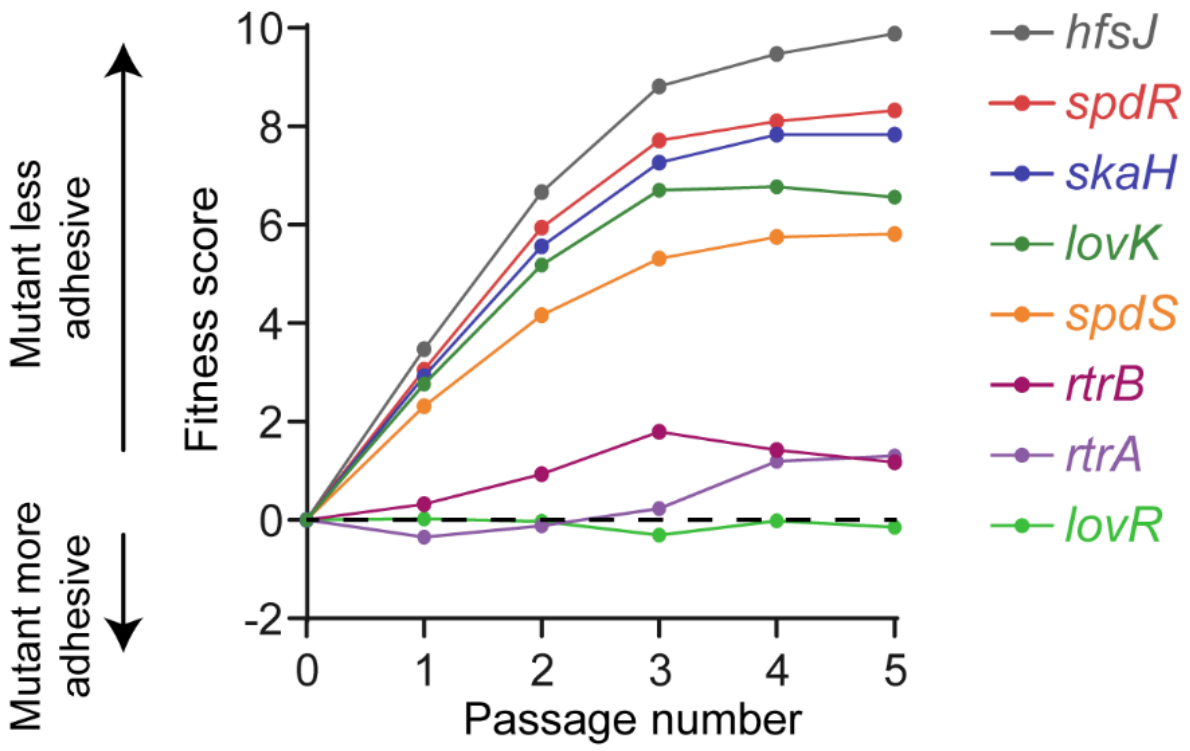


Figure 4.2: Global Tn-seq approach provides additional evidence that genes identified in this study play a role in determining enhanced adhesion in the LovK(H180A) strain. This preliminary BarSeq experiment was performed as in (Hershey et al., 2019), with the exception that I used the Tn-mutagenized *lovK*(H180A) strain and M2X medium. Briefly, a *C. crescentus lovK*(H180A) transposon mutant library was grown and passaged in M2X medium in the presence of cheesecloth to provide surface area for cell attachment. Cells were periodically sampled from the broth over several days. Hyper-adhesive mutants are titrated out of the broth by the cheesecloth and, consequently, have negative fitness scores (in which the fitness score is a $\log_2$ ratio of mutant strain abundance relative to the average strain). Conversely, hypo-adhesive mutants are enriched in the broth resulting in positive fitness scores. Strains bearing transposon insertions in TCS genes and transcription factors identified in our study have positive fitness scores. *hfsJ* mutants, which do not produce holdfast and are non-adhesive, are presented as a reference. *lovR* mutants, have no effect on adhesion in a *lovK*(H180A) background, in concordance with results showing that *lovR* is not required for the *lovK*(H180A) hyper-adhesive phenotype (Fig 2.3A).

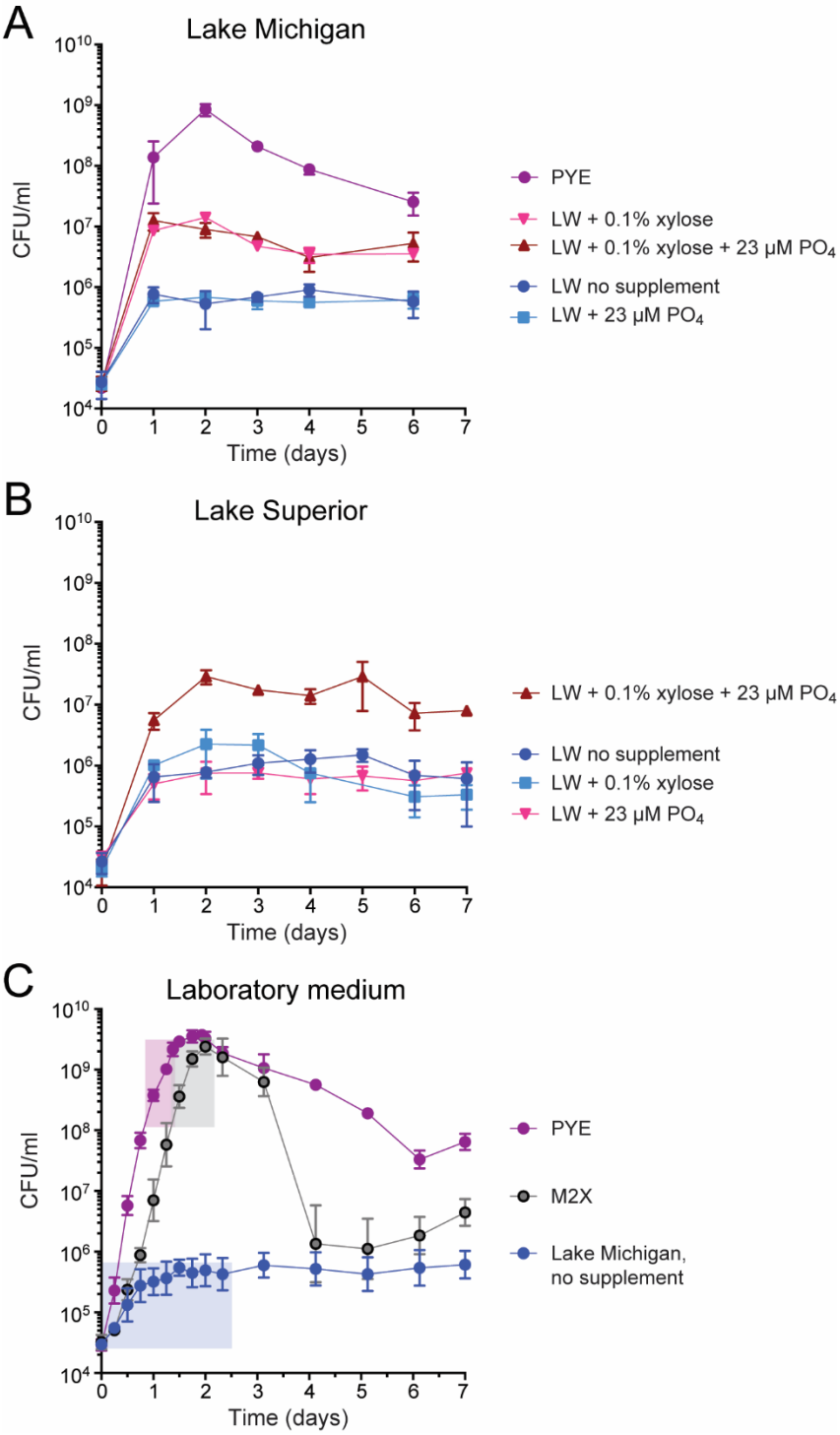


Figure A3.1: Growth of *Caulobacter* in laboratory medium, and supplemented or unsupplemented water from two Great Lakes. A and B) Overnight cultures washed with filtered lake water (LW) were inoculated into 5 mL of filtered water from Lake

Figure A3.1: Growth of *Caulobacter* in laboratory medium, and supplemented or unsupplemented water from two Great Lakes (continued). Michigan (A) or Lake Superior (B). Growth assays in water supplemented with carbon (0.1% w/v xylose) and/or phosphorus (23 μM K_2HPO_4) as indicated; growth was monitored every 24 hours by enumeration of colony forming units (CFUs) by dilution plating. Lake water growth is compared to growth in a laboratory peptone yeast extract (PYE) medium. Data represent mean ± standard deviation of 5 replicates per condition. **C)** Fine scale growth of *Caulobacter* in PYE, M2-xylose defined medium (M2X), and filtered Lake Michigan water. Cells were grown as in A and B and monitored by enumerating CFUs. Data represent mean ± standard deviation of 5 replicates per condition. Boxes represent the approximate region of the growth curve (cell density and incubation time) in which the barcoded Tn-Himar mutant library (BarSeq) pools were cultivated.

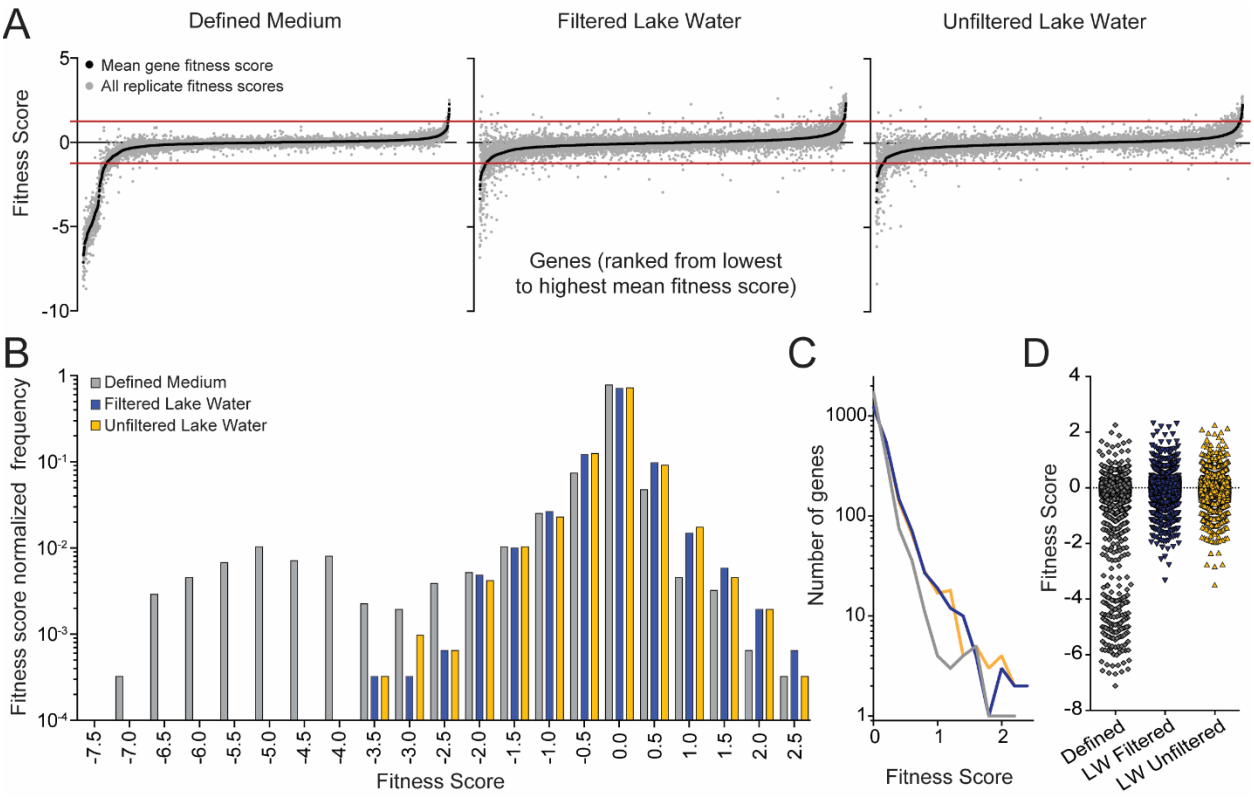
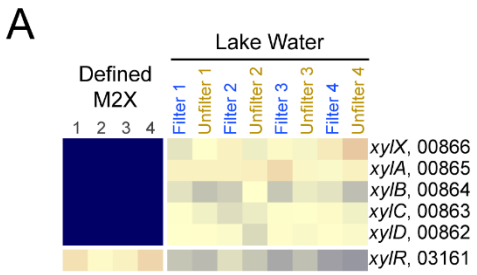


Figure A3.2: *Caulobacter* gene fitness score summary after cultivation in defined medium, filtered, or unfiltered Lake Michigan water. A) Rank ordered mean fitness scores of each scorable *Caulobacter* gene across each of the four replicate experiments for each growth condition is plotted; black = mean fitness score; gray = independent replicate fitness scores. Red lines represent ± 3σ from the mean score of the entire dataset (which is approximately zero; genes with fitness scores less than -2.5 in M2X were excluded from this determination). **B)** Distributions of mean gene fitness scores for each condition: defined M2X medium (gray), filtered Lake Michigan water (blue), and unfiltered Lake Michigan water (yellow). **C)** Distribution of mean gene fitness values between 0 and +2.5 plotted for each condition; defined M2X medium (gray), filtered Lake Michigan water (blue), and unfiltered Lake Michigan water (yellow). **D)** Genes fitness score distribution scores plotted for each of the three cultivation conditions: defined M2X medium (gray), filtered Lake Michigan water (blue), and unfiltered Lake Michigan water (yellow).



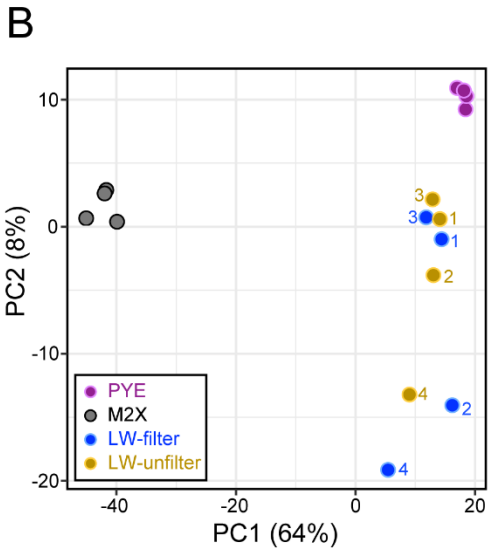


Figure A3.3: Functional validation of the barcoded Tn-Seq approach; PCA profile of Tn strain fitness in complex medium, minimal defined medium and Lake Michigan water. A) Heatmap of gene fitness scores for six xylose utilization genes (*xylX*, *xylA*, *xylB*, *xylC*, *xylD*, *xylR*) from each replicate experiment of cultivation of the Tn-Himar mutant library in M2X defined medium, filtered lake water, and unfiltered lake water. Fitness score color scale bar is shown below the panel. **B)** Principal component (PC) analysis (PCA carried out in ClustVis (Metsalu and Vilo, 2015)) of genomic-scale fitness values for the barcoded *Caulobacter* Tn-Himar mutant library cultivated in complex peptone yeast extract (PYE) medium (reference set), defined M2X medium, filtered Lake Michigan water, and unfiltered Lake Michigan water. The plot shows PCA values for individual samples from each cultivation condition. Percent of variance contributed by the first two PCs is noted on the axes. PCA plot is based on all fitness score values for all genes in the Tn-Himar datasets.

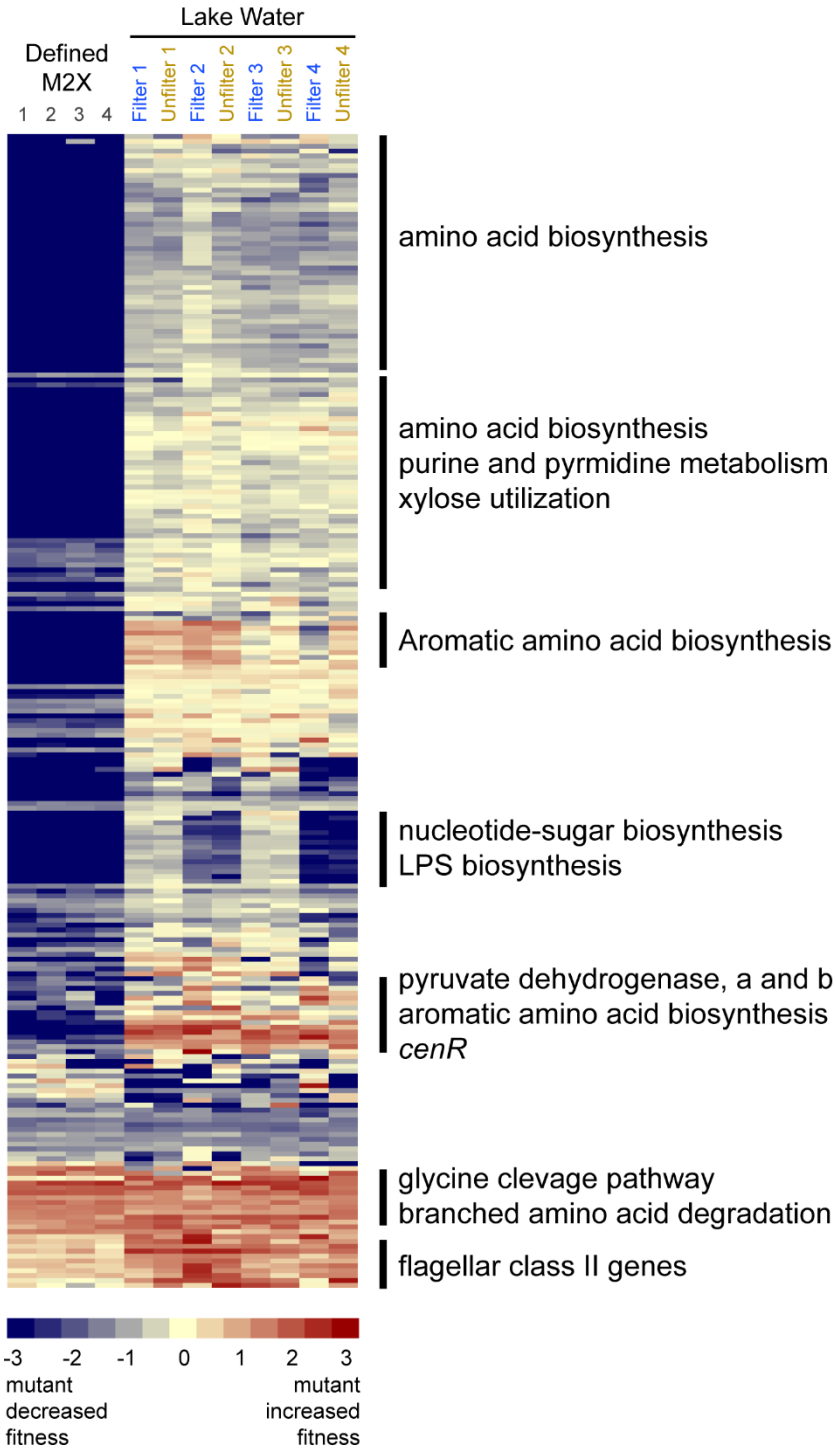


Figure A3.4: Functional summary of mutant strains with diminished or enhanced fitness in minimal defined medium and Lake Michigan water. Heatmap of fitness scores for genes with mean fitness scores higher than 1.2 or lower than -1.2 (approximates ± 3σ fitness score cutoff) in at least one cultivation condition. The sigma

Figure A3.4: Functional summary of mutant strains with diminished or enhanced fitness in minimal defined medium and Lake Michigan water (continued). cutoff was based on the major fitness score distribution centered on zero (see Figure 2B). Genes are hierarchically clustered using Cluster 3.0 (average linkage) and visualized using TreeView, and fitness scores for each replicate experiment are color coded on scale bar shown below. General functions of genes within particular regions of this 256 gene cluster are noted; this entire figure is expanded and split into three fully annotated clusters in Fig. S5, with gene names included. The full cluster of genes with mean fitness scores higher than +1.2 or lower than -1.2 in the lake water conditions (excluding the M2X data) is presented in Fig. S6.

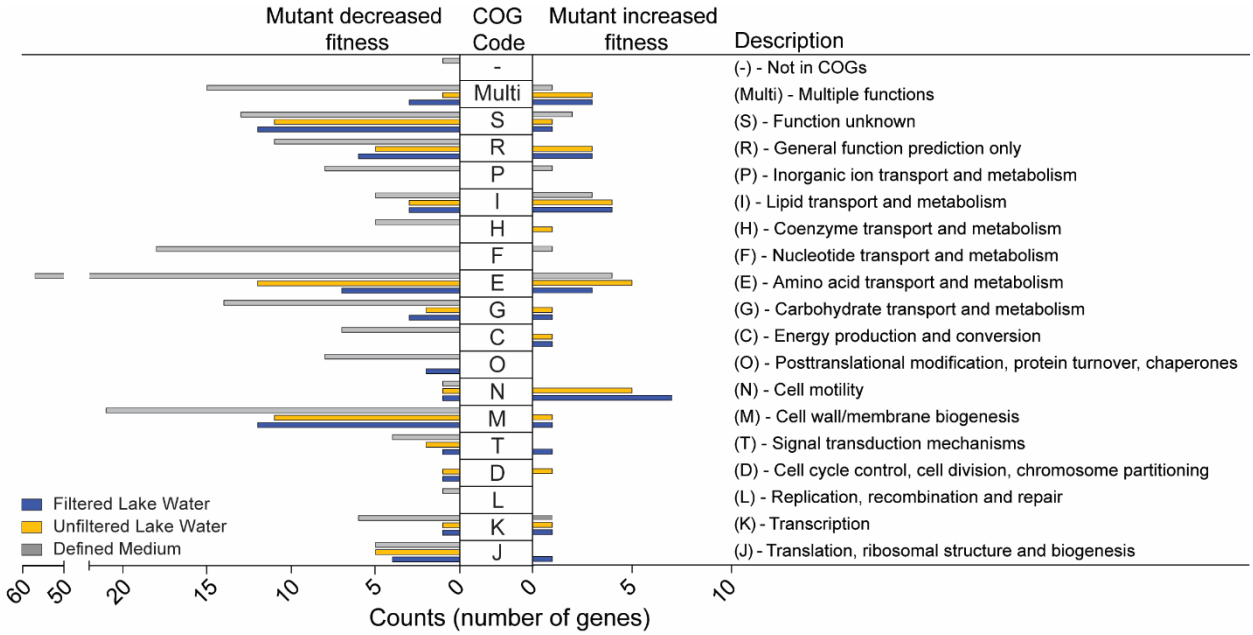


Figure A3.5: COG analysis of Tn-Himar gene fitness data. The analysis includes genes with fitness scores of absolute value greater than 1.2 (which approximates a ± 3σ fitness score cutoff) in each condition. Each gene was assigned a cluster of orthologous group (COG) functional category, obtained through the NCBI COG site (Galperin et al., 2015). The number of genes in each COG category is plotted; genes with negative fitness values (left) and genes with positive fitness values (right).

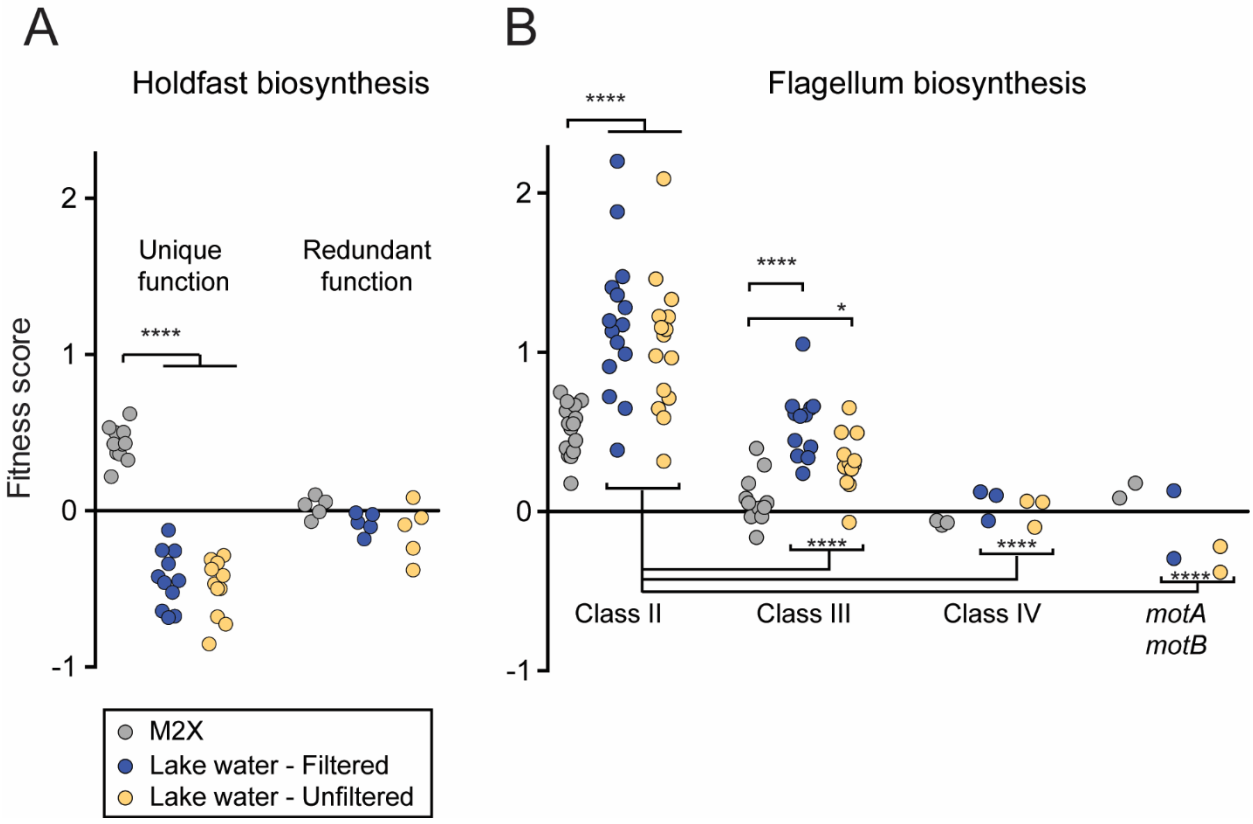


Figure A3.6: Genes with functions in flagellum and holdfast biosynthesis influence fitness in Lake Michigan water. A and B) Mean fitness scores of genes involved in (A) holdfast biosynthesis (17 genes) and (B) flagellum biosynthesis (38 genes) are plotted for each cultivation condition (four independent growth replicates): defined M2X medium (gray); filtered lake water (blue); and unfiltered lake water (yellow). Holdfast genes with unique function (i.e. single gene deletions have a holdfast defect) and redundant function (i.e. single mutants have no holdfast defect) are shown separately. Class II, Class III, and Class IV genes in the flagellar hierarchy and *motA*/*motB* stator complex genes are also shown separately. Clusters of holdfast and flagellum fitness score data, with individually annotated genes, are shown in Fig. S4. One-way ANOVA was applied to assess differences in fitness scores between marked groups; Tukey's post test (**** $p < 0.0001$; * $p < 0.05$).

Appendix B: Tables

Table 2.1: Genetic lesions in mutants identified in the chloramphenicol growth selections

Mutant	Locus	Description	Type of mutation	[a]Nucleotide change	[b]Amino acid change
Strains with lesions in *tipR* and *acrB2* (other lesions in these strains not listed)					
1-2	*tipR*	transcription factor	SNP	A928097T	154* stop
1-17	*tipR*	transcription factor	2 SNP	T928226C, T928844A	S197P, L203Q
	acrB2	efflux pump	SNP	T925807A	Q361L
2-2-17	*tipR*	transcription factor	SNP	T928233A	V199E
	acrB2	efflux pump	SNP	G925825A	A156V
2-2-20	*tipR*	transcription factor	3 SNP	T928209A, T928226A, T928242A	L191Q, S197T, L202Q
	acrB2	efflux pump	SNP	A923992G	V767A
2-1-20	*tipR*	transcription factor	Deletion	ΔC927943	Frameshift after L101
2-1-6	*tipR*	transcription factor	Deletion	ΔC927943	Frameshift after L101
Strains without lesions in *tipR* and *acrB2*					
2-1-16	***CC_0248***	*spdS* histidine kinase	Deletion	ΔT260815-C260833	Frameshift after Q27
	CC_1056	conserved hypothetical cytosolic protein	SNP	T1213999C	W7R
	non-coding	non-coding region	Insertion	+C after 2867910	non-coding
2-1-1	***CC_3555-CC_3565***	multiple genes replaced by IS3 transposon. Includes *CC_3560*.	Deletion/ Replacement	ΔC3830950-G3839676	N/A
	CC_1056	conserved hypothetical cytosolic protein	SNP	T1213999C	W7R
	non-coding	non-coding region	Insertion	+C after 2867910	5' non-coding
2-1-14	***CC_3555-CC_3565***	multiple genes replaced by IS3 transposon. Includes *CC_3560*.	Deletion/ Replacement	ΔC3830950-G3839676	N/A
	CC_1056	conserved hypothetical cytosolic protein	SNP	T1213999C	W7R
	non-coding	non-coding region	Insertion	+C after 2867910	5' non-coding
Modified selection strategy with two copies of *tipR* in the parent strain					
TipR#5-3	***CC_3560***	hybrid histidine kinase	SNP	A3834540G	H82R
	CC_3444	conserved hypothetical protein	SNP	C3713704G	R36P
TipR#7-1	***CC_3560***	hybrid histidine kinase	SNP	A3834540G	H82R
TipR#7-8	***CC_3560***	hybrid histidine kinase	SNP	A3834540G	H82R
TipR#7-2	***CC_3560***	hybrid histidine kinase	SNP	A3834540G	H82R

Table 2.1: Genetic lesions in mutants identified in the chloramphenicol growth selections (continued)

Mutant	Locus	Description	Type of mutation	[a]Nucleotide change	[b]Amino acid change
	CC_3444	conserved hypothetical protein	SNP	C3713704G	R36P
	CC_1729	pyruvate dehydrogenase complex	SNP	T1927487C	No change
TipR#7-18	***CC_3560***	hybrid histidine kinase	SNP	A3834540G	H82R
	CC_1729	pyruvate dehydrogenase complex	SNP	T1927487C	No change
TipR#7-20	***CC_3560***	hybrid histidine kinase	SNP	A3834540G	H82R
	CC_1729	pyruvate dehydrogenase complex	SNP	T1927487C	No change
TipR#8-5	***CC_3560***	hybrid histidine kinase	SNP	A3834540G	H82R
	CC_2978	hypothetical protein	SNP	G3223720A	A27T
	CC_1729	pyruvate dehydrogenase complex	SNP	T1927487C	No change
	CC_1765-CC_1777	multiple genes replaced by IS3 transposon	Deletion/ Replacement	ΔA1974346-C1986866	N/A
TipR#8-6	***CC_3560***	hybrid histidine kinase	SNP	A3834540G	H82R
	CC_3444	conserved hypothetical protein	SNP	C3713704G	R36P
	CC_1729	pyruvate dehydrogenase complex	SNP	T1927487C	No change
TipR#8-19	***CC_3560***	hybrid histidine kinase	SNP	A3834540G	H82R
	CC_3444	conserved hypothetical protein	SNP	C3713704G	R36P

[a]Nucleotide positions refer to the NA1000 genome sequence (genbank accession CP001340).
[b]Based on the NA1000 annotations in genbank accession CP001340.

Table 3.1: Transcripts with more than 2-fold difference between *spdR*(D64E) and Δ*spdR*

[a]Locus	Description	[b]Fold change	Predicted SpdR binding motif
spdR	two-component receiver protein SpdR	637.6	
CCNA_02114	hypothetical protein	357.3	TCTGCGTCCAGCGGTCGCATG
CCNA_R0130	-	145.4	CCTGCGCCGGTTTGTCGCAGT
CCNA_02238	hypothetical protein	72.6	GCCCCGACCTTATGTCGCAGG
CCNA_03249	TonB-dependent receptor	48.9	ATTAGGACACAATGTCGCACC
CCNA_03468	methyl-accepting chemotaxis protein	44.8	GGCGAGACAATCTGTCGCGGG
cheY	chemotaxis protein cheY	30.6	CTCGTGACATTTGGCCGCACC
CCNA_03096	TonB-dependent receptor	28.8	TTTGCGGCGTCTTGCCCCGAT
CCNA_00627	hypothetical protein	28.4	ACCCGGACAAAGCGCCGCA
CCNA_03447	hypothetical protein	25.4	ACCCGGACAAAGCGCCGCATG
CCNA_R0162	-	24.8	CATGCGACCGAACGCCTCGTT
CCNA_02319	hypothetical protein	18.2	CTTGGGGCGATAGGTCGCGGC
cheW	chemotaxis protein cheW	18.1	
CCNA_00969	two component sensor histidine kinase	17.9	AGCGCGATCATAGGTCGCAGG
CCNA_01270	hypothetical protein	17.4	
CCNA_00940	sensory box GGDEF/EAL protein	15.4	TCCGCGACAATCCGCCGCAGC
cheBII	receiver domain-glutamate methylesterase CheBII	15.4	
CCNA_00437	methyl-accepting chemotaxis protein	15.1	GGTGCGGCGAGGCGCCGCACT
CCNA_00629	methyl-accepting chemotaxis protein	14.8	
CCNA_00689	PAS-family hybrid histidine kinase/receiver protein	14.5	GACGCGACCCATTGTCGCATC
cheAII	chemotaxis histidine kinase protein cheAII	14.1	GCTGGGGCGTCAAGCCGCAGC
CCNA_01269	GcrA superfamily protein	13.8	
CCNA_00551	hypothetical protein	12.8	CCTGCGTCAGCGTGTCGCAGG
CCNA_02606	hybrid sensor histidine kinase/receiver domain protein	11.1	CCTGCGGCGTTCTGGCGCACA
CCNA_03267 (rtrA)	Xre-family transcriptional regulator	10.8	ATCGCGACGTAACGCCGCGCC
CCNA_03829	hypothetical protein	10.5	GCTGCGACACCGCGCCGCAAT
CCNA_03154	PAS-family sensor histidine kinase/receiver protein	10.4	GCCGCGACGGTTTGCCTCATG
CCNA_00632	chemotaxis receiver domain protein CheYII	10.1	CCCCGGACAATTTGCCGGTGC
CCNA_00064	methyl-accepting chemotaxis protein	9.6	TCTGCGACATCATGCCACCCA

Table 3.1: Transcripts with more than 2-fold difference between *spdR*(D64E) and Δ*spdR* (continued)

[a]Locus	Description	[b]Fold change	Predicted SpdR binding motif
CCNA_00900	CHASE4-family GGDEF/EAL protein	9.0	CCTGCGACAAGTTGCCTCAGG
CCNA_00848	leucine-responsive regulatory protein	9.0	
CCNA_02945	PAS-family sensor histidine kinase	8.7	CGCGGGACAAATCGCCGCATC
CCNA_03087	H-NOX family heme binding protein	8.3	
CCNA_03088	two-component sensor histidine kinase	7.0	CGCGCGACAGTTTCACGAGGT
CCNA_03901	hypothetical protein	6.9	
CCNA_02715	MASE1-family sensor histidine kinase	6.2	
CCNA_03295	PAS-family sensor histidine kinase	6.2	GCCGAGGCGATATGCCGCAAC
CCNA_00634	chemotaxis protein methyltransferase	5.9	
CCNA_03368	hypothetical protein	5.8	
CCNA_03896	hypothetical protein	5.6	
CCNA_00623	PAS-family sensor histidine kinase	5.3	CCTGCGGCCAATAGTCTCAGC
CCNA_03248	TonB-dependent receptor	5.2	ACCACGACATAATGTCGCACC
CCNA_02116	hypothetical protein	5.2	CGCGCGACCTAATGTCGCAGC
CCNA_03325	hypothetical protein	5.0	
CCNA_03459	methyl-accepting chemotaxis protein	4.7	TGAGCGACAAAAGGCCACACT
CCNA_01007	hypothetical protein	4.6	CGCGCGACAGGTCGTCCCGTT
CCNA_03246	hypothetical protein	4.5	
CCNA_03367	chemotaxis receiver domain protein CheYIII	4.4	TCTGCGACAATTCGCCCCACA
CCNA_R0121	-	4.4	
CCNA_03108	TonB-dependent outer membrane receptor	4.1	
CCNA_02415 (rtrB)	Xre-family transcriptional regulator	4.1	ACCGCGACACTTCAGCGCGAA
CCNA_00986	general stress protein 17O	4.0	CGTGCGGCAGAACGACAACTT
cspD	cold shock protein cspD	3.9	GCTGCGACGAAATGTCGCGGT
CCNA_02286	hypothetical protein	3.8	
CCNA_03660	Usg protein	3.7	CCCGCGACAACGCGCCGCTTG
CCNA_03926	hypothetical protein	3.7	GCCGGGTCAATGGGTCGCGGA
CCNA_03066	GAF/PAS-family sensor histidine kinase	3.7	TCTGCGTCAATTGGCAGCATC
CCNA_00983	hybrid histidine kinase/receiver protein	3.6	
CCNA_00348	methyl-accepting chemotaxis protein	3.4	GATGCGACACACTGTCGCGAA

Table 3.1: Transcripts with more than 2-fold difference between *spdR*(D64E) and Δ*spdR* (continued)

[a]Locus	Description	[b]Fold change	Predicted SpdR binding motif
CCNA_02395	hypothetical protein	3.3	CCCGCGATAATTCGTCGCACA
CCNA_03526	hypothetical protein	3.2	ACCGCGACATTGCGCCCCCTT
dgcA	GGDEF diguanylate cyclase DgcA	3.2	
CCNA_03687	carbonic anhydrase	3.1	AACGTGTCGAATTGCCGCTGT
CCNA_00728	hypothetical protein	3.1	GGTGCGTCAATCCGACGCAGC
fixK	hypoxia transcriptional regulator FixK	3.1	
CCNA_01163	ice nucleation protein	3.0	TCTGCGGCCGTGTGTCTCGTT
CCNA_00626	methyl-accepting chemotaxis protein	3.0	
malA	TonB-dependent maltose outer membrane transporter malA	3.0	
CCNA_03247	methyl-accepting chemotaxis protein	3.0	
CCNA_03951	hypothetical protein	3.0	
CCNA_02163	hypothetical protein	3.0	GGCGCGGCATGTCAGCGCAGG
CCNA_01140	PAS-containing EAL phosphodiesterase	2.9	
CCNA_00243	hypothetical protein	2.8	
CCNA_00624	hypothetical protein with pentapeptide repeats	2.7	
CCNA_03936	hypothetical protein	2.7	
creS	intermediate filament-like cell shape determinant creS	2.7	GCCGCGACGCCCGGTCGCGGC
CCNA_00637	tetratricopeptide repeat family protein	2.7	
CCNA_00532	hypothetical protein, RDD superfamily	2.6	
CCNA_01165	FkbM methyltransferase family protein	2.6	
fliI	flagellum-specific ATP synthase fliL	2.5	CCTGCGACCAATTGCCCGGCA
CCNA_00638	heme:hemopexin-binding protein	2.5	
CCNA_02001	phosphoserine phosphatase	2.5	GGCGAGACGATCAGCCGCACC
CCNA_02597	hypothetical protein	2.5	
CCNA_01100	acylamino-acid-releasing enzyme	2.5	
CCNA_01278	hypothetical protein	2.5	ACTGAGACGTTGGGTCGCACG
CCNA_02922	hypothetical protein	2.5	TCGCGGACAACTTGTCGCAAG
CCNA_02523	hypothetical protein	2.4	
CCNA_03250	EAL-domain diguanylate phosphodiesterase	2.4	GGCGCGGCAGGCTGTCGCAGG
CCNA_02939	hypothetical protein	2.3	CCTGCGACCAAGCGCCTCCTT

Table 3.1: Transcripts with more than 2-fold difference between *spdR*(D64E) and Δ*spdR* (continued)

[a]Locus	Description	[b]Fold change	Predicted SpdR binding motif
CCNA_00639	heme:hemopexin-binding protein	2.3	
CCNA_00769	TetR family transcriptional regulator	2.3	
CCNA_00768	hypothetical protein	2.3	
CCNA_00770	TetR family transcriptional regulator	2.3	
CCNA_02402	methyl-accepting chemotaxis protein	2.3	GGTGCGGCCACGGGGCGCATC
CCNA_01973	hypothetical protein	2.3	
CCNA_02951	WbqC-like family protein	2.3	
sapA	secreted protease precursor sapA	2.3	
CCNA_01164	O-linked N-acetylglucosamine transferase	2.3	
CCNA_01684	phenylalanine-4-hydroxylase	2.2	
cheYIII	cheYIII	2.2	
CCNA_00443	chemotaxis protein cheW	2.2	
CCNA_01465	methyl-accepting chemotaxis protein	2.2	ATTGAGGCCCAGCGGCGCACG
CCNA_03790	hypothetical protein	2.2	GGTGCGCTCTCAGGTCGCGGA
CCNA_03605	hypothetical protein	2.1	GCTGCGCGCTTTCGCCGCCGT
fliM	flagellar motor switch protein fliM	2.1	
CCNA_00444	chemotaxis protein methyltransferase	2.1	
CCNA_02982	porin	2.1	CGTGCGGCAGCGTGCCGCCCT
CCNA_R0190	-	2.1	
CCNA_02137	endoglucanase H	2.1	CCTGCGTTCCAGCGTCGCCGC
CCNA_02743	TonB dependent receptor	2.1	
hipB	HTH transcriptional regulator	2.1	
CCNA_01649	transporter, drug/metabolite exporter family	2.1	GGTGCGACAGGTTGTCGGCGT
flmF	flagellin modification protein FlmF	2.1	GGTGCGCCGCCTCGTCGACGA
cheU	cheU protein	2.0	CCGTCGACAACGCGCCGCAGG
fliL	flagellum-specific ATP synthase fliL	2.0	
CCNA_03082	hypothetical protein	2.0	
pflI	polar flagellum positioning protein pflI	2.0	CCGCTGACACGCGGCCGCATG
CCNA_03353	metallophosphatase domain protein	2.0	AAGCGGAGAATTTGCCGCACG
flmE	S-adenosylmethionine-dependent methyltransferase related protein	2.0	CGTGCGACAACCTGACCCACC
CCNA_00426	very-short-patch-repair endonuclease	-2.0	

Table 3.1: Transcripts with more than 2-fold difference between *spdR*(D64E) and Δ*spdR* (continued)

[a]Locus	Description	[b]Fold change	Predicted SpdR binding motif
CCNA_01237	hypothetical protein	-2.0	
CCNA_03568	LemA protein	-2.0	
CCNA_00583	lysozyme family protein	-2.0	
CCNA_01260	hypothetical protein	-2.0	
CCNA_02931	flgE-related flagellar hook protein	-2.1	TCTGCGGCGGCGTATCCAAGG
CCNA_00589	hypothetical protein	-2.1	
CCNA_03929	hypothetical protein	-2.1	CCTGCGCCCCCTCGCGGGAGC
CCNA_R0199	-	-2.1	
CCNA_01841	isocitrate lyase	-2.1	
CCNA_03918	hypothetical protein	-2.1	GGTCGGGTGAGGGGTCGCGCT
CCNA_R0064	Met tRNA	-2.1	
CCNA_03935	hypothetical protein	-2.1	
CCNA_03592	outer membrane lipoprotein Blc	-2.1	CCTGCGACGTCGTCCCGCCCC
CCNA_00585	hypothetical protein	-2.2	
CCNA_00355	hypothetical protein with pentapeptide repeats	-2.2	CGTGCGCGCTGTTCTCGCGTC
CCNA_R0082	Met tRNA	-2.2	
CCNA_02242	PHB granule-associated protein, phasin2	-2.2	
CCNA_R0065	5S ribosomal RNA	-2.2	
CCNA_R0083	5S ribosomal RNA	-2.2	
CCNA_03678	putative outer membrane protein	-2.2	
CCNA_00784	peptidoglycan-associated outer membrane lipoprotein	-2.3	CATGCGGTTGCTGGTCGGCGT
CCNA_R0052	Leu tRNA	-2.3	
CCNA_03982	hypothetical protein	-2.3	GCTGCAGCCTCTCGTCGGCGG
CCNA_02106	TonB-dependent outer membrane receptor	-2.3	
CCNA_01078	methyltransferase	-2.3	
CCNA_R0151	-	-2.3	
CCNA_03826	hypothetical protein	-2.4	CGCGCGCCAAAGCGCCAGAGC
CCNA_R0040	-	-2.5	
CCNA_R0108	-	-2.5	
CCNA_02309	EF hand domain protein	-2.5	
CCNA_02531	proline-rich hypothetical protein	-2.5	GCGGCGAGCAGATGGCGCACT
CCNA_01601	hypothetical protein	-2.6	GCGGGGGCGAGAAGGCGCATG

Table 3.1: Transcripts with more than 2-fold difference between *spdR*(D64E) and Δ*spdR* (continued)

[a]Locus	Description	[b]Fold change	Predicted SpdR binding motif
CCNA_03825	hypothetical protein	-2.6	CGCGCGAGGGCGCGCCGCGCA
CCNA_01072	hypothetical protein	-2.6	
CCNA_01236	general stress protein GsiB	-2.9	
CCNA_01247	CESA-like glycosyltransferase	-3.0	
CCNA_R0089	Ala tRNA	-3.1	
CCNA_R0092	-	-3.3	TCCCCGAGATGAGCCCGCGCG
CCNA_02532	hypothetical protein	-3.4	
CCNA_00859	acetyltransferase, GNAT family	-3.5	GCCGTGACATTGGCGCGCGGC
CCNA_02342	hypothetical protein with pentapeptide repeats	-3.7	TTTGCGACAATCCGACGCAGT
CCNA_R0088	-	-3.9	
CCNA_00882	hypothetical protein	-4.0	
CCNA_03970	hypothetical protein	-4.9	TCCGCGTCGATAGCGCGCGGT
hfiA	holdfast inhibitor protein HfiA	-5.1	
fljO	flagellin FljO	-5.4	GGAGCGTCATATTGTCGCGGT
CCNA_03941	hypothetical protein	-10.5	GGCGCGGCGGGTTTTCGCGCG

[a]Based on the NA1000 annotation CP001340.
[b]2-fold change in transcripts cut-off.

Table 3.2: Transcripts with more than 1.5-fold difference between an overexpression strain and empty vector control

Locus	Description	[a]*rtrA*$^{++}$ fold change	[a]*rtrB*$^{++}$ fold change	[a]*rtrA*$^{++}$ *rtrB*$^{++}$ fold change
CCNA_03967	hypothetical protein	-1.16092	-445.305	-405.668
CCNA_03969	hypothetical protein	-18.183	-303.077	-276.559
CCNA_03385	flagellin N-methylase	-175.495	-187.21	-170.687
CCNA_01247	CESA-like glycosyltransferase	-37.81	-3.82858	-41.7303
CCNA_03826	pyocin large subunit family protein	-54.1735	-2.52614	-29.9874
cdzC	killing protein CdzC	-19.002	-4.30194	-26.6956
CCNA_00594	hypothetical protein	-16.1998	-4.3436	-22.7064
CCNA_00593	hypothetical protein	-16.5629	-4.13872	-21.5602
cdzB	type I secretion protein CdzB, adaptor	-18.0044	-4.00752	-20.2388
CCNA_00426	very-short-patch-repair endonuclease	-28.6733	-2.98228	-20.214
CCNA_00591	hypothetical protein	-19.1802	-3.64131	-19.6617
CCNA_00592	ferritin superfamily protein	-19.131	-3.16274	-18.7987
CCNA_00589	hypothetical protein	-19.1599	-4.17794	-18.1569
CCNA_00590	FixC-family flavoprotein dehydrogenase	-14.6532	-3.34114	-15.4275
cdzD	killing protein CdzD	-14.5189	-4.22024	-13.9709
cdzA	type I secretion protein CdzA, transmembrane ATPase/peptidase	-12.0433	-3.39684	-13.6124
CCNA_R0120		-10.5563	-1.94986	-12.8372
dprA	DNA recombination-mediator protein A, SMF family	-8.53176	-2.14227	-11.9216
CCNA_R0053		-10.2467	-9.44552	-11.2547
CCNA_03825	hypothetical protein	-16.8295	-2.04006	-11.2116
CCNA_01260	hypothetical protein	-14.5269	-1.67799	-9.41929
CCNA_01536	hypothetical protein	-13.2171	-2.17344	-9.39295
CCNA_00588	SapA domain protein	-11.7365	-2.35406	-9.03862
CCNA_00744	GIY-YIG endonuclease domain protein	-9.71477	-2.62556	-8.74108
CCNA_01405	GIY-YIG domain protein	-15.4365	-1.38669	-8.54829
CCNA_00967	TetR-family transcriptional regulator	-6.82062	-2.48291	-8.44422
CCNA_02762	DNA repair protein RadC	-7.5963	-1.68764	-8.31119
CCNA_00582	hypothetical protein	-8.01597	-2.50894	-7.91664
cdzI	immunity protein CdzI	-5.85362	-3.41123	-7.36412
CCNA_R0111		-12.8697	-4.38559	-6.95635
CCNA_02831	hypothetical protein	-7.99405	-1.95281	-6.82327
CCNA_00579	endonuclease	-5.24647	-2.67814	-6.59066
CCNA_03886	trypsin-like peptidase domain family protein	-6.5734	-1.82059	-6.46238

Table 3.2: Transcripts with more than 1.5-fold difference between an overexpression strain and empty vector control (continued)

Locus	Description	[a]*rtrA*$^{++}$ fold change	[a]*rtrB*$^{++}$ fold change	[a]*rtrA*$^{++}$ *rtrB*$^{++}$ fold change
CCNA_R0117		-7.51586	-2.58173	-6.39016
CCNA_03935	hypothetical protein	-31.4082	-3.89289	-6.3782
CCNA_03887	hypothetical protein	-9.11432	-1.24676	-6.29809
CCNA_00600	hypothetical protein	-11.383	-1.30573	-6.11514
CCNA_00601	MoxR-like ATPase	-8.53201	-1.47124	-5.50489
CCNA_02528	PlsY-like acyl-phosphate glycerol 3-phosphate acyltransferase	-4.82391	-1.80561	-5.46074
CCNA_03911	hypothetical protein	-7.27285	-1.35061	-5.38495
CCNA_00603	integral membrane protein	-7.57234	-1.35264	-5.34485
CCNA_00718	abortive infection protein	-6.28983	-1.73376	-5.32013
CCNA_02312	TraB superfamily protein	-5.44134	-2.08393	-5.13282
CCNA_00020	hypothetical protein	-5.41036	-1.53431	-4.82358
chlI	magnesium chelatase subunit ChlI-like ATPase	-4.38449	-2.42955	-4.71512
CCNA_00580	AAA ATPase domain protein	-5.26887	-2.09358	-4.6519
CCNA_00716	hypothetical protein	-4.0276	-1.528	-4.60114
CCNA_00583	lysozyme family protein	-3.94349	-1.94254	-4.56711
CCNA_03983	HicA-related toxin-antitoxin protein	-3.53513	-1.49878	-4.47185
CCNA_00873	amidophosphoribosyltransferase family protein	-3.94564	-1.9165	-4.46769
CCNA_02779	LysM family peptidoglycan binding protein	-7.53303	-1.44467	-4.4077
CCNA_00602	von Willebrand factor type A domain protein	-4.96487	-1.5622	-4.40753
CCNA_02661	M16 family peptidase	-4.78864	-1.46641	-4.39769
CCNA_01121	MA superfamily peptidase	-5.81333	-1.49263	-4.37226
CCNA_02944	hypothetical protein	-3.91295	-2.02106	-4.36435
CCNA_R0137		-4.24795	-1.38312	-4.25907
CCNA_01766	CsgG-related curli protein	-4.45468	-1.58446	-4.01052
CCNA_03839	acylamino-acid-releasing enzyme	-6.02783	-1.26955	-3.82057
CCNA_03981	hypothetical protein	-3.82386	-1.5578	-3.74013
CCNA_03929	hypothetical protein	-6.1087	-1.22352	-3.73555
CCNA_03603	beta-lactamase family protein	-6.05913	-1.22518	-3.65493
CCNA_01122	MA superfamily peptidase	-5.46477	-1.30228	-3.64913
CCNA_03928	hypothetical protein	-2.16478	-1.02438	-3.55701
hfiA	holdfast inhibitor protein HfiA	-3.32219	-1.88386	-3.31298
CCNA_02905	very short patch repair (Vsr) endonuclease	-4.03074	-1.35086	-3.31256
CCNA_02760	hypothetical protein	-4.34423	-1.21652	-3.23944
CCNA_03396	PDZ domain trypsin-like serine protease	-5.53515	-1.1195	-3.19315

Table 3.2: Transcripts with more than 1.5-fold difference between an overexpression strain and empty vector control (continued)

Locus	Description	[a]*rtrA*$^{++}$ fold change	[a]*rtrB*$^{++}$ fold change	[a]*rtrA*$^{++}$ *rtrB*$^{++}$ fold change
CCNA_03966	hypothetical protein	-5.91234	-1.17072	-3.16657
CCNA_03976	hypothetical protein	-3.8616	-1.21744	-3.11552
CCNA_00604	RDD family membrane protein	-3.45701	-1.31012	-3.06032
CCNA_03927	hypothetical protein	-4.8542	-1.40185	-3.05745
CCNA_00019	molybdopterin-guanine dinucleotide biosynthesis protein A	-4.4207	-1.33925	-2.96622
CCNA_03255	PemK-family protein	-2.42475	-1.36001	-2.89453
CCNA_02761	hypothetical protein	-3.90078	-1.25396	-2.87092
CCNA_00587	alpha/beta hydrolase family protein	-2.64831	-1.73479	-2.79866
CCNA_02672	hypothetical protein	-3.90726	-1.13453	-2.70534
CCNA_R0040		-2.86905	-1.29707	-2.64609
CCNA_00745	hypothetical protein	-1.6491	-1.28749	-2.63544
CCNA_01981	DNA translocation competence protein ComA	-3.19424	-1.35091	-2.62188
CCNA_00585	hypothetical protein	-2.2287	-1.57575	-2.57936
CCNA_02931	flgE-related flagellar hook protein	-2.25565	-2.5108	-2.56292
CCNA_01980	PEP phosphonomutase related enzyme	-2.82166	-1.32392	-2.5534
CCNA_00026	PAS-family sensor histidine kinase	-4.92714	1.019844	-2.55283
CCNA_00665	GAF-family sensor histidine kinase	-2.89147	-1.03952	-2.55207
CCNA_02987	hypothetical protein	-3.46987	-1.24083	-2.49335
CCNA_03265	CBS pair-family sensor histidine kinase/receiver domain protein	-1.98964	-1.4073	-2.45326
CCNA_01083	HATPase superfamily protein	-3.98422	1.165997	-2.45216
CCNA_02311	hypothetical protein	-2.41661	-1.91942	-2.42974
CCNA_03930	hypothetical protein	-7.11893	2.841146	-2.32698
CCNA_02061	hypothetical protein	-2.4687	-2.14838	-2.31567
CCNA_03485	membrane protease family, stomatin/prohibitin-like protein	-2.22554	-1.10519	-2.29712
CCNA_00605	conserved hypothetical membrane protein	-2.15247	-1.34802	-2.29357
CCNA_03355	acylamino-acid-releasing enzyme	-3.28971	-1.01075	-2.25581
CCNA_00859	acetyltransferase, GNAT family	-2.00605	-1.62835	-2.22114
CCNA_02763	bacteriophage P4 integrase	-2.26904	-1.22692	-2.21674
CCNA_03484	NfeD-related membrane protease regulator	-3.28999	-1.17721	-2.1476
CCNA_02800	HTH/XRE-family transcriptional regulator	-2.25405	-1.08614	-2.12558

Table 3.2: Transcripts with more than 1.5-fold difference between an overexpression strain and empty vector control (continued)

Locus	Description	[a]*rtrA*$^{++}$ fold change	[a]*rtrB*$^{++}$ fold change	[a]*rtrA*$^{++}$ *rtrB*$^{++}$ fold change
CCNA_03348	hypothetical protein	-2.31776	-1.07519	-2.12268
CCNA_02409	hybrid sensor histidine kinase/receiver protein	-2.04778	-1.75829	-2.10871
CCNA_02976	MCP-signal associated domain protein	-2.7083	-1.93927	-2.07299
relB-4	anti-toxin protein relB-4	-1.93829	-1.35977	-2.055
CCNA_01119	hypothetical protein	-2.26049	-1.67957	-2.03784
CCNA_02728	hypothetical protein	-2.69737	-1.08826	-2.02307
CCNA_03915	hypothetical protein	-1.68535	-2.00989	-2.01348
CCNA_00910	erythromycin esterase family protein	-1.92113	-1.00779	-1.97346
CCNA_03407	sodium export permease protein	-1.61638	-1.31812	-1.9147
CCNA_R0197		-1.43561	-1.50756	-1.89925
CCNA_00629	methyl-accepting chemotaxis protein	-1.65957	-1.46065	-1.87189
CCNA_00437	methyl-accepting chemotaxis protein	-2.97818	-1.00506	-1.82951
CCNA_01084	hypothetical protein	-2.42489	-1.03271	-1.82135
CCNA_03408	ABC transporter ATP-binding protein	-1.65386	-1.26358	-1.7971
CCNA_00912	beta-lactamase family protein	-1.52627	1.006758	-1.79436
CCNA_R0005		-1.96797	-1.76593	-1.7887
fljN	flagellin FljN	-1.96388	-2.08029	-1.78493
CCNA_02105	hypothetical protein	-2.08021	-1.5061	-1.7815
CCNA_02589	BolA-related transcriptional regulator	-1.74959	-1.15819	-1.78115
CCNA_02590	aldo-keto reductase family protein	-1.43701	-1.18469	-1.7655
CCNA_01111	CCNA_01111-DnaA-related protein	-2.97736	-1.48694	-1.75791
CCNA_00438	hypothetical protein	-3.23453	-1.31746	-1.75686
CCNA_00877	hypothetical protein	-1.95841	-1.04153	-1.73587
CCNA_03279	short chain dehydrogenase	-2.03314	-1.06758	-1.72189
CCNA_03426	MarR family repressor protein	-2.67239	-1.20566	-1.72176
CCNA_03325	hypothetical protein	-2.24285	-1.25267	-1.70578
CCNA_03367	chemotaxis receiver domain protein CheYIII	-1.69577	-1.03698	-1.69266
CCNA_00913	hypothetical protein	-2.16697	1.001786	-1.68947
CCNA_02106	TonB-dependent outer membrane receptor	-1.53242	-1.33174	-1.67025
CCNA_R0083		-2.06243	-1.46817	-1.65872
CCNA_01841	isocitrate lyase	-1.72065	-1.57927	-1.65465
hfsI	secreted polysaccharide polymerase hfsI	-2.52619	1.040102	-1.65422
CCNA_R0065		-2.05248	-1.46363	-1.65267
CCNA_00656	type I restriction-modification system methylation subunit	-1.6725	1.164265	-1.65175

Table 3.2: Transcripts with more than 1.5-fold difference between an overexpression strain and empty vector control (continued)

Locus	Description	[a]*rtrA*$^{++}$ fold change	[a]*rtrB*$^{++}$ fold change	[a]*rtrA*$^{++}$ *rtrB*$^{++}$ fold change
CCNA_01843	malate synthase	-1.90266	-1.4789	-1.63957
CCNA_00901	alpha/beta hydrolase family protein	-1.98701	1.082097	-1.63069
CCNA_01842	hypothetical protein	-1.62129	-1.45956	-1.62082
CCNA_00732	GumN superfamily protein	-1.64352	1.029045	-1.61652
CCNA_02692	YceI-like domain protein	-1.57223	1.020112	-1.61532
CCNA_03020	phosphatidylinositol-specific phospholipase C1-like protein	-1.76785	-1.25074	-1.61372
CCNA_02597	hypothetical protein	-1.48426	1.376344	-1.60816
CCNA_00734	GumN superfamily protein	-1.54513	-1.12632	-1.60146
CCNA_00911	AAA domain ATPase	-1.9323	1.15578	-1.59518
fljM	flagellin FljM	-1.56546	-1.51269	-1.58984
CCNA_00658	5-methylcytosine-specific restriction enzyme MRR	-1.7054	1.022728	-1.58771
CCNA_02172	ABC-type transporter, permease component	-1.62711	1.056174	-1.58324
CCNA_02986	phage-related conserved hypothetical protein	-1.96636	-1.01276	-1.56282
CCNA_01186	hypothetical protein	-2.49306	-1.55484	-1.5626
CCNA_01259	class II aldolase and Adducin protein	-1.46178	-1.12242	-1.56212
CCNA_01764	DNA polymerase III, chi subunit	-1.71333	-1.30257	-1.55727
CCNA_00224	TonB-dependent receptor protein	-2.52603	1.135621	-1.55246
CCNA_01521	putative exported protein	-1.74562	-1.07673	-1.52786
CCNA_02815	ice nucleation protein	-1.52543	-1.06615	-1.52638
CCNA_01100	acylamino-acid-releasing enzyme	-2.41807	1.101409	-1.52386
CCNA_02588	aminopeptidase	-1.37696	-1.06958	-1.51841
CCNA_03482	ABC transporter permease protein	-1.41348	-1.11619	-1.51148
CCNA_01117	hypothetical protein	-1.52574	-1.26952	-1.50307
hdaA		-1.49338	-1.36665	-1.50297
CCNA_02816	hypothetical protein	-1.66156	-1.13339	-1.49833
hfsK	GNAT-family holdfast biogenesis protein HfsK	-1.68931	1.122044	-1.49813
CCNA_02962	HicB-family RNAse H fold protein	-1.90047	-1.26645	-1.46878
sppA	signal peptide peptidase A	-1.52636	-1.13646	-1.46624
fliX	flagellar assembly regulator FliX	-1.61805	-1.24154	-1.46581
dgrB	diguanylate receptor protein DgrB	-1.89044	-1.62382	-1.46287
CCNA_00659	type I restriction-modification system restriction subunit	-1.74557	1.072472	-1.43795
CCNA_02347	phosphomannomutase/phosphoglucomutase	-1.7725	-1.16297	-1.42833

Table 3.2: Transcripts with more than 1.5-fold difference between an overexpression strain and empty vector control (continued)

Locus	Description	[a]*rtrA*$^{++}$ fold change	[a]*rtrB*$^{++}$ fold change	[a]*rtrA*$^{++}$ *rtrB*$^{++}$ fold change
CCNA_02693	M10 peptidase-related protein	-1.76749	1.486081	-1.42496
CCNA_03446	feruloyl-CoA synthetase	-1.61726	-1.11305	-1.38591
CCNA_02433	hypothetical protein	-1.69437	-1.53394	-1.38531
CCNA_01551	hypothetical protein	-1.54825	-1.03225	-1.38091
CCNA_02845	two-component response regulator	-1.54808	-1.07223	-1.36939
CCNA_03896	hypothetical protein	-1.68381	-1.33309	-1.36818
CCNA_02534	hypothetical protein	-1.65226	-1.23007	-1.354
CCNA_02945	PAS-family sensor histidine kinase	-1.67313	-1.21848	-1.35335
CCNA_00143	ribosomal RNA small subunit methyltransferase I	-1.78076	-1.12358	-1.34816
CCNA_02036	DNA-binding protein HU	-1.80368	-1.23304	-1.3437
CCNA_01480	hypothetical protein	-1.91283	1.053747	-1.33814
CCNA_01742	hypothetical protein	-1.75622	1.02525	-1.33796
CCNA_01481	cyclase homology domain protein	-1.85481	1.019147	-1.31997
CCNA_03417	hydrolase	-2.0116	-1.34006	-1.31839
CCNA_02935	methyl-accepting chemotaxis protein	-1.51829	1.090402	-1.31683
cleC	single domain receiver protein CleC	-1.50804	1.111178	-1.31417
CCNA_01749	acyl carrier protein	-1.43468	-1.51051	-1.31061
CCNA_02416	DNA-binding protein HU	-1.63728	-1.2791	-1.30547
CCNA_03676	GNAT acetyltransferase protein	-1.79388	1.336841	-1.29542
CCNA_00248	sensor histidine protein kinase	-1.717	1.079786	-1.2664
CCNA_00350	hypothetical protein	-1.52891	-1.08264	-1.26582
CCNA_00777	PAS family GGDEF/EAL protein	-1.52844	1.155311	-1.22816
CCNA_03368	hypothetical protein	-1.58997	-1.24408	-1.22318
flaF	flagellar biosynthesis regulatory protein flaF	-1.65874	-1.40684	-1.22131
CCNA_02160	hypothetical protein	-1.88624	-1.052	-1.19779
CCNA_00743	NnrU family membrane protein	-1.64218	-1.0366	-1.15607
CCNA_03087	H-NOX family heme binding protein	-1.98191	-1.0428	-1.12278
CCNA_00784	peptidoglycan-associated outer membrane lipoprotein	1.542551	-1.38057	-1.09404
CCNA_03078	Arc/MetJ-family transcriptional regulator	-1.75023	-1.10522	-1.08677
CCNA_03907	hypothetical protein	-1.58314	-1.17244	-1.07033
CCNA_R0061		1.628959	-1.01136	-1.03829
CCNA_01189	TonB-dependent receptor	1.559239	-1.10677	-1.02629
CCNA_01145	monoamine oxidase	1.535174	-1.02784	-1.02573

Table 3.2: Transcripts with more than 1.5-fold difference between an overexpression strain and empty vector control (continued)

Locus	Description	[a]*rtrA*$^{++}$ fold change	[a]*rtrB*$^{++}$ fold change	[a]*rtrA*$^{++}$ *rtrB*$^{++}$ fold change
CCNA_02805	hypothetical protein	-1.04427	1.504918	-1.00437
rplV	LSU ribosomal protein L22P	1.52475	-1.03841	1.005255
phd	SpoVT-AbrB family transcription factor, phd antitoxin	-1.54291	-1.08347	1.017382
rplF	LSU ribosomal protein L6P	1.54454	-1.00411	1.117202
CCNA_03960	hypothetical protein	1.672156	1.113885	1.143414
CCNA_00443	chemotaxis protein cheW	1.72038	1.182622	1.173134
CCNA_00006	enoyl-CoA hydratase	2.051563	-1.05454	1.17933
rpsN	SSU ribosomal protein S14P	1.606182	1.015248	1.180786
CCNA_00820	3-ketoacyl-CoA thiolase	-1.72825	1.161046	1.206018
CCNA_03778	integral membrane protein	1.513968	-1.0895	1.207079
nstA	redox cell cycle regulatory protein NstA	1.562739	-1.08817	1.215886
rpsS	SSU ribosomal protein S19P	2.278916	1.072265	1.223334
CCNA_02517	hypothetical protein	2.175107	1.28588	1.248962
CCNA_03214	hypothetical protein	1.807092	1.875869	1.253632
CCNA_01027	acetyltransferase	1.249165	1.745041	1.265576
CCNA_00490	conserved hypothetical membrane protein	1.541416	1.37445	1.290226
CCNA_02201	nucleoside 2-deoxyribosyltransferase	1.627929	1.223568	1.290592
CCNA_01453	electron transport complex I subunit, NADH-ubiquinone oxidoreductase	1.661732	1.026997	1.294573
CCNA_R0134		2.661268	1.572757	1.300357
CCNA_00126	glutathione S-transferase	1.605733	1.300445	1.306091
CCNA_00514	pirin	1.517306	1.301031	1.310394
CCNA_01094	hypothetical protein	1.806806	-1.09125	1.331333
CCNA_01223	acyl-CoA synthetase	1.740864	1.084783	1.331834
CCNA_02893	beta-xylosidase/alpha-L-arabinofuranosidase	1.5819	1.04922	1.336333
CCNA_02743	TonB dependent receptor	1.942327	-1.00509	1.341668
CCNA_00687	hypothetical protein	1.618203	1.274294	1.343046
CCNA_00105	MarC family integral membrane protein	1.780088	1.219696	1.343455
CCNA_03248	TonB-dependent receptor	1.556889	-1.03641	1.352202
CCNA_01945	anhydro-N-acetylmuramic acid kinase	1.474363	1.636028	1.373699
relE1		1.127084	1.804422	1.378098
CCNA_00415	hypothetical protein	1.708882	1.231268	1.390638
CCNA_01417	5-aminolevulinic acid synthase	1.560444	1.327202	1.400116
CCNA_03556	multidrug efflux system protein	1.11857	1.614779	1.421512

Table 3.2: Transcripts with more than 1.5-fold difference between an overexpression strain and empty vector control (continued)

Locus	Description	[a]*rtrA*$^{++}$ fold change	[a]*rtrB*$^{++}$ fold change	[a]*rtrA*$^{++}$ *rtrB*$^{++}$ fold change
CCNA_02723	PAS domain protein	1.5566	1.74538	1.435726
CCNA_00374	putative ATP synthase protein I	1.605697	1.256226	1.445514
CCNA_00793	phosphatidylglycerol glycosyltransferase	1.772162	1.20901	1.471418
CCNA_01727	methyl-accepting chemotaxis protein	2.19394	-1.05868	1.48623
CCNA_02895	TonB-dependent receptor	1.764582	1.120446	1.494983
stpX	stalk-specific protein X	1.612182	1.31723	1.504537
CCNA_03307	hypothetical protein	1.568295	1.428304	1.504901
CCNA_03945	hypothetical protein	2.208101	1.121742	1.515915
CCNA_02852	hypothetical protein	1.415851	1.282358	1.522684
ldpA		1.450845	1.387809	1.561841
CCNA_02465	UDP-glucose 6-dehydrogenase	1.815398	1.352016	1.59713
CCNA_02462	hypothetical protein	1.592066	2.364457	1.605633
CCNA_03600	integral membrane protein	1.754478	1.027318	1.606104
CCNA_02297	hypoxanthine-guanine phosphoribosyltransferase	1.536941	1.32635	1.613369
CCNA_01705	hypothetical protein	1.185971	1.738595	1.629531
CCNA_03995	hypothetical protein	2.148261	1.345171	1.630472
CCNA_02940	methyl-accepting chemotaxis protein	1.955339	1.190272	1.662328
CCNA_01756	hypothetical protein	1.140285	1.889422	1.675378
CCNA_00067	YciI-like protein	2.522433	1.804456	1.689927
CCNA_01437	transglycosylase associated protein	1.862212	1.832753	1.707536
CCNA_03327	hypothetical protein	1.97643	1.646559	1.765228
CCNA_01607	SNARE associated protein family	1.565941	1.376694	1.782787
CCNA_00821	hypothetical protein	-1.86781	1.801942	1.856867
CCNA_03724	hypothetical protein	1.651613	1.64869	1.857471
CCNA_02755	HipB-family transcriptional regulator	2.955531	-1.0639	1.883106
CCNA_02286	hypothetical protein	3.247769	3.681264	1.899063
CCNA_02939	hypothetical protein	1.969125	1.409099	1.958075
CCNA_03082	hypothetical protein	1.540508	1.744387	1.960756
CCNA_03889	hypothetical protein	-1.18641	1.381717	2.031284
CCNA_03917	hypothetical protein	2.086909	2.020467	2.079783
CCNA_02715	MASE1-family sensor histidine kinase	2.620766	1.187018	2.084265
CCNA_03487	zonular occludens toxin, zot-like protein	1.881537	1.469193	2.099074
CCNA_03169	SH3 domain protein	2.232052	1.842789	2.104819
CCNA_00879	PAS-family sensor histidine kinase	1.8012	1.693737	2.244145

Table 3.2: Transcripts with more than 1.5-fold difference between an overexpression strain and empty vector control (continued)

Locus	Description	[a]*rtrA*$^{++}$ fold change	[a]*rtrB*$^{++}$ fold change	[a]*rtrA*$^{++}$ *rtrB*$^{++}$ fold change
CCNA_00867	hypothetical protein	1.14599	1.514781	2.281105
CCNA_00689	PAS-family hybrid histidine kinase/receiver protein	2.906498	1.328666	2.287945
CCNA_01606	Soj/ParA-related ATPase protein	1.964212	1.915675	2.375459
CCNA_03116	hypothetical protein	15.62271	5.15988	2.413485
CCNA_00243	hypothetical protein	2.506292	2.215929	2.467401
fixK	hypoxia transcriptional regulator FixK	2.670775	2.576993	2.584901
CCNA_00788	transporter, major facilitator superfamily	2.334694	3.063141	2.589424
CCNA_03246	hypothetical protein	3.308859	1.173695	2.611941
hipB	HTH transcriptional regulator	2.555504	1.679218	2.616267
CCNA_01684	phenylalanine-4-hydroxylase	2.138845	2.945352	2.658
CCNA_R0185		1.739005	2.967444	2.669954
CCNA_03249	TonB-dependent receptor	3.331869	1.312961	2.68012
CCNA_00822	hypothetical protein	-1.31671	1.591859	2.806675
CCNA_00177	type II secretion pathway protein I	3.633884	2.78581	2.895252
CCNA_00823	LuxR-like DNA-binding protein	1.372682	1.631918	3.484258
CCNA_03923	hypothetical protein	2.29574	3.234311	3.925742
CCNA_03941	hypothetical protein	1.645782	2.257895	4.335826
CCNA_03901	hypothetical protein	6.05203	4.182527	5.633486
CCNA_R0004		18.32111	4.203535	12.06986

[a]Criteria to be considered as differentially expressed: overexpression strain should show a fold-change > 1.5 or < -1.5 compared to empty vector control, and P-value < 0.001.

Table 5.1: *Caulobacter crescentus* strains used in Chapters II and III

Strain #	Genotype	Source
FC19	Wild type CB15	(Poindexter, 1964)
FC438	CB15 *xylX*::pMT585-*lovR vanR*::pMT528-*lovK*	(Purcell et al., 2007)
FC1486	CB15 *lovK*(H180A)	(Reyes Ruiz et al., 2019)
FC1949	CB15 / pRKlac290-P*hfiA*	(Fiebig et al., 2014)
FC1951	CB15 *xylX*::pMT585-*lovR vanR*::pMT528-*lovK* / pRKlac290-P*hfiA*	(Fiebig et al., 2014)
FC3261	CB15 *lovK*(H180A) / pRKlac290-P*hfiA*	(Reyes Ruiz et al., 2019)
FC792	CB15 Δ*lovK/lovR*	(Foreman et al., 2012)
FC1467	CB15 Δ*lovK/lovR vanR*::pMT528-*lovK*	(Reyes Ruiz et al., 2019)
FC1470	CB15 Δ*lovK/lovR vanR*::pMT528-*lovK*(H180A)	(Reyes Ruiz et al., 2019)
FC3262	CB15 *lovK*(H180A) *hfiA*::pMT635-P*hfiA-cat*	(Reyes Ruiz et al., 2019)
FC3263	CB15 *lovK*(H180A) *hfiA*::pMT635-P*hfiA-cat tipR*::pMT585-P*tipR-tipR*	(Reyes Ruiz et al., 2019)
FC3264	CB15 Δ*tipR*	(Reyes Ruiz et al., 2019)
FC3265	CB15 *lovK*(H180A) Δ*tipR*	(Reyes Ruiz et al., 2019)
FC3143	CB15 *lovK*(H180A) Δ*skaH*	(Reyes Ruiz et al., 2019)
FC3146	CB15 *lovK*(H180A) Δ*spdR*	(Reyes Ruiz et al., 2019)
FC3266	CB15 *lovK*(H180A) Δ*skaH* / pRKlac290-P*hfiA*	(Reyes Ruiz et al., 2019)
FC3267	CB15 *lovK*(H180A) *skaH*(H285A) / pRKlac290-P*hfiA*	(Reyes Ruiz et al., 2019)
FC3268	CB15 *lovK*(H180A) *skaH*(D550A) / pRKlac290-P*hfiA*	(Reyes Ruiz et al., 2019)
FC3269	CB15 *lovK*(H180A) Δ*spdR* / pRKlac290-P*hfiA*	(Reyes Ruiz et al., 2019)
FC3270	CB15 *lovK*(H180A) *spdR*(D64A) / pRKlac290-P*hfiA*	(Reyes Ruiz et al., 2019)
FC3271	CB15 *lovK*(H180A) Δ*spdS* / pRKlac290-P*hfiA*	(Reyes Ruiz et al., 2019)
FC3135	CB15 Δ*skaH*	(Reyes Ruiz et al., 2019)
FC3138	CB15 Δ*spdR*	(Reyes Ruiz et al., 2019)
FC3141	CB15 Δ*spdS*	(Reyes Ruiz et al., 2019)
FC3272	CB15 Δ*skaH* / pRKlac290-P*hfiA*	(Reyes Ruiz et al., 2019)
FC3273	CB15 *skaH*(H285A) / pRKlac290-P*hfiA*	(Reyes Ruiz et al., 2019)
FC3274	CB15 *skaH*(D550A) / pRKlac290-P*hfiA*	(Reyes Ruiz et al., 2019)
FC3275	CB15 Δ*spdR* / pRKlac290-P*hfiA*	(Reyes Ruiz et al., 2019)
FC3276	CB15 *spdR*(D64A) / pRKlac290-P*hfiA*	(Reyes Ruiz et al., 2019)
FC3277	CB15 Δ*spdS* / pRKlac290-P*hfiA*	(Reyes Ruiz et al., 2019)
FC3278	CB15 *lovK*(H180A) Δ*skaH xylX*::pMT680	(Reyes Ruiz et al., 2019)
FC3279	CB15 *lovK*(H180A) Δ*skaH xylX*::pMT680-*skaH*	(Reyes Ruiz et al., 2019)
FC3280	CB15 *lovK*(H180A) Δ*spdR xylX*::pMT680	(Reyes Ruiz et al., 2019)
FC3281	CB15 *lovK*(H180A) Δ*spdR xylX*::pMT680-*spdR*	(Reyes Ruiz et al., 2019)

Table 5.1: *Caulobacter crescentus* strains used in Chapters II and III (continued)

Strain #	Genotype	Source
FC3282	CB15 *lovK*(H180A) Δ*spdS xylX*::pMT680	(Reyes Ruiz et al., 2019)
FC3283	CB15 *lovK*(H180A) Δ*spdS xylX*::pMT680-*spdS*	(Reyes Ruiz et al., 2019)
FC3284	CB15 *lovK*(H180A) Δ*skaH xylX*::pMT680 / pRKlac290-P*hfiA*	(Reyes Ruiz et al., 2019)
FC3285	CB15 *lovK*(H180A) Δ*skaH xylX*::pMT680-*skaH* / pRKlac290-P*hfiA*	(Reyes Ruiz et al., 2019)
FC3286	CB15 *lovK*(H180A) Δ*spdR xylX*::pMT680 / pRKlac290-P*hfiA*	(Reyes Ruiz et al., 2019)
FC3287	CB15 *lovK*(H180A) Δ*spdR xylX*::pMT680-*spdR* / pRKlac290-P*hfiA*	(Reyes Ruiz et al., 2019)
FC3288	CB15 *lovK*(H180A) Δ*spdS xylX*::pMT680 / pRKlac290-P*hfiA*	(Reyes Ruiz et al., 2019)
FC3289	CB15 *lovK*(H180A) Δ*spdS xylX*::pMT680-*spdS* / pRKlac290-P*hfiA*	(Reyes Ruiz et al., 2019)
FC3144	CB15 *lovK*(H180A) *skaH*(H285A)	(Reyes Ruiz et al., 2019)
FC3145	CB15 *lovK*(H180A) *skaH*(D550A)	(Reyes Ruiz et al., 2019)
FC3147	CB15 *lovK*(H180A) *spdR*(D64A)	(Reyes Ruiz et al., 2019)
FC3136	CB15 *skaH*(H285A)	(Reyes Ruiz et al., 2019)
FC3137	CB15 *skaH*(D550A)	(Reyes Ruiz et al., 2019)
FC3139	CB15 *spdR*(D64A)	(Reyes Ruiz et al., 2019)
FC3290	CB15 *vanR*::pMT570-MBP-*skaH vanR*::pMT528-HA-*lovK*	(Reyes Ruiz et al., 2019)
FC3291	CB15 *vanR*::pMT570-MBP-*skaH vanR*::pMT528-*spdS*-3xFLAG	(Reyes Ruiz et al., 2019)
FC3292	CB15 *vanR*::pMT570-MBP *vanR*::pMT528-HA-*lovK*	(Reyes Ruiz et al., 2019)
FC3293	CB15 *vanR*::pMT570-MBP *vanR*::pMT528-*spdS*-3xFLAG	(Reyes Ruiz et al., 2019)
FC3140	CB15 *spdR*(D64E)	(Reyes Ruiz et al., 2019)
FC3294	CB15 *spdR*(D64E) / pRKlac290-P*hfiA*	(Reyes Ruiz et al., 2019)
FC3295	CB15 *xylX*::pMT680 *xylX*::pMT585 / pRKlac290-P*hfiA*	(Reyes Ruiz et al., 2019)
FC3296	CB15 *lovK*(H180A) *xylX*::pMT680 *xylX*::pMT585 / pRKlac290-P*hfiA*	(Reyes Ruiz et al., 2019)
FC3297	CB15 *xylX*::pMT680-*rtrA xylX*::pMT585 / pRKlac290-P*hfiA*	(Reyes Ruiz et al., 2019)
FC3298	CB15 *xylX*::pMT680 *xylX*::pMT585-*rtrB* / pRKlac290-P*hfiA*	(Reyes Ruiz et al., 2019)
FC3299	CB15 *xylX*::pMT680-*rtrA xylX*::pMT585-*rtrB* / pRKlac290-P*hfiA*	(Reyes Ruiz et al., 2019)
FC3300	CB15 *xylX*::pMT680 *xylX*::pMT585	(Reyes Ruiz et al., 2019)
FC3301	CB15 *lovK*(H180A) *xylX*::pMT680 *xylX*::pMT585	(Reyes Ruiz et al., 2019)
FC3302	CB15 *xylX*::pMT680-*rtrA xylX*::pMT585	(Reyes Ruiz et al., 2019)
FC3303	CB15 *xylX*::pMT680 *xylX*::pMT585-*rtrB*	(Reyes Ruiz et al., 2019)
FC3304	CB15 *xylX*::pMT680-*rtrA xylX*::pMT585-*rtrB*	(Reyes Ruiz et al., 2019)
FC3305	CB15 Δ*spdS* Δ*hfsJ*	(Reyes Ruiz et al., 2019)
FC1974	CB15 Δ*hfsJ*	(Fiebig et al., 2014)
FC3357	CB15 Δ*rtrA*	(Reyes Ruiz et al., 2019)

Table 5.1: *Caulobacter crescentus* strains used in Chapters II and III (continued)

Strain #	Genotype	Source
FC3358	CB15 Δ*rtrB*	(Reyes Ruiz et al., 2019)
FC3359	CB15 Δ*rtrA* Δ*rtrB*	(Reyes Ruiz et al., 2019)
FC3360	CB15 *spdR*(D64E) Δ*rtrA*	(Reyes Ruiz et al., 2019)
FC3361	CB15 *spdR*(D64E) Δ*rtrB*	(Reyes Ruiz et al., 2019)
FC3362	CB15 *spdR*(D64E) Δ*rtrA* Δ*rtrB*	(Reyes Ruiz et al., 2019)
FC3363	CB15 *lovK*(H180A) *vanR*::pMT528	(Reyes Ruiz et al., 2019)
FC3364	CB15 *lovK*(H180A) *vanR*::pMT528-*lovK*	(Reyes Ruiz et al., 2019)
FC3365	CB15 Δ*spdS* Δ*spdR*	(Reyes Ruiz et al., 2019)

Table 5.2: *Escherichia coli* strains, plasmids and primers used in Chapter II and III

Strain #	Genotype	Description	Source
FC929	Top10		Invitrogen
FC3	MT607 / pRK600		-
FC803	Rosetta (DE3) pLysS		Novagen
Plasmids for expression in *Caulobacter crescentus*			
MTLS 4389	Top10 / pMT680 (pXGFPC-6)		(Thanbichler et al., 2007)
MTLS 4259 (FC339)	Top10 / pMT585 (pXGFPC-2)		(Thanbichler et al., 2007)
MTLS 4368	Top10 / pMT570 (pVGFPC-5)		(Thanbichler et al., 2007)
FC338	Top10 / pMT528 (pVCHYC-1)		(Thanbichler et al., 2007)
MTLS 4321	DH10B / pMT635 (pMCS-4)		(Thanbichler et al., 2007)
FC364	Top10 / pMT585-*lovR*		(Purcell et al., 2007)
FC1660	Top10 / pMT528-*lovK*		(Purcell et al., 2007)
FC490	Top10 / pMT528-*lovK(H180A)*	UP F: cttacatATGGAAGACTATTCGG UP R*: GTTCTTCACACGggcATCCA DN F*: GGATgccCGTGTGAAGAACA DN R: aattggtaccATGATCAAAAGCGTCCTTGG	(Reyes Ruiz et al., 2019)
FC3232	Top10 / pMT680-*spdR*	F: ttatcatatgGCGGATATCGGAGAACT R: ttatgaattcAACTTAAACGCTTCTACCGC	(Reyes Ruiz et al., 2019)
FC3233	Top10 / pMT680-*spdS*	F: ttatcatatgGCTGACGTTTCGTTTGC R: ttatgaattcATCACAACTCAAGCCCCTGT	(Reyes Ruiz et al., 2019)
FC3234	Top10 / pMT680-*skaH*	F: ttatcatatgGACGACGCCCCAAGAG R: ttatggtaccTGGTTCAGGGCGTCTTCG	(Reyes Ruiz et al., 2019)
FC3235	Top10 / pMT680-*rtrA*	F: ttatcatatgAGCGAAGAAACCGAAGGC R: ttatggtaccCGGCGAGTGTTGAGCTCAG	(Reyes Ruiz et al., 2019)
FC3236	Top10 / pMT585-*rtrB*	F: ttatcatatgAACGATCGACCCGCA R: ttatggtaccCAAGGTCCAAGAACAGGGCT	(Reyes Ruiz et al., 2019)
FC2109	Top10 / pMT635-P*hfiA-cat*	P*hfiA* amplified from *Caulobacter crescentus* DNA: UP F: taatcatatgTCGTCTGGGCGTCCCTGAT UP R*: TTTCTCCATCACTGACAACATCCTGTC *cat* amplified from pMT680: DN F*: ATGTTGTCAGTGATGGAGAAAAAAATCACTGGA DN R: taatggtaccTCATCATGCCTAGGTTACGC	(Reyes Ruiz et al., 2019)
FC3237	Top10 / pMT585-P*tipR-tipR*	F: ttatcatatgACCCATGCCCATGCCCATG R: ttatgaattcGCGTCTAGCTCAGCAGGT	(Reyes Ruiz et al., 2019)
FC3238	Top10 / pMT570-*MBP-skaH*		(Reyes Ruiz et al., 2019)
FC3239	Top10 / pMT570-*MBP*	F: ttatcatatgAAAATCGAAGAAGGTAAA R: ttatggtaccTCAAGTCTGCGCGTCTTTCAGG	(Reyes Ruiz et al., 2019)

Table 5.2: *Escherichia coli* strains, plasmids and primers used in Chapter II and III (continued)

Strain #	Genotype	Description	Source
pAF453	Top10 / pMT528-*HA-lovK*	PCR #1: amplified from *Caulobacter crescentus* DNA with F1 and R PCR #2: amplify with F2 and R using PCR #1 as template F1: TGTTCCAGATTACGCAGGTGAAGACTATTCGGAA F2: gaggaaacgcatatgTACCCATACGATGTTCCAGAT R: aattggtaccATGATCAAAAGCGTCCTTGG	(Reyes Ruiz et al., 2019)
FC3240	Top10 / pMT528-*spdS-3xFLAG*		(Reyes Ruiz et al., 2019)
Allele replacement plasmids			
FC55	DH10B / pNPTS138	Allele replacement plasmid, KanR, SacB	
pAF494	Top10 / pNPTS138-Δ*hfsJ*	For in-frame knockout of *hfsJ*	(Fiebig et al., 2014)
FC3241	Top10 / pNPTS138-Δ*skaH*	UP F: ttataagcttCGCCAGCGTCATCAGAGTTT UP R*: CGTTGTAGGCGTCGTCCATGATTCGA DN F*: CGACGCCTACAACGTGCACCAGATCGC DN R: ttatggatccGGACGTCGCCGGGATGAG	(Reyes Ruiz et al., 2019)
FC3242	Top10 / pNPTS138-Δ*spdR*	UP F: ttataagcttATGTCGACTCGGTGTTCCTG UP R*: TGCGGCGATCCTGGCGAAACGGGC DN F*: AGGATCGCCGCAACCAGTTCTCCG DN R: ttatggatccAGCGCATGAGCCTGGTTC	(Reyes Ruiz et al., 2019)
FC3243	Top10 / pNPTS138-Δ*spdS*	UP F: ttataagcttTCGGCGCTCATCGGATTTTC UP R*: GTTTGCCGGACATACAGGGGCTTGAG DN F*: TGTCCGGCAAACGAAACGTCAGCCAT DN R: ttatggatccGCAGGTTGTAGGGGATCACC	(Reyes Ruiz et al., 2019)
FC3244	Top10 / pNPTS138-Δ*tipR*	UP F: ttataagcttTCTCGACCAGGGCCTTGTAG UP R*: AGGTGCCTTATGGAAGTCCTGTGCGGC DN F*: TCCATAAGGCACCTGCTGAGCTAGAC DN R: ttatggatccCCTGGGCCGCATCGTCTC	(Reyes Ruiz et al., 2019)
pAF444	Top10 / pNPTS138-*lovK(H180A)*	UP F: aataggatccGATACGTTGAGCGTCCA UP R*: GTTCTTCACACGggcATCCA DN F*: GGATgccCGTGTGAAGAACA DN R: tcatgaattcAAAAGCGTCCTTGG	(Reyes Ruiz et al., 2019)
FC3245	Top10 / pNPTS138-*skaH(H285A)*	UP F: ttataagcttCGTCCCGATCCGCGATTG UP R*: TTCTTGATCCGGGCGTCCA DN F*: AGCTGGACGCCCGGATCAAGA DN R: ttatggatccTTCAACAATACGCTCGGGC	(Reyes Ruiz et al., 2019)
FC3246	Top10 / pNPTS138-*skaH(D550A)*	UP F: ttataagcttGAACTGTTCCTGACGCCCC UP R*: TTGAGATCGGCGGCGAGCACC DN F*: GTGCTCGCCGCCGATCTCAA DN R: ttatggatccGCGTGGCTTTCCAGATTGTG	(Reyes Ruiz et al., 2019)

Table 5.2: *Escherichia coli* strains, plasmids and primers used in Chapter II and III (continued)

Strain #	Genotype	Description	Source
FC3247	Top10 / pNPTS138-*spdR(D64A)*	UP F: ttataagcttAGCGACGAGAGCAAGTTCAT UP R*: ATGCTGTTCTGGCCATGCGG DN F*: GCATGGCCAGAACAGCATG DN R: ttatgaattcTTCACGCCATGACCTCGATC	(Reyes Ruiz et al., 2019)
FC3248	Top10 / pNPTS138-*spdR(D64E)*	UP F: ttataagcttAGCGACGAGAGCAAGTTCAT UP R*: ATGCTGTTCTGGAGATGCGG DN F*: GCATCTCCAGAACAGCATG DN R: ttatgaattcTTCACGCCATGACCTCGATC	(Reyes Ruiz et al., 2019)
FC3354	Top10 / pNPTS138-Δ*rtrA*	UP F: ttatgaattcTTCTCGAAGACGAAGTGGCC UP R*: GGCCCCGCCTTCGGTTTCTTCGCTCA DN F*: GAAGGCGGGGCCTGAGCTCAACACTC DN R: ttatggatccAAGAAGGTTGGCCTGATCCC	(Reyes Ruiz et al., 2019)
FC3355	Top10 / pNPTS138-Δ*rtrB*	UP F: ttataagcttGCGCTTCATGATGTCGAGGT UP R*: CATGGCCTGCTGAAAGGTGATGCCA DN F*: CAGCAGGCCATGGTGGACGACGACG DN R: ttatctgcagAAGCCGACCTTCTTGCTGG	(Reyes Ruiz et al., 2019)
FC3356	Top10 / pNPTS138-Δ*spdS*Δ*spdR*	UP F: ttataagcttATGTCGACTCGGTGTTCCTG UP R*: GTTTGCGATCCTGGCGAAACGGGC DN F*: AGGATCGCAAACGAAACGTCAGCCAT DN R: ttatggatccTCAGGTAAATCGCCGTGGAA	(Reyes Ruiz et al., 2019)
***lacZ* fusions plasmids**			
FC54	S17-1 / pRKlac290		(Gober and Shapiro, 1992)
FC1948	Top10 / pRKlac290-P*hfiA*		(Fiebig et al., 2014)
Heterologous protein expression plasmids (expression was conducted in Rosetta(DE3)pLysS strains)			
FC155	DH10B / pET28a		-
FC3249	DH5α / pET23b-His6Sumo		(Stein et al., 2016)
FC3250	Top10 / pET28a-His6-*rtrB*	F: ttatcatatgAACGATCGACCCGCA R: ttatggtaccCAAGGTCCAAGAACAGGGCT	(Reyes Ruiz et al., 2019)
	Top10 / pET23b-His6Sumo-*rtrA*	F: ttataccggtGGAAGCGAAGAAACCGAAGGC R: ttatgcggccgcTCAGGCCCCGTCGCGCGA	(Reyes Ruiz et al., 2019)
FC3252	Top10 / pET28a-His6-*spdR(D64E)*	F: ttatcatatgGCGGATATCGGAGAACT R: ttatgaattcAACTTAAACGCTTCTACCGC	(Reyes Ruiz et al., 2019)
Bacterial two-hybrid plasmids			
FC257	BTH101 (F- cya-99, araD139, galE15, galK16, rpsL1 (Str resistant), hsdR2, mcrA1, mcrB1)		(Karimova et al., 1998)
FC255	Top10 / pKT25		(Karimova et al., 1998)

Table 5.2: *Escherichia coli* strains, plasmids and primers used in Chapter II and III (continued)

Strain #	Genotype	Description	Source
FC253	Top10 / pUT18c		(Karimova et al., 1998)
FC256	Top10 / pKT25-zip		(Karimova et al., 1998)
FC254	Top10 / pUT18c-zip		(Karimova et al., 1998)
FC3253	Top10 / pKT25-*skaH*	F: ttatggatccTCGAATCATGGACGACGCC R: ttatggtaccCGGATGATCATGCCCGTCC	(Reyes Ruiz et al., 2019)
FC3254	Top10 / pKT25-*lovK*	F: ttatggatccTGGGAGCTTAGCCTTGGAAG R: ttatggtaccTGGCTATTGCGTCCCATTGA	(Reyes Ruiz et al., 2019)
FC3255	Top10 / pKT25-*spdS**	F: ttatggatccTACGGGTCTCTACGCATGG R: ttatggtaccGCATCACAACTCAAGCCCCT	(Reyes Ruiz et al., 2019)
FC3256	Top10 / pKT25-*skaHΔREC*	F: ttatggatccTCGAATCATGGACGACGCC R: ttatggtaccCTACGGGGCGCCTTGAATCATC	(Reyes Ruiz et al., 2019)
FC3257	Top10 / pUT18c-*skaH*	F: ttatggatccTCGAATCATGGACGACGCC R: ttatggtaccCGGATGATCATGCCCGTCC	(Reyes Ruiz et al., 2019)
FC3258	Top10 / pUT18c-*lovK*	F: ttatggatccTGGGAGCTTAGCCTTGGAAG R: ttatggtaccTGGCTATTGCGTCCCATTGA	(Reyes Ruiz et al., 2019)
FC3259	Top10 / pUT18c-*spdS**	F: ttatggatccTACGGGTCTCTACGCATGG R: ttatggtaccGCATCACAACTCAAGCCCCT	(Reyes Ruiz et al., 2019)
FC3260	Top10 / pUT18c-*skaHΔREC*	F: ttatggatccTCGAATCATGGACGACGCC R: ttatggtaccCTACGGGGCGCCTTGAATCATC	(Reyes Ruiz et al., 2019)
FC3346	Top10 / pKT25-*skaH(H285A)*	F: ttatggatccTCGAATCATGGACGACGCC R: ttatggtaccCGGATGATCATGCCCGTCC	(Reyes Ruiz et al., 2019)
FC3347	Top10 / pKT25-*skaH(D550)*	F: ttatggatccTCGAATCATGGACGACGCC R: ttatggtaccCGGATGATCATGCCCGTCC	(Reyes Ruiz et al., 2019)
FC3348	Top10 / pKT25-*lovK(H180A)*	F: ttatggatccTGGGAGCTTAGCCTTGGAAG R: ttatggtaccTGGCTATTGCGTCCCATTGA	(Reyes Ruiz et al., 2019)
FC3349	Top10 / pKT25-*spdS(H248A)**	F: ttatggatccTACGGGTCTCTACGCATGG R: ttatggtaccGCATCACAACTCAAGCCCCT	(Reyes Ruiz et al., 2019)
FC3350	Top10 / pUT18c-*skaH(H285A)*	F: ttatggatccTCGAATCATGGACGACGCC R: ttatggtaccCGGATGATCATGCCCGTCC	(Reyes Ruiz et al., 2019)
FC3351	Top10 / pUT18c-*skaH(D550A)*	F: ttatggatccTCGAATCATGGACGACGCC R: ttatggtaccCGGATGATCATGCCCGTCC	(Reyes Ruiz et al., 2019)
FC3352	Top10 / pUT18c-*lovK(H180A)*	F: ttatggatccTGGGAGCTTAGCCTTGGAAG R: ttatggtaccTGGCTATTGCGTCCCATTGA	(Reyes Ruiz et al., 2019)
FC3353	Top10 / pUT18c-*spdS(H248A)**	F: ttatggatccTACGGGTCTCTACGCATGG R: ttatggtaccGCATCACAACTCAAGCCCCT	(Reyes Ruiz et al., 2019)

Table A3.1: Analysis of Lake Michigan water used for barcoded Tn-Himar fitness experiments

Water collection[a]	Nov. 30	Dec. 6	Dec. 9	Dec. 12
Date of experiment	Dec. 2	Dec. 6	Dec. 9	Dec. 12
Water temperature (°C)	7.5	7	3	2
Air temperature (°C)	7.8	4.4	-3.3	-7.2
pH	5.8	5.8	5.8	5.8
Phosphate	5 ppm	5 ppm	5 ppm	5 ppm
Nitrate/Nitrite[b]	undetectable	undetectable	undetectable	undetectable

[a]Water collection occurred in 2016 at Promontory Point, Hyde Park, Chicago, Illinois, USA.
[b]Limit of detection: Nitrite 0.15 mg/L, Nitrate 1 mg/L

Table A3.2: Transposon library statistics

Library	Unique insertion sites	Total TA sites	Percent sites hit	Average transposons per ORF	Mean reads
Tn5[a]	115,788	N/A	2.9	30	90–150 (per Tn)
Tn-Himar[b] (BarSeq)	46,395	49,437	93.8	24	7612 (per gene)

[a]Tn5 can insert into any dinucleotide site. We identified 115,788 insertion sites in an initial population of ~300,000 clones.
[b]TnHimar specifically inserts into TA dinucleotides.

Table A3.3: Average fitness scores for hypothetical genes and genes of unknown function with fitness scores less than -1.2 and greater than +1.2 (bold) in at least one condition

Locus number	Description	Defined	Filtered Lake Water	Unfiltered Lake Water
CCNA_00375	No conserved domains	**-1.21**	**-2.57**	**-2.83**
CCNA_01724	COG4649, TPR_21 pfam09976	**-2.58**	**-1.25**	**-1.28**
CCNA_03860	COG3786, YkuD superfamily	**1.33**	**1.53**	**1.41**
CCNA_03864	DUF3576, pfam12100	**-1.71**	**-1.32**	**-1.57**
CCNA_03909	No conserved domains	**-1.20**	**-1.76**	**-1.84**
CCNA_00927	No conserved domains	**-2.56**	**-2.11**	-0.67
CCNA_02875	No conserved domains	**-1.54**	**-2.19**	-0.39
CCNA_00895	No conserved domains	-0.19	**-2.14**	**-1.44**
CCNA_00913	No conserved domains	-0.28	**-2.48**	**-1.21**
CCNA_00519	No conserved domains	**-1.66**	-0.37	-0.61
CCNA_01176	DUF2849, pfam11011	**-4.98**	-0.81	-0.84
CCNA_01178	DUF934, pfam06073, COG3749	**-4.36**	-0.60	-0.64
CCNA_01676	TamB, COG2911, pfam04357	**-6.40**	-0.52	-0.74
CCNA_01219	No conserved domains	**1.68**	-0.70	0.78
CCNA_02669	Uncharacterized membrane protein, DUF3422, pfam11902, COG4949	**-4.41**	0.38	0.20
CCNA_03273	COG4944, DUF1109, pfam06532 NrsF, anti-sigF	**-1.63**	-0.39	-0.27
CCNA_03692	COG1975, XdhC/CoxF family	**-4.06**	-0.98	-0.65
CCNA_02796	No conserved domains	-0.45	**-1.20**	-0.52
CCNA_03420	No conserved domains	0.40	**-1.46**	0.42
CCNA_03883	No conserved domains	0.05	**-1.25**	0.99

Table A3.3: Average fitness scores for hypothetical genes and genes of unknown function with fitness scores less than -1.2 and greater than +1.2 (bold) in at least one condition (continued)

Locus number	Description	Defined	Filtered Lake Water	Unfiltered Lake Water
CCNA_03984	No conserved domains	0.01	**-1.22**	-0.25
CCNA_01286	No conserved domains	0.25	-0.25	**-1.27**
CCNA_02053	No conserved domains	-0.56	-0.74	**-1.34**
CCNA_02160	No conserved domains	0.26	-1.02	**-1.38**
CCNA_03015	No conserved domains	-0.15	0.57	**-1.46**
CCNA_03945	No conserved domains	-0.43	0.24	**-1.90**

Appendix C: Genome-scale fitness profile of *Caulobacter crescentus* grown in natural freshwater

The work provided in this Appendix has been adapted from the manuscript: Genome-scale fitness profile of *Caulobacter crescentus* grown in natural freshwater by the authors Kristy L. Hentchel, Leila M. Reyes Ruiz, Patrick D. Curtis, Aretha Fiebig, Maureen L. Coleman, Sean Crosson. This manuscript was published in the ISME Journal (DOI: 10.1038/s41396-018-0295-6). I am a second author on this manuscript and my contributions to this work are as follows: *1)* I optimized the different growth conditions in which we cultivated *C. crescentus*, *2)* I contributed to the design of the laboratory experiments carried on for this manuscript, and *3)* I, together with Kristy Hentchel, carried on all the *C. crescentus* wild-type growth, and Barseq and Tn-Seq libraries growth, collection of the samples and DNA-extraction experiments.

Abstract

Bacterial genomes evolve in complex ecosystems and are best understood in this natural context, but replicating such conditions in the lab is challenging. We used transposon sequencing to define the fitness consequences of gene disruption in the bacterium *Caulobacter crescentus* grown in natural freshwater, compared to axenic growth in common laboratory media. Gene disruptions in amino acid and nucleotide biosynthesis pathways and in metabolic substrate transport machinery impaired fitness in both lake water and defined minimal medium relative to complex peptone broth. Fitness

in lake water was enhanced by insertions in genes required for flagellum biosynthesis and reduced by insertions in genes involved in biosynthesis of the holdfast surface adhesin. We further uncovered numerous hypothetical and uncharacterized genes for which disruption impaired fitness in lake water, defined minimal medium, or both. At the genome scale, the fitness profile of mutants cultivated in lake water was more similar to that in complex peptone broth than in defined minimal medium. Microfiltration of lake water did not significantly affect the terminal cell density or the fitness profile of the transposon mutant pool, suggesting that *Caulobacter* does not strongly interact with other microbes in this ecosystem on the measured timescale. Fitness of select mutants with defects in cell surface biosynthesis and environmental sensing were significantly more variable in lake water than in defined medium, presumably owing to day-to-day heterogeneity in the lake environment. This study reveals genetic interactions between *Caulobacter* and a natural freshwater environment, and provides a new avenue to study gene function in complex ecosystems.

Introduction

Environments inhabited by microbial cells have significant microscale heterogeneity. This is well recognized in biofilms, soils, and host-associated habitats (Vos et al., 2013;van Gestel et al., 2015;Cao and Goodrich-Blair, 2017). Free-living aquatic bacteria often encounter chemical gradients that can appear as ephemeral patches, arising from algal exudates, sinking particles, or lysis events (Azam and Malfatti, 2007;Stocker, 2012) and they may have to cope with prolonged periods of nutrient

scarcity. In addition, these bacteria face interspecies interactions, protistan predators, and viruses, as well as fluctuations in physical conditions such as temperature and light. These biotic and abiotic factors have driven myriad adaptations that enable survival and reproduction in natural environments.

In contrast with this natural complexity, studies on microbial physiology and gene function have traditionally relied on simplified experimental conditions. Thus it is not surprising that a large fraction of bacterial genes remain uncharacterized. Recently developed transposon sequencing (Tn-Seq) approaches (van Opijnen and Camilli, 2013) now enable rapid phenotypic assessment of thousands to millions of distinct mutant strains, and these methods have been used to interrogate gene function in a variety of in vitro and host-associated conditions (Kwon et al., 2016;Price et al., 2018). More recently, transposon mutagenesis approaches in which each transposon carries a unique 20-bp barcode sequence have been developed (Wetmore et al., 2015); each insertion site is associated with a short barcode, and the abundance of all mutant strains in the pool can be assessed by simple amplicon sequencing.

Here, we used a barcoded Tn-Seq approach to identify genes affecting fitness in *Caulobacter crescentus* strain CB15, cultivated in natural freshwater from Lake Michigan, Illinois, USA. As a well-characterized and genetically tractable bacterium originally isolated from a pond in California in 1960 (Poindexter, 1964), this strain is well suited for this study. Briefly, *C. crescentus* is among a group of dimorphic prosthecate (i.e. stalked) bacteria that attach to surfaces, often forming epibiotic interactions with algae and plant material (Poindexter, 2006). More broadly, members of the genus *Caulobacter* are

common in soil ecosystems, where they likely play an important role in plant matter decomposition (Wilhelm, 2018). In aquatic systems, *Caulobacter* interactions with substrates contribute to biopolymer mineralization, and have been proposed to enhance productivity of aquatic ecosystems (Allen, 1971;Poindexter, 2006). However, *C. crescentus* (hereafter referred to simply as *Caulobacter*) is typically grown in an artificial medium consisting of dilute peptone and yeast extract (PYE) or in a defined medium consisting of mineral salts and a single carbon source such as xylose (M2X) (Ely, 1991), neither of which adequately represents natural freshwater. PYE is replete with peptides, amino acids, and a range of carbon sources, while M2X requires *Caulobacter* to synthesize all cellular building blocks from salts and a simple sugar. Natural freshwaters, by contrast, contain an undefined, complex mixture of organic and inorganic nutrient sources (Drever, 1997). In many freshwater systems, essential nutrients including phosphorus and labile carbon do not accumulate to high concentrations (Carlson, 1977;Brunberg et al., 2002). We predicted that genes that are dispensable in PYE (Christen et al., 2011) or M2X medium would be important for fitness in complex natural freshwater, and that these genes would offer insights into *Caulobacter* physiology in a *bona fide* freshwater system.

Materials and methods

Bacterial strains and primers. Strains and primers used in this Appendix are listed in Table S1 of (Hentchel et al., 2019). All primers were synthesized by Integrated DNA Technologies (Coralville, IA, USA).

Growth media. *Caulobacter crescentus* strain CB15 (Poindexter, 1964) was grown in PYE medium [0.2% peptone (Fisher Scientific), 0.1% yeast extract (Fisher Scientific), 0.5 mM $MgSO_4$, 0.5 mM $CaCl_2$] or M2X minimal defined medium [6.1 mM Na_2HPO_4, 3.9 mM KH_2PO_4, 9.3 mM NH_4Cl, 0.5 mM $MgSO_4$, 10 μM $FeSO_4$ (EDTA chelate; Sigma Chemical Co.), 0.25 mM $CaCl_2$] supplemented with 0.15% xylose (Ely, 1991). *Escherichia coli* strains were grown in LB broth (1% peptone, 0.5% yeast extract, 0.5% NaCl). Solid growth media included 1.5% agar.

Lake water collection. Water from Lake Michigan was collected at Promontory Point, Chicago, Illinois, USA (Latitude: 41.794, Longitude: -87.579), on four dates in 2016 (Nov 30, Dec 6, Dec 9, and Dec 12). We measured water temperature, phosphate and nitrate/nitrite level (Aquacheck Water Quality Test Strips; Hach), and pH (pH indicator strips, Fisher Scientific) at the time of collection (Table A3.1). Lake water was filtered using Nalgene™ Rapid-Flow™ Sterile Disposable 0.1 μm Filter Units with PES Membrane (Thermo Scientific).

Measurement of Caulobacter *growth in lake water*. Colonies of *Caulobacter* were inoculated into 2 mL of PYE in glass culture tubes (13 × 100 mm) and grown overnight at 30°C with shaking at 200 rpm, for a total of five biological replicates. At saturation, 1 mL of culture was centrifuged at 8,000 ×*g* and washed twice in 1 mL of filtered lake water (see Fig A3.2 for information on lake water). The washed pellet was resuspended in

filtered lake water to a final OD_{660} of 0.1, and 0.5 µL (approximately 0.5–1 × 10^5 cells) was inoculated into 5 mL of filtered lake water in a glass culture tube (20 mm × 150 mm) in technical duplicate. Cultures were grown at 30°C with shaking at 200 rpm. To monitor growth, 20 µL of culture was removed at various time points, serially diluted, and titered onto PYE agar plates, which were incubated at 30°C for 2 days. Growth was monitored by enumeration of colony forming units (CFUs).

Construction of barcoded Tn-Himar mutant library. The recipient strain (*Caulobacter*) was grown overnight in 2 mL of PYE at 30°C with shaking at 200 rpm. This starter culture was used to inoculate 20 mL of PYE and grown at 30°C overnight with shaking at 200 rpm until saturated. The donor *E. coli* strain (APA752, gift from Adam Deutschbauer, University of California-Berkeley, USA), carrying the pKMW3 (kanamycin resistant) Himar transposon vector library (Wetmore et al., 2015), was inoculated into 20 mL of LB containing kanamycin (30 µg mL^{-1}) and diaminopimelate (DAP; 300 µM) and grown overnight at 37°C with shaking at 200 rpm; the *E. coli* Himar donor strain is a DAP auxotroph, and thus requires addition to the medium. To conjugate the barcoded transposon pool into *Caulobacter*, the recipient strain and donor strains were each centrifuged at 8,000 ×*g* for 2 min and resuspended in a total volume of 500 µL of PYE medium. The cultures were combined at a 10:1 ratio of recipient to donor and mixed by gentle pipetting. The mixed culture was centrifuged again at 8,000 ×*g*, and the supernatant decanted. The cells were resuspended in 30 µL of PYE, spotted onto a PYE agar plate containing diaminopimelate (300 µM), and incubated overnight at 30°C. After

growth, the mating spot was scraped from the plate and resuspended in 6.5 mL of PYE. This suspension was spread evenly (500 µL per plate) over 14 large (150 × 15 mm) PYE agar plates containing 25 µg mL^{-1} kanamycin and incubated for approximately 3 days at 30°C. Cells were harvested from all the plates and inoculated into 400 mL of PYE containing 5 µg mL^{-1} kanamycin. This cell mixture was grown at 30°C with shaking at 200 rpm for three doublings. Cells were centrifuged at 8,000 ×*g*, resuspended in 70 mL of PYE containing 15% glycerol, and stored as 1 mL aliquots at -80°C.

Mapping of the sites of Tn-Himar insertion in the Caulobacter *BarSeq library* (see Fig S1 in (Hentchel et al., 2019) for graphical overview). Genomic DNA was extracted using guanidium thiocyanate as previously described (Pitcher et al., 1989). The DNA was sheared (~300 bp fragments), cleaned with a standard bead protocol, end-repaired and A-tailed, and a custom double-stranded Y adapter was ligated. The custom adapter was prepared by annealing Mod2_TS_Univ and Mod2_TruSeq (Table S1 in (Hentchel et al., 2019)) as described (Wetmore et al., 2015). The sheared fragments containing transposons were enriched by PCR using the primers Nspacer_BarSeq_pHIMAR and P7_MOD_TS_index1 using GoTaq® Green Master Mix according to the manufacturer's protocol in a 100-µL volume with the following cycling conditions: 94°C for 2 min, 25 cycles at 94°C for 30 s, 65°C for 20 s, and 72°C for 30 s, followed by a final extension at 72°C for 10 min. After a second bead cleanup, the *Caulobacter* library was sequenced using a standard Illumina sequencing primer on an Illumina HiSeq2500 at the University of Chicago Genomics Facility with a 150-bp single-end read. The locations of Himar

transposon insertions were aligned and mapped using BLAT (Kent, 2002), and unique barcode sequences were associated with their corresponding genome insertion location using a custom Perl script (MapTnSeq.pl). Sets of barcodes that consistently map to one location in the genome were identified using a custom Perl script (DesignRandomPool.pl). This ensures that each unique barcode is properly assigned to a single insertion site. These scripts have been described by Wetmore and colleagues (Wetmore et al., 2015) and are available at https://bitbucket.org/berkeleylab/feba. For all analyses, reads were mapped to the *C. crescentus* NA1000 genome (accession CP001340) (Marks et al., 2010), which is more comprehensively annotated (Schrader et al., 2014) than the highly-related CB15 parent strain.

Cultivation of the Tn-Himar library. An aliquot of the *Caulobacter* library (2 mL) from a glycerol stock was inoculated into 18 mL of PYE, split into two tubes (20 × 150 mm) with 10 mL each, and grown in a cell culture roller drum (Fisher Scientific) at 30°C for 4 h. The tubes were then moved to a 30°C incubator with shaking at 200 rpm for an additional 2 h. Cultures were combined and centrifuged for 20 min at 3,000 ×*g* at 4°C. The cell pellet was resuspended and washed in 10 mL of filtered lake water, and centrifuged again at 3,000 ×*g* for 20 min at 4°C. The resulting pellet was resuspended in 5 mL of filtered lake water, and the OD_{660} measured. Flasks containing filtered or unfiltered lake water (7.5 L total volume per condition, divided over 3 flasks) were inoculated with the washed library with the aim of an initial starting concentration of approximately 2.5×10^7 total cells per flask (Fig S2 in (Hentchel et al., 2019)). Flasks were incubated at 30°C with shaking at

150 rpm. At 0 and 64 h, an aliquot of culture was removed from each flask for CFU enumeration on PYE agar plates (Fig S3 in (Hentchel et al., 2019)). After ~64 h of growth, cells from all three flasks were collected by filtration using an Express Plus Membrane 0.22 µm filter (Millipore). Filters were stored at -80°C until needed. To mimic saturating conditions with the same number of doublings in defined M2X and complex PYE laboratory medium as in lake water, we inoculated cultures at a concentration that after five doublings (the estimated number of doublings in lake water), the cultures reached saturation. Cells were pelleted at 10,000 ×g for 1 min and stored at -20°C. Genomic DNA from all samples was extracted using guanidium thiocyanate as previously described (Pitcher et al., 1989), with the exception that the lake water samples were lysed directly from the filters they were collected on. DNA quality and quantity was measured using a NanoDrop^{OneC} (Thermo Scientific).

Amplification and sequencing of Tn-Himar barcodes. PCR amplification for each sample was performed as previously described (Wetmore et al., 2015) (Fig S1 in (Hentchel et al., 2019)) using a standard reaction protocol for Q5 DNA polymerase (New England BioLabs) with the primers BarSeq_P1 and 1 of 16 forward primers (BarSeq_P2_IT001 to BarSeq_P2_IT016; Table S1 in (Hentchel et al., 2019)) containing unique 6-bp TruSeq indexes that were sequenced using a separate index primer. Cycling conditions were as follows: 98°C for 4 min followed by 25 cycles of 30s at 98°C, 30s at 55°C, and 30s at 72°C, followed by a final extension at 72°C for 5 min. PCR products were purified using GeneJET PCR Purification Kit (Thermo Scientific). Purified samples were run on a 2.5%

agarose gel to confirm correct product size (~200 bp). A total of 10 µL per purified PCR product was pooled, assessed for quality, and quantified using a Bioanalyzer. The amplified barcodes from the reference (PYE) and treatment (M2X, unfiltered lake water, and filtered lake water) were sequenced on an Illumina HiSeq4000 at the University of Chicago Genomics Facility, multiplexing all 16 samples in one lane with 50-bp single-end reads. All sequence data have been deposited in the NCBI Sequence Read Archive under BioProject accession PRJNA429486; BioSample accession SAMN08348121; SRA accession SRP128742.

Analysis of Tn-Himar strain fitness. We followed the fitness calculation protocol of Wetmore and colleagues (Wetmore et al., 2015), using scripts available at https://bitbucket.org/berkeleylab/feba. Briefly, the total count of each barcode in each sample was calculated using a Perl script (MultiCodes.pl) and, from this table of barcodes, strain fitness was calculated using an R script (FEBA.R). The fitness of each strain was calculated as a normalized $\log_2$ ratio of barcode counts in the treatment sample to counts in the PYE reference sample. The fitness of genes was calculated as the weighted average of strain fitness values, the weight being inversely proportional to a variance metric based on the total number of reads for each strain; this weighting is fully described by Wetmore and colleagues (Wetmore et al., 2015). Successful gene fitness calculations required at least 3 reads per strain and 30 reads for each of the 16 samples. Insertions in the first 10% or last 10% of a gene were not considered in gene fitness calculations. The

complete data set of fitness values for each condition is listed in Table S2 in (Hentchel et al., 2019).

To assess the distribution of fitness scores, we calculated the standard deviation for each condition using the frequency distribution of the mean fitness value of each gene (filtered lake water = 0.41, unfiltered lake water = 0.40, defined medium = 1.1). When the outlier region of the defined medium dataset (< -2.5) was removed, the calculated standard deviation was 0.36; therefore, a standard deviation of 0.4 was chosen and applied to all conditions. Genes with a mean fitness value approximated at ± 3σ from the mean (less than -1.2 and greater than +1.2) were selected for further examination. We also examined t-values, the fitness value of a gene divided by a variance metric, based on the total number of reads for each gene (as previously described (Wetmore et al., 2015)), to provide a metric to assess the significance of fitness values (Table S3 in (Hentchel et al., 2019)).

To identify genes showing differential fitness across lake water samples, we fit a linear model with two factors, sampling day and filtration treatment (filtered or unfiltered). The model was implemented using the functions *lmfit, eBayes, and topTable* in the R package *limma* (Ritchie et al., 2015). Genes were identified as having differential fitness across either sampling days or filtration treatment, with a false discovery rate threshold of 0.05.

Analysis of Caulobacter *Tn5-seq fitness.* A *Caulobacter* Tn5 insertion library containing an estimated 3×10^5 clones was constructed as previously described (Curtis and Brun,

2014). The lake water fitness experiment for the Tn5 library (from Lake Michigan water collected in April 2016) was performed similarly to the Tn-Himar library experiments with the following modifications: A total of 200 µL of the *Caulobacter* Tn5 library was inoculated into 20 mL of PYE for the initial outgrowth for 5 h, which was then inoculated into 2 L for the PYE and unfiltered lake water treatments, and 2 replicates of 2 L each for filtered lake water. Lake water cultures were harvested by filtration after 60 h of growth, and the PYE condition was filtered after 12 h to approximate the same number of doublings. However, our PYE cultures achieved over 6 doublings, versus 4 doublings for lake water.

A nested PCR approach was used to specifically amplify transposon-containing DNA fragments for sequencing. A low cycle PCR amplification for each sample was first performed using a standard reaction protocol for KOD Xtreme™ Hot Start Polymerase with 5% DMSO and 0.3 µM primer using the primers F1 and P7 (Curtis and Brun, 2014) (Table S1 in (Hentchel et al., 2019)). Cycling conditions were as follows: 95°C for 90 sec; 5 cycles of 95°C for 15 sec, 68°C for 30 sec, and 72°C for 30 sec; 13 cycles of 95°C 15 sec, 55°C 30 sec and 72°C 30 sec, followed by a final extension at 72°C for 5 min. Samples were treated with ExoSAP-IT™ PCR product cleanup reagent (Thermo Fisher Scientific) according to manufacturer's protocol. A second PCR step was performed with the transposon specific primer containing the adapter sequence using KOD Xtreme™ Hot Start Polymerase with 5% DMSO and 0.3 µM primer in a 62.5-µL reaction volume using the primers Tn5-left and P7 (Curtis and Brun, 2014) (Table S1 in (Hentchel et al., 2019)). Cycling conditions were as follows: 95°C for 3 min, 12 cycles of 95°C for 30 sec, 55°C for 30 sec, and 72°C for 30 sec, followed by a final extension at 72°C for 5 min. Product size

(~200 bp) was confirmed on a 1% agarose gel. After standard bead cleanup and Illumina library preparation, samples were sequenced at the University of Chicago Genomics Facility using a custom sequencing primer (Curtis and Brun, 2014) (Table S1 in (Hentchel et al., 2019)).

Fitness analysis was performed as previously described (DeJesus et al., 2015) using the TRANSIT software (available at https://github.com/mad-lab/transit). We used the permutation test in TRANSIT to quantify differences in sequencing read counts between our PYE and lake water conditions (DeJesus et al., 2015). The complete Tn5 dataset (Tables S4 & S5 in (Hentchel et al., 2019)), genes with differential fitness ($p<0.01$; Tables S5 & S6 in (Hentchel et al., 2019)), and genes shared between the Tn5 and Tn-Himar datasets (Table S7 in (Hentchel et al., 2019)) are listed in Supplemental Information. Raw Tn-seq data are deposited in the NCBI sequence read archive under BioProject accession PRJNA429486; BioSample accession SAMN08348191; SRA accession SRP128742.

Results

Growth of *Caulobacter* in natural freshwater

As a prerequisite to measuring strain fitness, we first sought to demonstrate *Caulobacter* growth in natural freshwater. We collected nearshore water from Lake Michigan, representing a typical oligotrophic freshwater system inhabited by *Caulobacter* spp. (McMillan and Stout, 1977;Lee et al., 2015). With no additional supplementation, filtered (0.1 μm) lake water supported *Caulobacter* growth to a maximal density of

approximately 5×10^5 CFU/mL (Fig A3.1A), from an initial inoculum of 2.5×10^4 CFU/mL. *Caulobacter* doubled 4–5 times at a rate of 0.14 hr^{-1} (doubling time 5 hr). Similar growth rates were observed in unfiltered lake water. Supplementation with 0.1% xylose increased the maximal density by about 10-fold, while addition of 23 μM K_2HPO_4 had no effect (Fig A3.1A), implying that carbon, but not phosphorus, limits *Caulobacter* growth in Lake Michigan water. For comparison, we also assayed *Caulobacter* growth in water collected from Lake Superior and found a similar growth yield (Fig A3.1B). Supplementation with either 23 μM K_2HPO_4 or 0.1% xylose did not significantly enhance *Caulobacter* growth, but together 0.1% xylose and 23 μM K_2HPO_4 enhanced growth by more than 10-fold, suggesting that both carbon and phosphorus limit growth in Lake Superior. By comparison, *Caulobacter* reached a density of 3×10^9 CFU/mL in PYE broth or in defined M2X medium (Fig A3.1C). Notably, cell density was stable for one week in lake water but declined by 2-3 orders of magnitude after 2 days of cultivation in artificial media (Fig A3.1). This finding is consistent with a report by Poindexter describing *Caulobacter* isolates that tolerated prolonged nutrient scarcity with little loss of viability (Poindexter, 2006). Based on our results, we chose to perform our genetic analysis in unsupplemented water from Lake Michigan.

A global Tn-sequencing approach identifies *Caulobacter* mutants with altered fitness in lake water

We sought to identify genes required for *Caulobacter* growth in natural freshwater, compared to defined M2X or complex PYE medium. To this end, we constructed mutant

libraries (Table A3.2) using two different transposons: Tn5, which inserts randomly, and Tn-Himar, which inserts specifically at TA dinucleotides, which occur on average every 82 bp in the *Caulobacter* genome. Each transposon in the pool of Himar transposons contains a unique 20-bp barcode sequence, which is mapped once to a specific insertion site in the genome and thereafter can be quantified by simple amplicon sequencing (Wetmore et al., 2015), see Fig. S1 in (Hentchel et al., 2019). Both transposon libraries were constructed by growing cells in PYE; hence insertions in genes essential for growth in PYE are not represented in either library.

We cultivated the *Caulobacter* Tn5 pool in PYE and filtered lake water (0.1 μm). Although Tn5 is capable of insertion at almost every position in the genome, our Tn5 library had lower site saturation than our Tn-Himar library, which limited the statistical power to identify significant fitness effects (DeJesus et al., 2015;Chao et al., 2016). We calculated mutant fitness and gene essentiality for all genes (Tables S5 & S6 in (Hentchel et al., 2019)**)** and identified 55 genes for which Tn5 disruption significantly diminished or enhanced growth in lake water relative to PYE (adjusted p-value cutoff < 0.01). Given the limited power of the Tn5 dataset, we focused our analyses on the Tn-Himar dataset, but include the Tn5 data in the supplemental material as they provide useful validation of the Tn-Himar data discussed hereforward.

The *Caulobacter* Tn-Himar library contained an estimated 2×10^6 clones, of which 7×10^4 passed the criteria for barcode mapping (Wetmore et al., 2015). Considering there are only ~ 5×10^4 TA insertion sites in the *Caulobacter* genome, it is clear that in this population we hit some sites more than once with unique barcodes. We cultivated this

library in four conditions: 1) complex PYE medium, 2) defined M2X medium, 3) filtered lake water, and 4) unfiltered lake water (Fig S2 in (Hentchel et al., 2019)). To ensure that we started the experiment with sufficient mutant strain diversity, we inoculated the same total number of cells (2.5×10^7) in each treatment and aimed for 4–5 doublings into the late exponential phase of growth (Fig A3.1 and Fig S2 in (Hentchel et al., 2019)). For PYE and M2X treatments, cells were grown in 1.5 mL volumes for 10 and 20 h, respectively. For lake water treatments cells were grown in three flasks each containing 2.5 L for 64 hours. By varying culture volume, we ensured an equal number of cell divisions across a similar phase of the growth curve. This approach required cultivation at different cell densities between conditions. After harvest, barcodes were analyzed as described (Wetmore et al., 2015), and strain fitness was calculated as the $\log_2$ of the ratio of barcode abundance in lake water (or M2X) to the control condition (PYE) (Wetmore et al., 2015). Given the 20–30-fold increase in cell number of the mutant pool, a Tn-Himar insertion strain that did not grow at all should have a fitness score around -4 to -5; more extreme (lower) fitness scores indicate strains that did not survive cultivation. The most extreme negative fitness scores in this dataset (i.e. < -4) likely reflect genes that are essential in a particular condition (Fig A3.2A and Fig A3.2D; Table S2 in (Hentchel et al., 2019)). The distributions of fitness scores in defined medium and the two lake water conditions are presented in Fig A3.2B and Fig A3.2C.

To validate our approach, we examined the fitness consequences of disrupting xylose utilization genes in the M2-xylose (M2X) growth condition. Genes in the *xylXABCD* operon are required for xylose utilization (Stephens et al., 2007a;Stephens et al., 2007b).

As expected, insertions in these genes generated fitness scores of -3.6 to -6.6 when the pool was cultivated in M2X (Fig A3.3A, Table S2 in (Hentchel et al., 2019)). Disruption of *xylR*, which functions as a transcriptional repressor of the xylose operon (Stephens et al., 2007b), resulted in a positive fitness score in M2X relative to PYE, indicating that derepression of the xylose utilization genes is advantageous when xylose is the sole carbon source. Disruption of the *xylXABCD* genes had little effect on fitness in lake water, which contains a range of carbon sources beyond xylose; disruption of *xylR* resulted in a modest fitness decrease in lake water relative to PYE (Fig A3.3A), suggesting a cost to constitutive expression of unused genes.

Increased variability of fitness scores in lake water

Compared to defined M2X medium, lake water is more heterogeneous over time and space. Our four lake water experiments used water collected on four days over a 2-week period and showed greater variability in strain fitness scores than our four independent M2X replicates (Fig A3.2A, Fig A4.3B and Fig A3.4). In addition, we fit a linear model to test the effects of two factors, sampling day and filtration condition, on strain fitness scores, and found a number of genes that differed significantly across days, including genes related to cell surface carbohydrate biosynthesis and environmental sensing and gene regulation (Table S8 in (Hentchel et al., 2019)). This variability likely reflects day-to-day differences in temperature, mixing, and biotic factors such as phage dynamics, though we cannot completely rule out technical day-to-day variations in sample processing. Future work with additional temporal replicates could discriminate genes

whose functions are consistently important from genes that are exploited under transient conditions in the lake.

Surprisingly, filtration (0.1μm) had little effect on the global fitness profile of *Caulobacter* (Fig A3.4), and our linear model approach did not identify any genes with differential fitness between filtered and unfiltered lake water. This result implies that particulates and other microorganisms present in the lake water did not affect strain growth, suggesting that *Caulobacter* is not in strong competition with other microbes for "common goods" in this system on the time scale of our experiment.

Fitness defects are more extreme in defined medium than in lake water

Transposon disruption of genes required for amino acid biosynthesis, nucleotide biosynthesis, lipopolysaccharide biosynthesis, and nucleotide sugar biosynthesis resulted in extreme (fitness score < -4) growth defects in M2X (Fig A3.3, Tables S2-S3 in (Hentchel et al., 2019)); these fitness scores provide evidence that strains harboring disruptions of these genes did not grow at all in M2X and thus likely comprise a strain/gene set that is essential in this defined condition. This result is not surprising, considering that growth in M2X medium requires de novo biosynthesis of diverse monomers and intermediates, many of which are supplied exogenously in the reference PYE condition. In many cases, strains with severe fitness defects in M2X also had reduced growth in lake water, but the fitness costs were less severe (Tables S2 and S9-S10 in (Hentchel et al., 2019)). We controlled the number of doublings (approximately 4–5) across all conditions, so the more pronounced fitness costs in defined medium

compared to lake water cannot be explained by differences in the number of doublings. Instead, these results imply that lake water is more similar to the reference condition PYE than M2X is to PYE, in terms of the metabolic demands it imposes on cells. This inference is supported by principal component analysis across all growth conditions (Fig A3.3B). Indeed, we expect that natural freshwater supplies diverse metabolites and growth substrates that may render some genes dispensable, whereas defined media provides fewer exogenous resources.

Pathways conferring differential fitness in natural freshwater and artificial media

To further explore the selective pressures faced by *Caulobacter* across these conditions, we focused on genes whose disruption induced large fitness effects, namely fitness scores less than -1.2 and greater than +1.2 (this approximates a $\pm 3\sigma$ cutoff). Based on this criterion, we identified 83 and 82 genes in the filtered and unfiltered lake water conditions, respectively, and 213 genes in the defined M2X medium (Table S9 in (Hentchel et al., 2019)). Genes with significant fitness values across all three conditions based on the t-statistic of Wetmore and colleagues (Wetmore et al., 2015) are outlined in Table S10 in (Hentchel et al., 2019). Broad functional patterns in our Tn-Himar dataset were assessed using clusters of orthologous group (COG) annotations (Galperin et al., 2015) (Fig A3.5). A full comparison of genes for which Tn-Himar disruption results in a specific advantage or disadvantage in M2X defined medium, but not in filtered or unfiltered Lake Michigan water (relative to complex PYE medium), and vice versa, are presented in Tables S11-S12 in (Hentchel et al., 2019). Genes that were not hit by Tn-

Himar, and thus not included in any of our analyses are included in Table S13 in (Hentchel et al., 2019). Many of these genes have been previously defined as essential (Christen et al., 2011). A clustered heatmap that contains genes with fitness scores less than -1.2 and greater than +1.2 from either the filtered or unfiltered lake water conditions is presented in Fig. S6 in (Hentchel et al., 2019).

Not surprisingly, the most negative fitness scores were observed for genes in amino acid and nucleotide biosynthesis (discussed above), and in genes required for transport of metabolic substrates into the cell (Tables S2 and S9 in (Hentchel et al., 2019)); we observed similar defects in our Tn5 dataset (Tables S4-S6 in (Hentchel et al., 2019)). In addition, disruption of genes encoding catabolic enzymes in the glycine cleavage pathway and in branched amino acid degradation led to an apparent enhancement of fitness in both M2X medium and in lake water relative to PYE, in both the Tn-Himar and Tn5 experiments (Fig A3.4; Tables S2 and S4-S7 in (Hentchel et al., 2019)). This result likely reflects the higher cost of deleting these catabolic genes in the reference PYE condition compared to M2X or lake water, and is consistent with transcriptional data showing that select amino acid degradation pathways — including glycine cleavage, histidine, branched chain, and phenylalanine degradation — are upregulated in PYE compared with M2X (Hottes et al., 2004).

Surprisingly, we found enhanced fitness for strains with disruptions in motility genes in lake water relative to PYE (Fig A3.4 and Fig A3.5). We more carefully examined the fitness scores of genes involved in synthesis and assembly of the flagellum (Fig A3.6). The flagellum is assembled in a regulated hierarchy of stages, which is well described in

Caulobacter (Xu et al., 1989;Benson et al., 1994;Ardissone and Viollier, 2015). Class II genes encode the inner components of the flagellum, including the export apparatus, and regulatory proteins that activate expression of class III and IV genes. Class III genes encode the basal body and hook structures. Completion of class III structures activates translation of class IV genes, which encode the subunits of the flagellar filament. Thus, defects in each class prevent expression of subsequent classes. Within each class of flagellar genes, we observed consistent fitness patterns, demonstrating the power of this method to capture even modest effects of gene disruption. Disruption of class II flagellar genes conferred an advantage that was significantly greater in lake water than in M2X compared to PYE (Fig A3.6B and Fig S2 in (Hentchel et al., 2019)). Disruption of class III genes followed similar trends, but with smaller magnitude effects. *Caulobacter* encodes six redundant class IV flagellin genes (Faulds-Pain et al., 2011), three of which are represented in our Tn-Himar pool and whose disruption, not surprisingly, had no effect on fitness. Disruption of the motor stator gene *motA* or *motB*, which results in a fully assembled but paralyzed flagellum (Johnson and Ely, 1979;Ely et al., 1984), did not affect fitness under our cultivation conditions. Together, these results suggest that the fitness advantage of flagellar gene disruption is not derived from energy saved in powering the flagellum, but rather in energy or resources saved in synthesizing and assembling the flagellum. In the lake water cultivations, we observed appreciable day-to-day variation in the fitness of each class of flagellar gene mutants (Fig S2 in (Hentchel et al., 2019)), which was particularly pronounced for class III genes. Patterns in this day-to-day

variability were consistent across members of each class, suggesting that this variability is driven by environmental factors rather than technical factors.

Fitness was also affected by the ability to synthesize the polar extracellular adhesin known as the holdfast (Fig 3.6A and Fig S2 in (Hentchel et al., 2019)). We systematically analyzed genes involved in synthesis, secretion, and attachment of the holdfast. Most holdfast genes yield partial to complete defects in holdfast development when disrupted (Toh et al., 2008); we categorized these genes as 'unique functions' genes. However, two sets of holdfast biosynthesis genes have redundant functions: two Wzy-family polymerase genes function in holdfast development and three paralogs of the HfsE glycosyltransferase have genetically redundant activities in holdfast synthesis (Toh et al., 2008). Disruption of genes in these redundant sets had no effect on fitness. Disruption of genes in the unique function group resulted in a modest but consistent fitness advantage in M2X and a fitness disadvantage in both filtered and unfiltered lake water, relative to PYE (Fig A3.6A and Fig S2A in (Hentchel et al., 2019)). For the group of all unique genes, the fitness consequence for loss of holdfast was significantly different between growth conditions ($p < 0.0001$) (Fig A3.6A).

Genes of unknown function contribute to fitness in natural freshwater

We hypothesized that many genes of unknown function play important roles in natural environmental contexts but not in typical laboratory media. Of all genes showing large fitness effects (± 3σ), hypothetical genes or genes of unknown function accounted for 16% (13/83) in filtered lake water, 15% (12/82) in unfiltered lake water, and 7% in

defined medium (15/213) (Table A3.3). Across these three conditions, five hypothetical genes were shared. *CCNA_03860* was the only hypothetical gene for which disruption provided a fitness benefit across all three conditions relative to PYE. CCNA_03860 contains a conserved domain belonging to the YkuD superfamily, which has been shown to have L,D-transpeptidase catalytic activity, providing an alternate pathway for peptidoglycan cross-linking (Bielnicki et al., 2006;Kumar et al., 2017). Disruption of *CCNA_01724*, *CCNA_03864*, *CCNA_03909*, and *CCNA_00375* resulted in reduced fitness across all three conditions relative to PYE. Hence using natural growth conditions may be critical for understanding the functions of many uncharacterized bacterial genes.

Discussion

Tn-seq fitness scores provide a window into cell-environment interactions

Bacterial genomes carry relatively little noncoding DNA. Genes that confer no fitness benefit tend to decay over time (Mira et al., 2001) implying that genes that are maintained are beneficial at least under some circumstances. Yet traditional microbial cultivation approaches often fail to yield discernable mutant phenotypes for many genes. One approach to overcome this challenge is to interrogate gene function in more relevant ecosystem contexts, embracing physicochemical complexity. The genome-scale fitness analysis of *Caulobacter* transposon mutants reported in this study provides new understanding of genes that affect growth in a *bona fide* freshwater environment. Disruption of genes involved in biosynthesis of non-aromatic amino acids, lipopolysaccharides, and nucleotide sugars results in large fitness defects in natural

freshwater compared to complex laboratory medium (PYE). Moreover, fitness effects were variable across temporal lake water replicates; this variability likely reflects physicochemical and biological variability in the lake and suggests an important role for transient response genes in fluctuating environments.

The fitness costs and benefits of motility and attachment in freshwater

The energetic cost of flagellar biosynthesis and motility is well established (Smith and Chapman, 2010;Martinez-Garcia et al., 2014). Our data indicate that transposon disruption of genes required for the synthesis of the single polar flagellum of *Caulobacter* enhanced fitness in lake water relative to PYE medium (Fig A3.6B). This is consistent with a *Salmonella* Tn-Seq study that revealed a fitness advantage in strains with disrupted flagellar genes (Langridge et al., 2009). Notably, we found that fitness effects were not uniform across all flagellar genes: disruption of class II genes, which has the greatest impact on flagellar gene expression, also led to greater effects on fitness, compared to class III and IV genes. The fitness enhancement in lake water is not due to the energy savings from motor rotation, as strains with insertions in the *motA* and *motB* stator genes, which assemble a full but non-rotary flagellum (Johnson and Ely, 1979), showed no fitness difference (Fig A3.6B). We conclude that the relative fitness advantage of flagellar gene disruption is related to the cost of biosynthesis of flagellar proteins. It seems certain that over longer cultivation timescales, and in more spatially complex environments, the *Caulobacter* flagellum provides a fitness advantage, as flagellar genes are maintained in natural freshwater environments.

Our data reveal that disruption of genes required for holdfast biosynthesis is disadvantageous when strains are cultivated in lake water relative to PYE. This fitness cost was evident in both filtered (particle-free) and unfiltered lake water relative to PYE (Fig A3.6A), suggesting that the effect is not due to adhesion to particles in the medium. Instead, it is possible that the holdfast confers a growth advantage by enabling adherence to the flask surface, where polymeric nutrients concentrate to form conditioning films (Loeb and Neihof, 1975;Schneider and Leis, 2003). In defined M2X medium, disruption of holdfast biosynthesis genes confers a fitness advantage (Fig A3.6A). In this medium, all the components are salts or simple sugars, which do not efficiently condition naïve surfaces (Loeb and Neihof, 1975;Schneider and Leis, 2003). In this case, surface attachment is apparently not advantageous, and holdfast biosynthesis comes at a cost.

Genetic evidence suggests a complex medium is a better freshwater analog than a defined mineral medium

Fitness defects of *Caulobacter* mutants were often more severe in a defined mineral xylose medium (M2X) than in lake water, relative to PYE. Moreover, the overall fitness profile of *Caulobacter* mutants cultivated in lake water more closely resembles that in PYE than in M2X, suggesting that dilute complex medium is a better proxy for natural freshwater. *Caulobacter* belongs to a group of dimorphic prosthecate (i.e. stalked) alphaproteobacteria that are often specialized for oligotrophic, dilute environments (Poindexter, 1964;2006). Indeed, the inhibition of growth and stalk development due to excess nutrients was the first physiological property of *Caulobacter* spp. to be described

(Loeffler, 1890). Complex and defined media of varying compositions have been outlined for cultivation of *Caulobacter* and related genera, but it is notable that dilute peptone (less than 0.2% w/v) generally supports growth of all dimorphic prosthecate bacteria (Poindexter, 2006). This observation supports the notion that the natural nutrient environment of this class of bacteria is best captured by cultivation in a dilute complex medium that contains amino acids and other trace complex biomolecular components. Our data also demonstrate that an M2-based medium exerts highly specific metabolic constraints and is likely not an ecologically or physiologically relevant growth condition.

An approach to study gene function in ecosystem context

The explosion of bacterial genome sequence information has far outpaced our ability to characterize gene function using traditional approaches, leading to the accumulation of thousands of 'unknown' protein families. Many of these families are conserved throughout the bacterial domain, which is evidence that they confer a selective benefit in particular conditions. This leads to the following question: under what circumstances do these conserved families provide a fitness advantage? At the onset of this study, we hypothesized that many of these unknown protein families would prove to be important in the natural ecological context of a bacterium. Among the genes whose disruption leads to the greatest fitness effects (± 3σ) in filtered lake water relative to PYE, approximately 15% are hypothetical or conserved genes of unknown function (Table A3.3, Tables S8–S9 in (Hentchel et al., 2019)). The approach we describe here indicates that these genes of unknown function play an important role in *Caulobacter* physiology in

a natural freshwater environment. Going forward, one can take advantage of lake-specific growth phenotypes to begin to define the functions of these genes in an ecologically relevant context.