# IDENTIFYING THE FUNCTIONS OF UNCHARACTERIZED GENES IN *BACILLUS SUBTILIS*

A Dissertation

by

SARAH HARTMAN

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

| | |
|---|---|
| Chair of Committee, | Jennifer K. Herman |
| Committee Members, | Ryland F. Young |
| | Hays S. Rye |
| | Joseph A. Sorg |
| Head of Department, | Dorothy E. Shippen |

August 2019

Major Subject: Biochemistry

# ABSTRACT

Bacteria encode a variety of ways to sense and respond to their dynamic environments. Consequently, in order to implement the appropriate response to execute changes in metabolism or subcellular organization, gene expression is tightly coordinated with the cell's ability to recognize a perturbation. For instance, the devastating effects of a toxic compound can be mitigated by expressing gene products to export or metabolize the compound. The Gram-positive model organism *Bacillus subtilis* utilizes a variety of gene expression programs in order to survive the diverse environments it encounters. Intriguingly however, nearly 40% of the genes in *B. subtilis* are either unannotated or annotated without experimental validation. Genes that play important roles in regulating gene expression, metabolism and regulating essential cell processes are likely found within this set. The lack of gene characterization is due in part to our inability to obtain tractable phenotypes from which to form testable hypotheses regarding gene function. To associate gene products with phenotypes, we utilized a gene misexpression library comprising more than 800 strains as a tool to uncover gene products that perturb growth and/or subcellular organization when artificially expressed. From this set, 10 DNA-binding proteins predicted to be transcription factors were selected for transcriptomic analysis and uncovered candidate regulatory targets for 8/10 of the DNA-binding proteins. Artificial expression of one of the transcriptional regulators, YxaD, resulted in cells unable to properly segregate chromosomes or replicate DNA, and confirmed a prior observation that YxaD is capable of interaction with ScpA, a subunit of the Structural Maintenance of Chromosome (SMC) complex.

Additionally, we showed that YxaD acts as a repressor of its own promoter as well as a divergently transcribed operon encoding the *cid*/*lrg* homologs, *yxaKC*. Transcriptional profiling revealed that *yxaKC* is expressed both in a specific region of a biofilm, and in glucose-containing medium, with expression peaking during stationary phase growth just as glucose is depleted. Using untargeted NMR-based metabolomics, we showed that cells artificially expressing YxaKC show enhanced export of the overflow metabolite 2-acetolactate. We hypothesize that YxaKC helps to maintain cellular homeostasis possibly by acting as a passive transporter of 2-acetolactate.

# DEDICATION

I would like to dedicate this dissertation to my mother, Liz Gray, and my father, Jim Hartman, who have always shown me an incredible amount of love, care, and support.

# ACKNOWLEDGEMENTS

I would like to thank my committee, Dr. Herman, Dr. Young, Dr. Rye and Dr. Sorg for providing their expertise, guidance, and constructive feedback on this project.

I would like to give a special thanks to my advisor, Jen Herman. Jen has always believed in me and my ability to complete this project. From my research experience, I have gained an incredible amount of independence and confidence in my abilities to reach far out of my comfort zone. Jen has also been tremendously supportive in my future aspirations, allowing me to attend conferences our lab does not normally attend and by offering various teaching opportunities, and I am thankful for that.

A big thank you goes out to the Herman lab's past and current members who have all taken time to show me some of the tools and techniques I've needed in the lab. I would also like to thank Cruz Perez and Taylor Brundage, an undergraduate and REU student I had the opportunity to mentor, for their help in the lab.

I would also like to thank the Center for Phage Technology (CPT) and Supergroup. Having the opportunity to present my research once a semester really helped me grow as a presenter and provided valuable insight into my project.

I would like to thank Nowlan Savage and Zane Taylor for teaching me how to use various NMR materials and software.

Thanks also to my friends I have met during my time in College Station and the great experiences I have had with them.

A very special shout out goes to one of my best friends, Allyssa Miller. Over the course of my ~6 years, Allyssa has taught me most of what I know in the lab. She very

quickly became one of my closest friends, and eventually became my bench-mate, travel partner, and neighbor and I am so incredibly thankful for her.

I would also like to thank my fiancé Drew and our dog Riley who have been incredibly supportive and encouraging. I am so excited to experience all our future endeavors.

I would like to thank my family. A big thank you goes to my sisters, Carley and Abby, for their love and laughter.

# CONTRIBUTORS AND FUNDING SOURCES

## Contributors

This work was supervised by a dissertation committee consisting of Professor Jennifer K. Herman (advisor), Professor Ryland F. Young, and Professor Hays S. Rye of the Department of Biochemistry and Biophysics and Professor Joseph A. Sorg of the Department of Biology at Texas A&M University.

The majority of construction and screening of the misexpression library in Chapter 2 (BEIGEL) was done by Allyssa K. Miller, Ben T. Mercado, and Faxin Zheng. Suppressor selections in Chapter 2 were performed by Yi Duan.

The instrumentation for the NMR experiments in Chapter 4 was performed by Dr. Xianzhong Xu and Dr. Jae-Hyun Cho in the Biomolecular NMR Laboratory at Texas A&M University.

All other work conducted for the dissertation was completed by the student independently.

## Funding Sources

Work in this dissertation was funded by that start-up funding of Dr. Jennifer Herman from the Department of Biochemistry and Biophysics with the funding number 02-248312. Additionally, the work in this dissertation was funded by Award 1514629 from the National Science Foundation. The authors are responsible for the contents in this dissertation. Information in this dissertation does not necessarily represent the official views of the Department of Biochemistry and Biophysics.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

# 1. INTRODUCTION

Advancements in next-generation sequencing (NGS) have led to increased sequence data output, with reduced costs, making projects involving large-scale DNA sequencing more feasible. Personalized medicine, in which healthcare is tailored to an individual based on genetics and environment, among other factors, was once a futuristic dream; however the cost to sequence a human genome has dropped to approximately $1,000 (https://www.genome.gov/sequencingcostsdata/), bringing personalized treatments closer to reality.

Currently, there are nearly 8,000 eukaryotic and 190,000 bacterial genomes in GenBank, with the number of sequenced genomes increasing exponentially. The International Nucleotide Sequence Database Collaboration (INSDC), a collaboration amongst the DNA Data Bank of Japan at the National Institute for Genetics in Mishima, Japan (DDBJ), the European Nucleotide Archive (ENA) at EMBL-EBI, and GenBank at NCBI, reported a near doubling in assembled/annotated sequenced bases from 2015 to 2017 alone; from 1.432 trillion to 2.650 trillion bases (Karsch-Mizrachi *et al.*, 2018). It is remarkable then, given the fundamental importance of understanding the function of gene products to understanding system-level "ome" data (genomes, transcriptomes, proteomes, metabolomes, etc.) and the large amount of sequencing data available, that approximately 30-50% of annotated genomes lack functional annotations (Hanson *et al.*, 2009). A recent search for “hypothetical proteins” in GenBank listed over 140 million gene products lacking both experimental validation and predictions based on homology. Surprisingly, nearly 30% of uncharacterized gene products have functions predicted to

be enzymatic (Ellens *et al.*, 2017). Furthermore, we have yet to attribute more than 800 out of 6,322 known enzymatic reactions (>12.5%) to gene products (Shearer *et al.*, 2014). This highlights that significant gaps in our understanding of metabolism still exist impacting scientific goals and future endeavors.

### 1.1. Implications for gene annotation and discovery

The lack of gene characterization presents major challenges to genome editing, metabolic engineering, and our understanding of the minimal components required for life. How can we artificially express genes in biological systems without functional annotations or understanding the effects artificial gene expression may have? The potential to use genome editing for disease treatments and preventions has been largely impacted by recent discoveries. CRISPR-based (Clustered Regularly Interspaced Short Palindromic Repeats) genome editing has transformed biological research and has the potential to transform medicine. However, as with any manipulation of DNA, the possibilities of off-target effects in eukaryotic organisms are still prevalent (Komor *et al.*, 2017) and the development of CRISPR requires further investigation. The precision of genome editing is expected to improve with the discovery of novel DNA-binding and/or cleavage proteins to be used in CRISPR-based systems.

Gene discovery is also important for the field of synthetic biology. Synthetic biology is an area of study in which scientists engineer biological systems to carry out new functions or produce desired compounds like pharmaceuticals, chemicals, and food additives. In general, biological components with a desired function, such as specific enzymes, are combined in combinatorial fashion to create new systems. One limitation

of synthetic biology is that it relies heavily on the availability of known components which are limited by gene discovery. For example, 12% of enzymatic activities occurring in nature have yet to be associated with a gene product (Shearer *et al.*, 2014). The discovery of the genes responsible for these enzymatic functions (Ellens *et al.*, 2017) could have a major impact on synthetic biology.

Closely related to the field of synthetic biology is synthetic genomics, which involves the synthesis or manipulation of organisms to define a minimal genome (Konig *et al.*, 2013). Early work performed in *Escherichia coli* K-12 revealed how deleting 12 K-islands, reducing the genome by 8.1% but did not affect growth on minimal media (Kolisnychenko *et al.*, 2002). To create this simplified organism, we first need to understand the minimal components a cell needs for life. In this way, synthetic biology and synthetic genomics converge in that they face a similar limitation – they both rely on functional gene discovery to expose the repertoire of biological components encoded by various organisms. One of the major challenges in characterizing gene functions is that they can be redundant and/or only important only under specific, often undefined growth conditions. Current approaches in gene discovery attempt to address these challenges (discussed (Brochado & Typas, 2013)).

### 1.2. Sequence annotations

Why are so many genes left unannotated or misannotated? At the onset of the genomic era, as sequences were deposited into databases, similarities, especially in protein coding sequence were observed, leading to the initial hypothesis that similar sequences would possess similar functions. Therefore, sequences were often annotated

based on sequence similarity to an experimentally characterized gene product. In 1990, the creation of the program BLAST (Basic Local Alignment Search Tool) enabled scientists to search their sequences against a database of published sequences. The BLAST tool allowed scientists to hypothesize function based on sequence similarity (Altschul *et al.*, 1990). Although the development of bioinformatic tools, like BLAST, have aided many projects, they also have limitations that introduce new problems. Even at their best, automated annotations are estimated to have an accuracy of only 70% (Bork, 2000). This leads to misannotations, which can impede research by leading researchers to follow unproductive research trajectories. The detrimental effect of incorrect annotations is exemplified in the case of the plant plastid terminal oxidase (PTOX)(expressed in stoma cells and function in photosynthetic electron transport) and alternative oxidases (AOX)(expressed in the mitochondria and involved in respiratory electron transport families). At least seven proteins belonging to the PTOX family were misannotated to belong to the AOX family, despite the fact that each have distinct functions and localizations (Nobre *et al.*, 2016). Inaccurate annotations can severely impact or delay scientific progress by introducing biases and leading researchers the wrong conclusions.

When enzymes are characterized, the enzyme is assigned a four-digit Enzyme Commission (EC) number for each reaction performed (Green & Karp, 2005). However, when an enzyme is only partially characterized, and it is only known that an enzyme catalyzes one reaction in a family of reactions, the enzyme is assigned a partial EC number. Later, additional enzyme functions may be discovered, but in the meantime putative enzymes may be assigned the same partial EC value as the characterized

enzyme. This can result in misleading annotations (Green & Karp, 2005). These partial, incorrect, or misleading annotations can be difficult to amend because they themselves go into databases that are used to curate newly sequenced genes. In addition, even if gene annotations are correct, the functional predictions for the unknown are only as good as the reference. Thus, although bioinformatics approaches are powerful tools, ultimately experimental validation is still required for gene characterization.

Many of the inaccuracies associated with automated annotation can be attributed to context-specific parameters that are not taken into consideration (e.g. different species, pathways, metabolic processes, growth conditions, etc.). For example, most prediction tools rely on data related to core metabolic pathways and lack predictions for secondary metabolism. For instance, a recent study compared four annotation programs and found that most agreed on predictions involving core metabolic pathways while nearly none agreed on predictions involved in degrading fatty acids, aromatic compounds or secondary metabolites (Griesemer *et al.*, 2018). Another challenge is that some protein functions are inherently difficult to predict from sequence alone such as whether a protein with a predicted transmembrane domain is a transporter if so, what the substrate is, and the substrate of a particular enzyme (Griesemer *et al.*, 2018). Some of these challenges have been mitigated by recent developments in transporter prediction software, for instance, the TransportDB 2.0 database uses an automated annotation pipeline called TransAAP to help predict if membrane proteins are transporters (Elbourne *et al.*, 2017). Furthermore, alternative splicing and post-translational modifications can significantly alter gene product function, making the in vivo function difficult if not impossible to predict (Bork, 2000). Accurate gene annotation will require

not only functional predictions, but must be determined experimentally and therefore, more tools are needed to accurately annotate genomes and further.

### 1.3. Current tools for gene characterization

Historically, gene discovery and characterization in bacteria involved forward genetic approaches, where a mutagen (chemical agent, transposon, virus, etc.) was used to enhance mutation rate. Then, depending on the mutation, a subsequent phenotype could be screened or selected for, and the location of the gene responsible for the phenotype mapped and identified. More recently however, the availability of sequencing information, robotics, and bioinformatics has enabled scientists to take reverse genetic approaches that utilize ordered loss-of-function (LOF) and/or gain-of-function (GOF) libraries. Libraries facilitate gene discovery of genes by permitting high throughput screening for phenotypes that respond to a genetic or chemical perturbation.

A vast array of LOF libraries are available including full non-essential knock-out collections, saturated transposon libraries, antisense RNA silencing, and more recently, CRISPR-Cas knockdown libraries (Brochado & Typas, 2013). Arrayed libraries have been created for organisms across all domains of life and are ubiquitously useful to measure the overall fitness of an organism under a given growth condition. Examples of libraries include 1) the International Mouse Phenotyping Consortium, which contains 5,000 knockout strains encompassing 20,000 protein-coding genes made with either a reporter gene or more recently using CRISPR-Cas9 (Meehan *et al.*, 2017, Chakravorty & Hegde, 2018), 2) the yeast deletion library comprised of over 21,000 mutants strains of 6,000 different open reading frames (Giaever & Nislow, 2014) 3) the Keio collection

with 3,985 single gene deletions in *E. coli* (Baba *et al.*, 2006), and more recently 4) erythromycin and kanamycin resistant knock-out libraries consisting of 3,970 genes in the Gram-positive model organism *Bacillus subtilis* (Koo *et al.*, 2017). Because ordered libraries are large, experiments often require high-throughput equipment, such as robotics to handle the large number of samples. This puts limits on the viability of such approaches to research groups with access to the technology required.

Limitations for using gene knockouts are prevalent when studying genes that are essential under a desired growth condition. As an alternative, gene knockdowns are often used. For instance, a *Drosophila* RNAi knockdown library was used to identify genes contributing to congenital heart disease (Zhu *et al.*, 2017a), which would likely not appear in a deletion library. Loss-of-function tools like morpholinos and antisense oligos that block translation, have been used in zebrafish, frogs, chicks, mice and other organisms where genetic techniques are less developed, to elucidate mechanisms of embryo development (Heasman, 2002, Blum *et al.*, 2015). More recently, our improved understanding of CRISPR has made the design and execution of targeted gene knockdowns more feasible for individual labs and knockdown libraries have been created in organisms such as mice (Chakravorty & Hegde, 2018), *Drosophila* (Bassett *et al.*, 2014), *B. subtilis* (Peters et al., 2016), and even all 20,500 protein-coding genes in human cell lines (Morgens *et al.*, 2017). Because many knockdown techniques involve plasmids, multiple gene products can be targeted at once with compatible vectors drastically enhancing the range of these applications.

### 1.3.1. Limitations of current approaches

Unfortunately, robotics and other expensive equipment required for large-scale screens may not be accessible, and as a result, alternative approaches have to be sought out. One approach is to pool libraries, either in their entirety or in groups or sub-collections. Pooled libraries are screened with reporters or by having the population of mutants to compete with one another and subsequently measuring the relative abundance of each mutant, generally by sequencing (Brochado & Typas, 2013). Many deletion libraries have been created with a unique small DNA sequence corresponding to each mutant, called a barcode, which can simplifies the process of identifying mutants (Hensel *et al.*, 1995, Giaever *et al.*, 2002, Koo *et al.*, 2017).

Essential genes are often described as genes that are required in a common lab medium, for instance, for bacteria this common growth medium is Lysogeny Broth (LB) (Koo *et al.*, 2017). However, LB medium is artificial and quite different from the environments bacteria evolved and naturally inhabit. Because of this, many of the genes that are still uncharacterized may not be necessary for growth in LB but may be required under another growth regime (i.e. higher or lower temperature or pH, different carbon sources, salt or ethanol stress, presence of toxins, etc.). Genetic knockouts and knockdowns reach additional limitations due to functional redundancies and homeostatic control mechanisms. For example, 10 of the 12 penicillin-binding proteins encoded by *E. coli* can be deleted before any discernible morphological defect is observed (Denome *et al.*, 1999). To overcome some of the challenges of screening for essential genes, a technique involving a transposon screen followed by deep sequencing (Tn-seq) was developed to identify synthetically lethal mutants (van Opijnen *et al.*, 2009, Meeske *et*

*al.*, 2015, Meeske *et al.*, 2016). For instance, Meeske et. al used Tn-seq to identify a novel lipid II flippase in *B. subtilis*. Meeske et al. found that although MurJ alone was sufficient for lipid II transport in *E. coli* (Sham et al., 2014), deleting all 10 MurJ-like proteins in *B. subtilis* resulted in no discernible effect, indicating that an additional factor(s) was able to perform the essential function (Meeske *et al.*, 2015). To identify the redundant factor, the four MurJ homologs (Δ4) most similar to *E. coli* MurJ were deleted in *B. subtilis*. The Δ4 and wild-type strains were subject to transposon mutagenesis; the cultures were then pooled, allowed to grow, followed by sequencing of both cultures. Transposon insertions that resulted in synthetic lethality were underrepresented in the Δ4 strain compared to wildtype. Tn-seq revealed a gene *amj,* which was later found to support lipid II flippase activity in *E. coli* (Sham et al., 2014). The identification and eventual characterization of Amj by Tn-seq is further support to use bioinformatics is a tool, but proper validation of gene function requires experimental evidence.

### 1.3.2. Gain-of-Function

The utilization of LOF libraries and ability to screen thousands of mutants for a desired phenotype has helped discover gene function as well as develop extensive genetic interaction maps (Babu *et al.*, 2011, Brochado & Typas, 2013). For instance, multiple pathways involved in cell envelope biogenesis were identified in *E. coli* by screening for compensatory relationships between mutant combinations of putative integral membrane proteins (Babu *et al.*, 2011). Assigning genes to pathways can provide mechanistic insight to a protein's function but stop short at functional annotation. A major limitation to using LOF libraries is the existence of redundancies to

where knockouts have no detectable phenotype. It can be nearly impossible to make and screen all possible combinations of mutations required for the phenotype being screened. Considering the example above, screening a library of two mutant combinations would not have identified 10 of the 12 PBP in *E. coli*, at least under the growth condition tested, as there would likely be no discernible phenotype (Denome *et al.*, 1999). In addition, many screens involving LOF libraries or knockouts measure a change in fitness corresponding to a mutation, but this often provides little insight into the genes possible function, and no details on mechanism. GOF approaches offer a complementary approach to loss-of-function studies as a way to gain insights into gene function.

GOF approaches involve the overexpression or misexpression of genes either on a plasmid or directly from the chromosome. We use the term misexpression as opposed to overexpression when a gene product is produced outside of its native context and we lack information about its native and artificially induced expression levels. Similar to gene knockouts, cell perturbation by misexpression creates a flux that can result in a tractable phenotype, which can then be used to provide insights into protein function. Artificial gene expression was used to identify a protein that regulates FtsZ in *B. subtilis*, called RefZ (Wagner-Herman *et al.*, 2012). Briefly, misexpression of RefZ led to cell filamentation, a phenotype often indicative of a defect in cell division. In support of a defect in cell division, it was found that RefZ misexpression results in the disruption of FtsZ (a cell division protein) and this phenotype was rescued by *ftsZ* suppressor mutations. Interestingly, a Δ*refZ* mutant does not display a growth or morphological phenotype on its own when grown in LB medium, but it was eventually determined to be important for positioning the location of the division septum relative to the chromosome

during sporulation (Miller *et al.*, 2016). Therefore, by using misexpression to obtain a phenotype, a testable hypothesis and eventual function could be assigned to an unknown gene. In addition GOF libraries have been constructed using "ORFeomes", libraries that contain all the open reading frames in an organism, and have been created in organisms such as yeast (Liu *et al.*, 1992, Gelperin *et al.*, 2005), *Pseudomonas aeruginosa* PA01 (Labaer *et al.*, 2004), *E. coli* (Kitagawa *et al.*, 2005), and *Neisseria gonorrhoeae* (Brettin *et al.*, 2005). Compatible LOF and GOF libraries can be combined creating 'synthetic toolboxes' enhancing the potential of gene characterization studies. Furthermore, many GOF libraries have been created on 'entry' vectors, which enable the gene(s) of interest to be easily shuttled into other vectors to examine protein expression, protein localization, and protein-protein interactions (Labaer *et al.*, 2004, Brettin *et al.*, 2005, Gelperin *et al.*, 2005, Kitagawa *et al.*, 2005, Brandner *et al.*, 2008). These libraries can also be combined with systems level "omics" approaches such as transcriptomics, metabolomics, transcriptomics, and proteomics to develop working hypotheses. Ultimately, functional gene discovery requires a multiprong strategy.

### 1.4. Application of "omics" Tools for Gene Discovery and Characterization

Gene or protein sequence alone does not provide mechanistic information as to how the gene product is functioning to contribute to the observed phenotype, though with the handle of an associated activity or phenotype, it is generally possible to drill down to the level of molecular mechanism using classical hypothesis-driven studies. Uncharacterized genes that lack associated activities or phenotypes pose a significant challenge to performing hypothesis-driven studies. One way to approach this, in

addition to GOF studies, is to obtain data from global 'omics' methodologies. Current "omics" methodologies include transcriptomics, metabolomics, proteomics, lipidomics, and glycomics, among others. Combining data sets in combinatorial fashion (functional genomics) can provide details about global changes that occur in an organism by changing conditions, treating with a chemical, or as a result of changes in genotype. Below I will review the omics most relevant to my dissertation.

### 1.4.1. Transcriptomics

The genetic material of an organism is considered quite static, with the exception of mutations that can occur as a result of environmental conditions or DNA replication. In contrast, expression of genes is primarily controlled at the level of transcription, with many genes only bring expressed when they are required. The transcriptome, or total RNA expressed in a population or single cell, is a function of genotype, cell type, environment, epigenetics, and stochasticity. Since the activity of gene products is associated with their expression, transcriptomics is extremely valuable for discovering genes that are expressed and function under a condition of interest. In addition, the precision of transcriptomics permits collection of temporal data, allowing deduction of what genes are expressed in direct response to a particular signal, as well as what genes are induced as a secondary consequence.

The field of transcriptomics has made significant advancements in the last few decades. The original method for measuring RNA, Northern blot hybridization, was limited to measuring an individual RNA with a specific probe (Alwine *et al.*, 1977). With advances in molecular genetic techniques such as polymerase chain reaction

(PCR), chemical DNA synthesis, and sequencing, came the advent of large-scale technologies like microarrays and RNA-seq which enabled transcription to be monitored globally (Passos, 2014). These techniques have been instrumental in understanding how cells exhibit precise control over gene expression, including identifying the regulons of various transcriptions factors.

Microarrays are large, ordered arrays of DNA probes (usually on a silica chip). To assess differential transcription, mRNA is collected from an experimental and a control and converted to uniquely labeled cDNA. The two samples are mixed in equimolar ratio to compete for hybridization with the ChIP DNA (Gershon, 2002). The amount of signal is proportional to the amount of hybridization, so differences in relative transcription in the two samples can be easily monitored. Microarrays made it possible to monitor the transcription of all genes simultaneously. Microarrays have also been used to identify large genomic deletions or single nucleotide polymorphisms (SNPs), making them a much less expensive and faster approach than whole genome sequencing in some cases (Gershon, 2002). There a few drawbacks to microarrays however, and it is no longer a common methodology. First, gene chips were only commercially available for organisms with the broadest user base. Although large companies like Thermo Fisher, Qiagen, and Bio-Rad offer hundreds of microarrays, those available are limited to common organisms or diagnostic tests. This means researchers working on less commonly studied organisms have to generate their own chips or have them custom made. Second, microarrays are typically made of DNA probes of 30-60 bp which limits the information available about the full-length mRNA, including the precise transcription start site. Finally, since the method relies on hybridization, it is relatively

noisy compared to next-generation methods developed. Relative differences in abundance generally need to be 2-3X different to detect with confidence.

An alternative approach, which has virtually replaced microarrays for transcriptomic studies is RNA-sequencing, or RNA-seq. In RNA-seq, isolated RNA is converted to cDNA and subjected to next-generation sequencing, providing both qualitative and quantitative information on highly and lowly expressed genes (Hrdlickova *et al.*, 2017). Once sequencing data is obtained, analysis often requires advanced software or coding experience. The first step is generally aligning the data to a reference genome. While many existing reference genomes are available and are useful, reference genomes are not required as individual transcripts can be assembled de novo. RNA-seq has the additional advantage in that because multiple samples can be differentially labeled with unique adapters containing barcodes, samples can be multiplexed and loaded into the same sequencing run (Hrdlickova *et al.*, 2017). The conversion of mRNA to cDNA prior to sequencing can introduce bias results. Further sample processing is sometimes required for instance, PCR amplification to generate large amounts of cDNA, a process that can lead to uneven amplification and skew results.

Eukaryotic mRNAs are polyadenylated, and this property can be used both to remove rRNAs and prime reverse transcription. Bacterial mRNAs, on the other hand, require specialized kits designed to deplete rRNAs and generate cDNA libraries. To control for experimental variables that might affect the results of RNA-seq studies, complementary methodologies are usually utilized to test the reproducibility. RNA levels can be monitored directly using RT-PCR or qPCR. Transcription from a promoter

can also be monitored *in vivo* using microscopy, flow cytometry, and microfluidics devices, methods that are scalable if global analyses are required (Ambriz-Avina *et al.*, 2014, Martins & Locke, 2015). Usually these methodologies require a reporter molecule be fused to a promoter of interest and measuring transcription indirectly through reporter protein production.

One benefit of reporter protein production is that it allows for single cell analysis. Flow cytometry followed by cell sorting, or fluorescence-activated cell sorting (FACS), is an exciting technique because it provides not only a method to look at distribution of transcription in single cells of a population, but also a way to separate and collect those cells for further analyses. This is of particular interest to researchers studying heterogeneity among cells within a population (Muller & Nebe-von-Caron, 2010). Rosenthal et. al used a fluorescent reporter followed by microscopy to show that *sucC* expression was highly heterogenous, with only ~3.8% of the population expressing detectable production of the reporter (Rosenthal et al., 2018). To further characterize this subpopulation and to understand how *sucC* is regulated, the authors separated *sucC*-expressing cells and subjected them to RNA-seq. Interestingly, *sucC* expression was correlated with competence genes and was later shown to be regulated by the master competence regulator, ComK. Even though the ComK regulon was previously described, it had not been previously shown to regulate genes involved in central carbon metabolism (Rosenthal *et al.*, 2018). These observations would likely not have been made using a whole-population approach because only a subset of cells in a population need to express the gene at any given time. Transcriptomic studies have the potential to discover genes or uncover regulatory networks. In addition to transcriptomics, additional

"-omics" tools such as metabolomic (described below) can used to aid functional discovery.

### 1.4.2. Metabolomics

Organisms utilize, produce, and secrete a variety of metabolites in response to external and internal cues. Metabolites act not only as intermediaries in metabolism, but also as important signals controlling metabolic flow along central pathways (Sonenshein, 2007). In addition, metabolites can interact with transmembrane proteins which, depending on the function of the protein, can affect substrate transport or initiate signaling cascades (Johnson *et al.*, 2016). Furthermore, the secretion of certain metabolites can affect intra- and extracellular environments. For example, during growth on acetogenic carbon sources, such as glucose, *E.coli* produces acetate which is secreted to regenerate NAD+ (Wolfe, 2005). As a result, secretion of acetate raises the intracellular pH while lowering the extracellular pH. Since metabolite pools have a primary role in fitness, there is a growing need to determine the function of metabolites and the pathways in which they are involved by linking metabolic information to phenotypes, followed by additional hypothesis-driven mechanistic and functional analyses.

The most common methods for metabolomics are untargeted and targeted mass spectrometry. First, particles are converted into ions by an ionization source, often Matrix-Assisted Laser Desorption/Ionization (MALDI) and electrospray ionization (ESI). Second, ions are sorted based on their mass/charge (*m/z*) with a mass analyzer such as time-of-flight (TOF), ion trap, and quadrupole. Finally, the *m/z* is measured with

an ion detector. MS technologies were greatly improved with tandem mass spectrometry (MS/MS) where after the first analyzer, ions are fragmented, separated, and detected again. In untargeted metabolomics, the spectrum obtained from a sample is compared to a database of available m/z ratios, allowing putative identifications (IDs) to be made. Since MS/MS provides limited structural information and metabolites generally do not contain unique peptide sequences, sometimes multiple compounds share identical m/z, complicating assignments. Although a wide range of metabolites are detectable, detection depends on many variables, including the abundance and lability of the metabolite, the method of extraction and chromatographic separation, and the type of MS. Therefore, untargeted metabolomics generally requires the use of multiple extraction and separation techniques to ensure the largest array of metabolites is sampled. In contrast, targeted metabolomics measures a specific metabolite of interest. Although targeted approaches are more sensitive and selective, they require prior knowledge and optimized methods for the specific metabolite (Johnson *et al.*, 2016) . Neither untargeted or targeted approaches offer information on metabolic rates, which is important for determining if a metabolic change is a direct cause of the perturbation or if it is a downstream effect. However, the general methodology can be coupled with stable-isotope labeling to follow metabolic flow and in terms of functional characterization, can aid in identifying the most likely pathway associated with a particular genotype or phenotype (Johnson *et al.*, 2016). One major challenge to metabolomics is metabolite identification and validation. This so called ‘dark matter’ makes up over 95% of spectra identified in an untargeted metabolomics experiment (da Silva *et al.*, 2015). To overcome the metabolite identification challenge, multiple methods can be combined to

examine both intracellular and extracellular (exometabolome) metabolomes including nuclear magnetic resonance (NMR), high performance liquid chromatography (HPLC), and Fourier transform infra-red spectroscopy (Bingol, 2018, Pinu & Villas-Boas, 2017) for metabolite identification.

Following metabolite identification, assigning a function to a particular metabolite can still be challenging. With advancements in data analysis and data storage, as well as the use of robotics, we now have the proper tools to obtain and combine big data from multiple omics approaches, allowing us to examine the relationships between the molecules that make up the cell and understand how they influence each other. Environments can be altered to determine how organisms respond, or we can look at entire communities to examine how signals from one cell or bacterium can influence others in a population. More targeted approaches can also be taken; for examples, deleting a gene with a potential enzyme activity and comparing the mutant metabolome to the wild type to identify a possible pathway involved in metabolism. Progress on the uncharacterized gene problem has been slow. Aside from being technically difficult, it is difficult to provide justification to study genes that lack obvious significance. From a technical standpoint, I have outlined above some of the approaches being used to gain the phenotypic insight needed for hypothesis-driven analyses. The justification challenge is in some ways more difficult, as impact has to be judged in retrospective, though the history of discovery and innovation in science research makes a good case for why we should study basic science.

### 1.5. Gene characterization in *Bacillus subtilis*

At the onset of the post-genomics era, most global studies initially focused on yeast. Later advances in bacterial genomics and molecular biology approaches allowed us to extend studies to other model organisms. The most well-studied model bacterium is the Gram-negative organism *E. coli*. The *E. coli* genome was closed in 1997. Despite the relatively small genome size (approximately 4,500 genes) and reduced complexity compared to yeast, many gene products remain experimentally uncharacterized in *E. coli*. Nevertheless, it was recently shown that the number of *E. coli* gene products with gene ontology (GO) terms associated with experimental evidence (albeit not necessarily an assigned function) has increased from 2,462 (Keseler *et al.*, 2013) to 3,350 (Keseler *et al.*, 2017) in just four years, reflecting an increase from 54% to 74%. Most of what we know about microbial metabolism is based on *E. coli* (and *Salmonella*). Our increase in understanding *E. coli*, however, has not necessarily led to a concomitant increase in our understanding of other bacteria. This is exemplified by a study that compared four annotation tools and found that the number of annotations derived by three out of four tools for *E. coli* was nearly double that of *Clostridium difficile* 630 (Griesemer *et al.*, 2018). The need for better representation among other organisms is emphasized by the number of discrepancies in essential processes between *E. coli* and other bacteria. For instance, DNA replication is generally thought of as somewhat uniform among bacteria, yet is quite different in the Gram-positive model organism, *B. subtilis*. Whereas DNA replication in *E. coli* is regulated by proteins that affect the levels of active replication initiator, DnaA-ATP, no regulators of DnaA-ATP hydrolysis are known in *B. subtilis* (Camara *et al.*, 2005, Jameson & Wilkinson, 2017). Instead, *B. subtilis* DnaA is

regulated by protein-protein interactions with the negative regulator YabA (Scholefield & Murray, 2013) and Soj (Scholefield *et al.*, 2012). Furthermore, unlike *E.coli*, DNA replication in *B. subtilis* is also regulated during the process of sporulation (Jameson & Wilkinson, 2017), whereas *E. coli* does not generate spores. Since gene identification and characterization are critical to understanding essential, fundamental processes as well as the more complex specialized processes such as survival in harsh environments and cellular communities, additional efforts must be made to identify gene functions.

Making gene characterization difficult is that some genes are expressed in a fraction of a population. This heterophysiology, thought to reflect "bet-hedging" (Seger, 1987), is evident in bacterial communities such as biofilms, and makes studies that utilize population-level data problematic. Biofilms are dense communities of cells held together by a matrix, usually consisting of extracellular polymeric substances, proteins, and DNA, making their removal difficult. Cell differentiation is thought to contribute to the resilience of biofilms by delegating certain energy-costly tasks (such as matrix production) only to a subpopulation (Kearns, 2008). This "division of labor" enables other subpopulations to divert energy to processes such as sporulation, motility, or competence which can enable those cells to proliferate if environmental conditions worsen (Kearns, 2008). Another challenge to studying these complex community structures is that extracellular factors like quorum-sensing molecules and metabolites act as mobile signals that can be produced and detected by different cell types in a community (Rutherford & Bassler, 2012). *B. subtilis* is an exceptional model organism for studying bacterial developmental processes such as biofilm formation as it can undergo numerous cell differentiation programs including sporulation, competence,

matrix producing and motile (Lopez & Kolter, 2010). Surprisingly, only 60% of the *B. subtilis* genome has been annotated using experimental evidence.

In this dissertation, I used a misexpression library from *B. subtilis* to identify putative DNA-binding proteins that, when misexpressed, lead to changes in subcellular organization. Specifically, alterations in DNA structure, cell morphogenesis and cell division. I analyzed a set of putative DNA-binding proteins and examined their effects on gene expression and subcellular organization when misexpressed. Eight proteins, *yxaD*, *yesS*, *yvmB*, *yhjH*, *ywhA*, *mdxR*, *ywbI*, and *ywgB* appear to act as transcriptional regulators and two putative DNA-binding proteins, *ycxD* and *ykoM* that do not appear to regulate transcription. In addition, I further characterized the MarR-like protein, YxaD. YxaD misexpression disrupts chromosome segregation and DNA replication. Since YxaD was shown to interact with ScpA and HolA by yeast two hybrid (), I examined the possible involvement of this interaction with YxaD's misexpression phenotype. By RNA-sequencing, I showed that YxaD represses the *yxaKC* operon. I identified and classed variants of YxaD that do not result in the misexpression phenotype. I further analyzed YxaD as a transcriptional regulator and showed that it binds to two sites in the intergenic region between *yxaD* and *yxaKC* to regulate expression of both operons. Lastly, I used NMR to analyze the excreted overflow metabolites from Δ*yxaD*, Δ*yxaKC* and wild type cells and found that Δ*yxaD* cells excreted more 2-ACL and less acetoin. These results suggest a possible role of YxaKC in regulated overflow metabolism and/or utilization of secreted metabolites

# 2. THE *BACILLUS* ECTOPIC INDUCIBLE GENE EXPRESSION LIBRARY (BEIGEL): A MULTIFUNCTIONAL RESEARCH TOOL FOR GENE DISCOVER

## 2.1. Introduction

Post-genomic techniques and methodologies provide incredible insight into systems biology and allow us to examine global effects in response to a change or perturbation. For example, we now have the ability to sequence an organism's genome and monitor the transcriptome, proteome and metabolome, enabling systems-level analyses for understanding differences between genetically distinct isolates and responses to changing nutrient availability and other environmental signals. Even with these advances, our inability to link gene products with observed biological variations remains a major challenge for post-genomic biology. The problem is underscored by the fact that nearly half of the genes in even some of the most well studied organisms remain uncharacterized, are misannotated, or are annotated based solely on bioinformatic approaches without experimental validation (Bork, 2000), (Gerdes *et al.*, 2011), (Riley *et al.*, 2006). 30% of these are estimated to be enzymes (Ellens *et al.*, 2017), consistent with the fact that a large number of enzymatic reactions are known to exist that have yet to be associated with a particular gene product (Hanson *et al.*, 2009), (Sorokina *et al.*, 2014). Many of the uncharacterized genes are conserved across multiple genera or even all domains of life, suggesting they encode for important functions that have yet to be elucidated.

Functional characterization of uncharacterized genes remains the major challenge to post-genomic biology and it has been challenging for several reasons. First,

uncharacterized genes often lack obvious knockout phenotypes because of functional redundancy and buffering mechanisms within cells. For instance, a study in *E. coli* found that 10 out the 12 genes encoding known penicillin binding proteins needed to be deleted before an obvious phenotype was detected (Denome *et al.*, 1999). Moreover, standard laboratory conditions are not often adequate to detect a knockout phenotype due to context specific functions. For instance, the genes that make up SMC condensin complexes in *B. subtilis*, *smc*, *scpA*, and *scpB*, are required for growth on LB media at 37 °C but are not required at low temperatures (Gruber *et al.*, 2014), whereas *B. subtilis yvcK* is not essential for growth on LB but is required when grown on gluconeogenic substrates (Gorke *et al.*, 2005). Interestingly, these studies highlight the importance of studying genes in various growth conditions as genes considered to be 'essential' in the laboratory might not be in native environments and *vice versa*. The construction and analysis of a *B. subtilis* knockout library found that out of an estimated 4,245 encoded genes, only 257 were considered to be essential for growth on LB media at 37 °C (Koo *et al.*, 2017); however this number would be expected to go up or down depending on growth context. The lack of easily discernible phenotypes is a problem because it increases the difficulty of forming testable hypotheses regarding an unknown gene's function.

Gene characterization in the post-genomic era has seen some recent success by assaying gene knockout libraries for fitness in response to changes in growth conditions and compounds (Nichols *et al.*, 2011, Koo *et al.*, 2017). These datasets, termed 'phenomics', are particularly useful for revealing how uncharacterized genes are situated within pathways. One study examined an *E.coli* mutant library under different growth

conditions and was able to class 25% of *E. coli's* unknown genes into categories (i.e. heat shock) (Nichols *et al.*, 2011). While these designations are useful, this information alone often does not provide enough information to inform specific hypotheses regarding function.

An additional strategy to link gene products to phenotypes is by screening for gain-of-function phenotypes following misexpression. Artificial expression of gene at either higher than wild-type levels or in a non-native expression context can perturb homeostasis and thus produce phenotypes that can then be utilized to form testable hypotheses regarding gene function. For this study, we utilized misexpression strains from a library of *B. subtilis* misexpression strains the *Bacillus* Ectopic Inducible Gene Expression Library (BEIGEL) (Figure 2.1) (Duan *et al.*, 2016b) . The BEIGEL comprises 810 strains, each with one uncharacterized gene under the control of an inducible promoter. Each strain has a misexpression construct integrated into the *B. subtilis* chromosome at an ectopic locus providing titratable expression from a single gene copy.

We screened the BEIGEL for phenotypes associated with misexpression that affect growth and subsequently, those that perturb DNA replication, cell division, and overall cell morphology. Forty-nine strains with phenotypes were identified. In addition, we combined phenotypic data with suppressors selections followed by whole-genome sequencing to probe genetic interaction partners of a subset of the induced genes. Of the 49 strains with phenotypes, 26 corresponded to uncharacterized DNA-binding proteins. Ten of these, none of which had obvious knockout phenotypes under standard laboratory growth conditions, were selected for further analysis. Many DNA-binding proteins act as

transcription factors, whereas others, such as *E. coli* SlmA (Bernhardt & de Boer, 2005) and *B. subtilis* RefZ (Miller *et al.*, 2016) have been shown to use their DNA-binding activity to localize another activity to a specific subcellular location. To assess if the DNA-binding proteins identified in the screen acted as transcription factors, RNA-seq was performed on a subset, *mdxR, ycxD, yesS, yhjH, ykoM, yvmB, ywbI, ywgB, ywhA, yxaD*. For each strain, we analyzed transcription for wild type, an unmarked deletion strain, and the BEIGEL strain following induction. The strains were grown in expression conditions where the gene of interest was previously shown to be expressed (Nicolas *et al.*, 2012). We identified eight genes, *mdxR, yesS, yhjH, yvmB, ywbI, ywgB, ywhA,* and *yxaD* that appeared to affect transcription and two genes, *ycxD* and *ykoM* that I not show altered transcription relative to wild type.

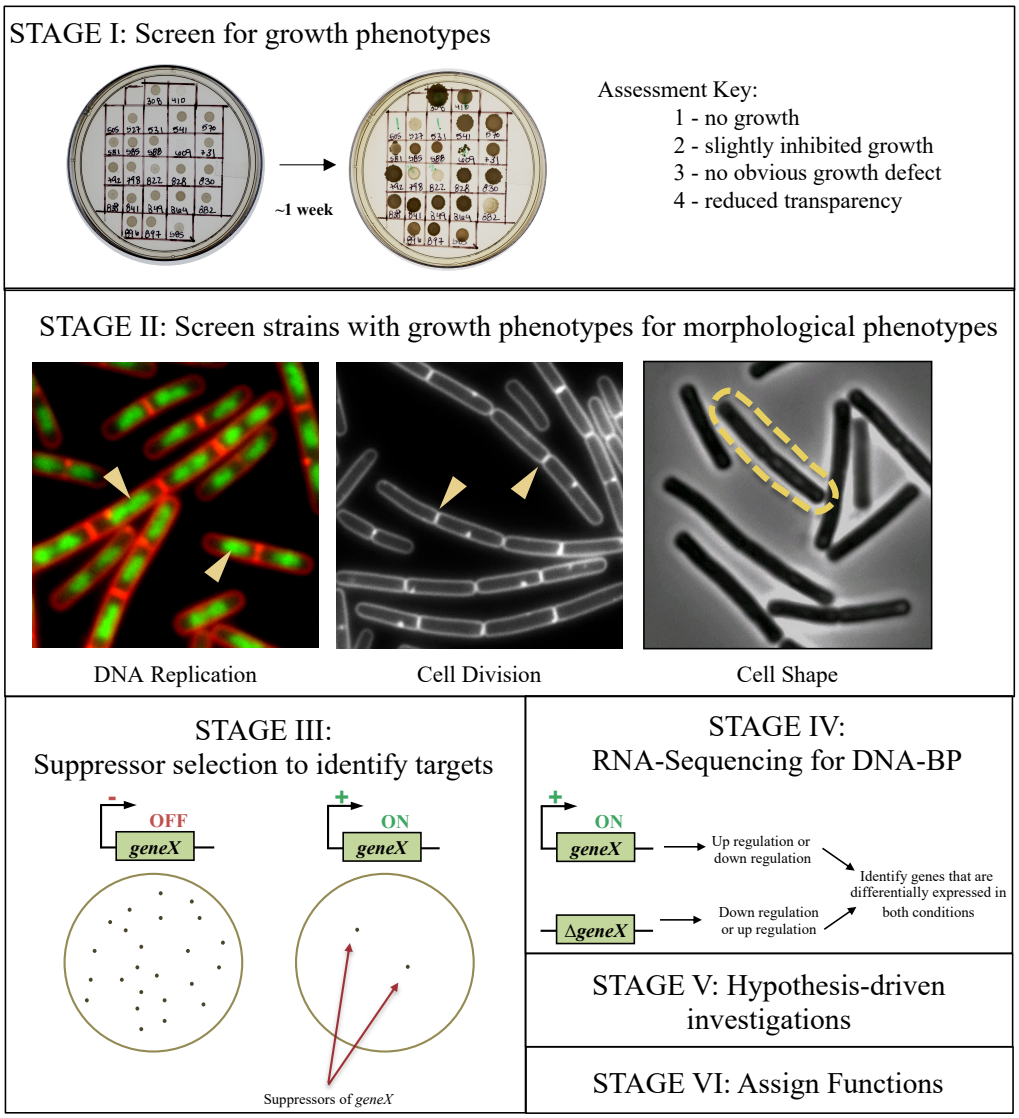



**Figure 2.1 BEIGEL strategy for gene characterization.**
Stage I: screen BEIGEL for growth phenotypes on plates containing inducer. Stage II: screen strains with growth phenotypes for morphological phenotypes by fluorescence microscopy. Stage III: suppressor selection to identify genetic targets of misexpressed genes. Stage IV: RNA-sequencing performed on putative DNA-binding proteins. Stage V: form testable hypotheses and perform experiments to stage VI: assign gene functions.

## 2.2. Materials and Methods

### 2.2.1. Gibson Assembly

UP fragments were generated by an initial assembly of the fragments: *amyEdown* PCR amplified with OAM009 and OAM010 from *B. subtilis* 168 and *lacI* (OAM011 and OAM012) from pDR111 to generate a fragment of 5,612 bp. DOWN fragments were generated by a similar initial assembly of *amyEup* OAM001 and OAM002 from *B. subtilis* 168, *specR-optRBS-lacO-* $P_{hyperspank}$ using OAM013 and OAM014 from pDR111 generating a 6,012 bp fragment. UP and DOWN fragments were gel purified using the QIAQuick gel purification system and used as templates in a final PCR using primers OAM001 and OAM012 for UP and OAM011 and OAM010 to generate final DOWN. The UP and DOWN constructs were introduced into the chromosome of *Bs168* using Gibson assembly with UP and DOWN fragments and a test gene product PCR amplified with extensions complementary to UP and DOWN fragments at an equimolar ratio and subsequently transformed into *Bs168* competent cells. Bs168 with the correct insertion were selected for on plates containing spectinomycin (100 µg/ml). Correct insertion into the *amyE* locus was verified by patching colonies on starch containing plates and by PCR amplifying the gene of interest. Once confirmed, genomic DNA of this strain was used as template to generate stocks of UP (OAM010 and OAM013) and DOWN (OAM001 and OAM014). Stocks of UP and DOWN were diluted to a working concentration of 20 ng/µl enabling more efficient normalization of BEIGEL gene fragments to equimolar ratios.

**Table 2.1 Strains used in Chapter 2**

| Strain | Description | Reference |
|---|---|---|
| *Bacillus subtilis* 168 | | |
| BAM083 | *amyE::Phyperspank-optRBS (spec)* | This study |
| BAM049 | *amyE::Phyperspank-ykoM (spec)* | This study |
| BSH050 | *amyE::Phyperspank-yxaD (GTG start) (spec)* | This study |
| BEA089 | *amyE::Phyperspank-mdxR (spec)* | This study |
| BEA121 | *amyE::Phyperspank-ycxD (spec)* | This study |
| BEA136 | *amyE::Phyperspank-yesS (spec)* | This study |
| BEA144 | *amyE::Phyperspank-yhjH (spec)* | This study |
| BEA170 | *amyE::Phyperspank-yvmB (spec)* | This study |
| BEA174 | *amyE::Phyperspank-ywbI (spec)* | This study |
| BEA176 | *amyE::Phyperspank-ywgB (spec)* | This study |
| BEA177 | *amyE::Phyperspank-ywhA (spec)* | This study |
| BSH062 | Δy*cxD* | This study |
| BSH063 | Δ*mdxR* | This study |
| BSH064 | Δ*yvmB* | This study |
| BSH065 | Δ*ywgB* | This study |
| BSH067 | Δ*ywbI* | This study |
| BSH068 | Δ*yhjH* | This study |
| BSH069 | Δ*ykoM* | This study |
| BSH070 | Δ*yxaD* | This study |
| BSH071 | Δ*yesS* | This study |
| BSH072 | Δ*ywhA* | This study |
| BYD080 | *amyE::Phyperspank-mdxR (spec),*<br>*sacA::Phyperspank-lacZ (erm),*<br>*yycR::Phyperspank-mdxR (cat)* | |
| BYD034 | *amyE::Phyperspank-yttP (spec),*<br>*yhdG::Phyperspank-yttP (phleo),*<br>*ycgO::Phyperspank-yttP (tet),*<br>*sacA::Phyperspank-lacZ (erm)* | |
| BYD048 | *amyE::Phyperspank-yodL (spec),*<br>*ycgO::Phyperspank-yodL (tet),*<br>*yhdG::Phyperspank-yodL (phleo),*<br>*sacA::Phyperspank-lacZ (erm)* | |
| BYD076 | *amyE::Phyperspank-yisK (spec),*<br>*yhdG::Phyperspank-yisK (phleo),*<br>*yycR::Phyperspank-yisK (cat),*<br>*sacA::Phyperspank-lacZ (erm)* | |
| BEA052 | *amyE::Phyperspank-yisK (spec)* | |
| BEA098 | *amyE::Phyperspank-yodL (spec)* | |
| BJH026 | *amyE::Phyperspank-yttP (spec)* | |

**Table 2.2 Oligonucleotides used in Chapter 2**

| Oligo | Sequence 5'-3' |
|---|---|
| OAM001 | agaagcgttagcggcagcaagtgat |
| OAM002 | ccatgtctgcccgtatttcgcgtaaggaaatccattatgtactatttcgatcagaccag |
| OAM009 | gaaaacaataaacccttgcatagggggatcgggcaaggctagacgggacttacc |
| OAM010 | atggacacaacaacagcaaaacaggc |
| OAM011 | TAATGGATTTCCTTACGCGAAATA |
| OAM012 | GCTAGccgCATGCAAGCTAATT |
| OAM013 | agtagttCCTCCTTAtgtAAGC |
| OAM014 | GATCCCCCTATGCAAGGGTTTATT |
| OJH001 | CATATGTAAGATTTAAATGCAACCG |
| OJH002 | CTACAAGGTGTGGCATAATGTGT |

### 2.2.2. BEIGEL Construction

BEIGEL primers were designed using a computer algorithm. Lyophilized primer sets were provided in 96-well plate formats by IDT and resuspended to a final concentration of 10 µM each primer. BEIGEL genes were PCR amplified using Phusion HF Polymerase and primer sets for each gene in 25 µl reactions. Purified PCR products were normalized by a standard molar concentration for all BEIGEL PCR fragments (based on a test quantification) of 1000 bp at a concentration of 30 ng/µl. BEIGEL gene products were assembled with UP and DOWN fragments in 96-well format using Gibson assembly using a ratio of 1 µl, 2 µl and 2 µl respectively. Fifteen µl reactions were incubated for 1 hr at 50 °C. Assembled fragments were then transformed into *B. subtilis* and selected for on LB plates containing 100 µg/ml spectinomycin and 0.2% glucose (w/v) (LB-S-G). Single colonies were streaked for isolation on LB-S-G plates and patched on starch containing plates to check for integration at the *amyE* locus. Single isolates were inoculated in 5 ml LB + 0.2% glucose (w/v) and 5 µl was immediately spotted on LB plates containing 100 µg/ml spectinomycin and 1.0 mM

isopropyl-β-D-thiogalactopyranoside (IPTG). Insertions of the correct gene fragments were confirmed by PCR amplifying the gene region in *amyE* with OJH001 and OJH002.

### 2.2.3. Microscopy-based morphological screen

BEIGEL strains were streaked from frozen glycerol stocks onto LB-S-G plates. Single colonies were inoculated in 5 ml LB containing 0.2% (w/v) glucose and grown to mid exponential at 37 °C. Cultures were back diluted in 25 ml LB media to an $OD_{600}$=0.00625 in 250 ml baffled flasks at 37 °C shaking in a water bath set to 280 rpm. Cultures were induced with 1.0 mM IPTG at $OD_{600}$ = 0.05. Samples for microscopy were taken 90' post induction unless otherwise indicated. Fluorescence microscopy was performed with a Nikon Ti-E microscope equipped with a CFI Plan Apo lambda DM 100× objective, Prior Scientific Lumen 200 illumination system, C-FL UV-2E/C 4′,6-diamidino-2-phenylindole, and a CoolSNAP HQ2 monochrome camera. All images were captured with NIS Elements Advanced Research (version 4.10), and processed with NIS Elements and ImageJ64 (W., 1997-2015). For image capture, 1 ml of cells were pellet at 6,010 x g for 1 min in a tabletop microfuge at room temperature, supernatants were removed by aspiration, and pellets were resuspended in 7 µL of 1X PBS containing 0.02 mM 1-(4-(trimethylamino)phenyl)-6-phenylhexa-1,3,5-triene (TMA-DPH) or 7 µl 1X PBS containing DAPI DNA stain (2 µg/ml) (Molecular Probes) and FM4-64 membrane stain (3 µg/ml) (Molecular Probes). Cells were mounted on glass slides with polylysine-treated coverslips.

### 2.2.4. Spontaneous suppressor selection

Independent cultures were started from six single colonies of BYD034, BYD048, BYD076, or BYD080. Individual colonies were used to inoculate 5 ml LB cultures. Cultures were grown at 37 °C for 6 hrs when 0.3 µl of each culture was diluted in 100 µl LB and subsequently plated on LB agar plates with 100 µg/ml spectinomycin (LB-spec) and 1 mM IPTG. Plates were incubated overnight at 37 °C until suppressor colonies grew. Suppressors were patched on LB-spec and LB-spec with 1mM IPTG, 40 µg/ml 5-bromo-4- chloro-3-indolyl-β-D-galactopyranoside (X-gal). The patched plates were grown at 37 °C overnight. Only patches that appeared blue in the presence of X-gal were selected for further analysis. In addition, each misexpression cassette was moved into a clean genetic background to ensure the original construct remained functional in the presence of IPTG.

### 2.2.5. RNA sequencing

Individual colonies were used to inoculate independent cultures of $P_{hy}$-*mdxR*, $P_{hy}$-*ycxD*, $P_{hy}$-*yesS*, $P_{hy}$-*yhjH*, $P_{hy}$-*ykoM*, $P_{hy}$-*yvmB*, $P_{hy}$-*ywbI*, $P_{hy}$-*ywgB*, $P_{hy}$-*ywhA*, $P_{hy}$-*yxaD*, and $P_{hy}$-empty in 5 ml LB + 0.2% (w/v) glucose and grown to mid exponential. Cultures were then back diluted into a 25 ml LB to a final $OD_{600}$ = 0.01875 in 250 ml baffled flasks at 37 °C shaking in a water bath set to 280 rpm. After 1 hr 15 minutes, cultures at $OD_{600}$ = 0.3 were induced with 1.0 mM IPTG. After 15 minutes, 500 µl samples of culture were added to 1 ml RNA Protect Reagent (Qiagen). For RNA isolated from deletions in stationary or exponential phase, cultures were grown similar to $P_{hy}$- cultures except grown in LB + 0.3% (w/v) glucose (Nicolas *et al.*, 2012). For growth in

sporulation, cultures were grown as previously in LB (no glucose) to and $OD_{600}$ = 0.5. Sporulation was induced by resuspension at 37 °C according to the Sterlini-Mandelstam method and samples for RNA-seq were taken 1 hr or 6 hr into growth (Sterlini & Mandelstam, 1969). Five-hundred µl samples were collected 1 hr or 6 hr into sporulation as indicated. RNA was isolated using the RNeasy mini kit (Qiagen), with the exception that the 6 hr sporulation samples were vortexed for 1 hr with lysozyme treatment. RNA isolation was followed by DNase I treatment (Qiagen) to remove DNA and ribosomal RNA was removed using Ribo-Zero rRNA Removal Kit (Gram-Positive Bacteria) (Illumina). 50-bp single end read libraries were prepped with a TruSeq Stranded Total RNA Kit (Illumina) and sequenced on an Illumina HiSeq 2500. Reads were mapped to each open reading frame (ORF) in the *B. subtilis* 168 genome (GenBank: NC_000964.3) using with kallisto (Bray *et al.*, 2016) and used edgeR (Robinson *et al.*, 2010) for differential gene expression analysis. Lowly expressed ORFs were filtered (<1 count per million). The single-factor exact test was used and differentially expressed genes were reported with a false discovery rate cutoff of <0.05.

## 2.3. Results

### 2.3.1. Generation of the *Bacillus* Ectopic Inducible Gene Expression Library (BEIGEL)

To create the BEIGEL, we selected genes annotated in BsubCyc (Caspi *et al.*, 2014) as hypothetical (131), conserved hypothetical (537), putative transcriptional regulators (100), putative integral membrane proteins (101), and putative membrane proteins (21) (See Table 2.3). While this is a substantial fraction of the total *B. subtilis*

genes (~20%, or 894/4175), it is not inclusive of all genes annotated as experimentally uncharacterized.

To generate the large number of misexpression constructs while avoiding passage through *E. coli*, we assembled each misexpression construct in vitro using the Gibson assembly method (Gibson *et al.*, 2009) and took advantage of the natural competence of *B. subtilis* to transform and select for integration of each linear construct into the genome through homologous recombination at an ectopic locus (Figure 2.2). Genome integration also mitigated the potential off-target effects associated with plasmid-borne, multicopy expression systems, the final constructs were introduced in single copy into the chromosome (Figure 2.2). Successful transformants were judged by spectinomycin resistance and the integration of the appropriate size DNA fragment in *amyE*. In total, we readily obtained for 810/894 strains, and the remaining 84 strains were not successfully generated for unknown reasons, despite multiple attempts.

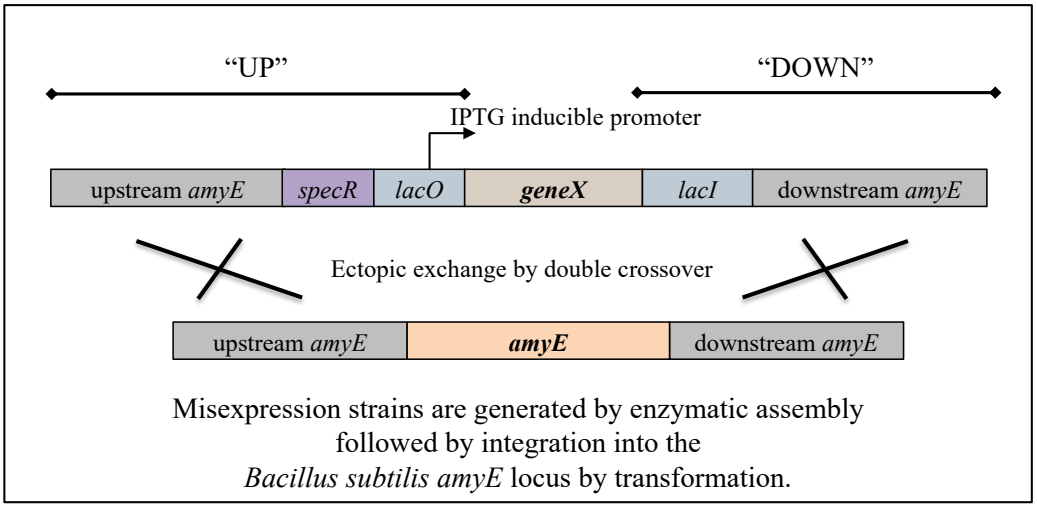



**Figure 2.2 BEIGEL construction.**
PCR generated "UP" and "DOWN" fragments were assembled to an uncharacterized gene of interest, *geneX*, by Gisbson assembly. Constructs were integrated into to the *B. subtilis* chromosome by double crossover integration at the *amyE* locus.

### 2.3.2. Primary screen to identify gene products that resulted in growth defects

To discern which gene products were associated with disruptions in growth, we performed a primary screen on LB solid medium containing inducer. Growth phenotypes were evaluated as a primary screen because in a prototype screen, reduced growth phenotypes were most strongly associated with morphological perturbation phenotypes related to the nucleoid, cell division, and morphogenesis. Under one inducing growth condition, we identified 35/810 misexpression strains that did not grow, 33/810 that grew slow or with reduced viability, and 7 strains that grew well, but exhibited increased colony transparency after 72 hr indicating delayed lysis (Table 2.3).

**Table 2.3 Current status of BEIGEL**

| **Total strains in BEIGEL** | 810 |
|---|---|
| No growth | 35/810 |
| Slow growth | 33/810 |
| Transparent after 72 hr | 7/810 |
| Total growth phenotype when misexpressed | 75/810 |
| Morphological phenotype when misexpressed | 49/75 |
| Morphological phenotype without growth phenotype | 0/22 |
| Suppressors selection analysis | 6 |

### 2.3.3. Secondary screen to identify gene products that perturb subcellular organization

To identify genes that affect DNA replication, cell division and cell shape, we carried out a microscopy-based screen of the 35/35 strains with no growth phenotypes, 33/33 strains with slow growth phenotypes, and 7/7 strains that produced transparent colonies following extended growth. Additionally, we screened 22 strains that did not have a growth phenotype. To screen for morphological phenotypes, strains were grown in rich liquid media under rapid growth conditions to mid-exponential phase, back-diluted, and expression was induced with IPTG. Samples of each strain were taken 30, 60, 90, or 120 min post-induction, stained with both membrane and DNA stains, and imaged. A montage of the observed phenotypes is shown in Figure 2.3. In total, 49 strains displayed obvious changes in cell shape, cell division (longer or shorter cells), and/or DNA structure after 90’ or 120’ induction. 48/97 strains do not show obvious changes in cell structure under the tested growth condition (Table 2.4).

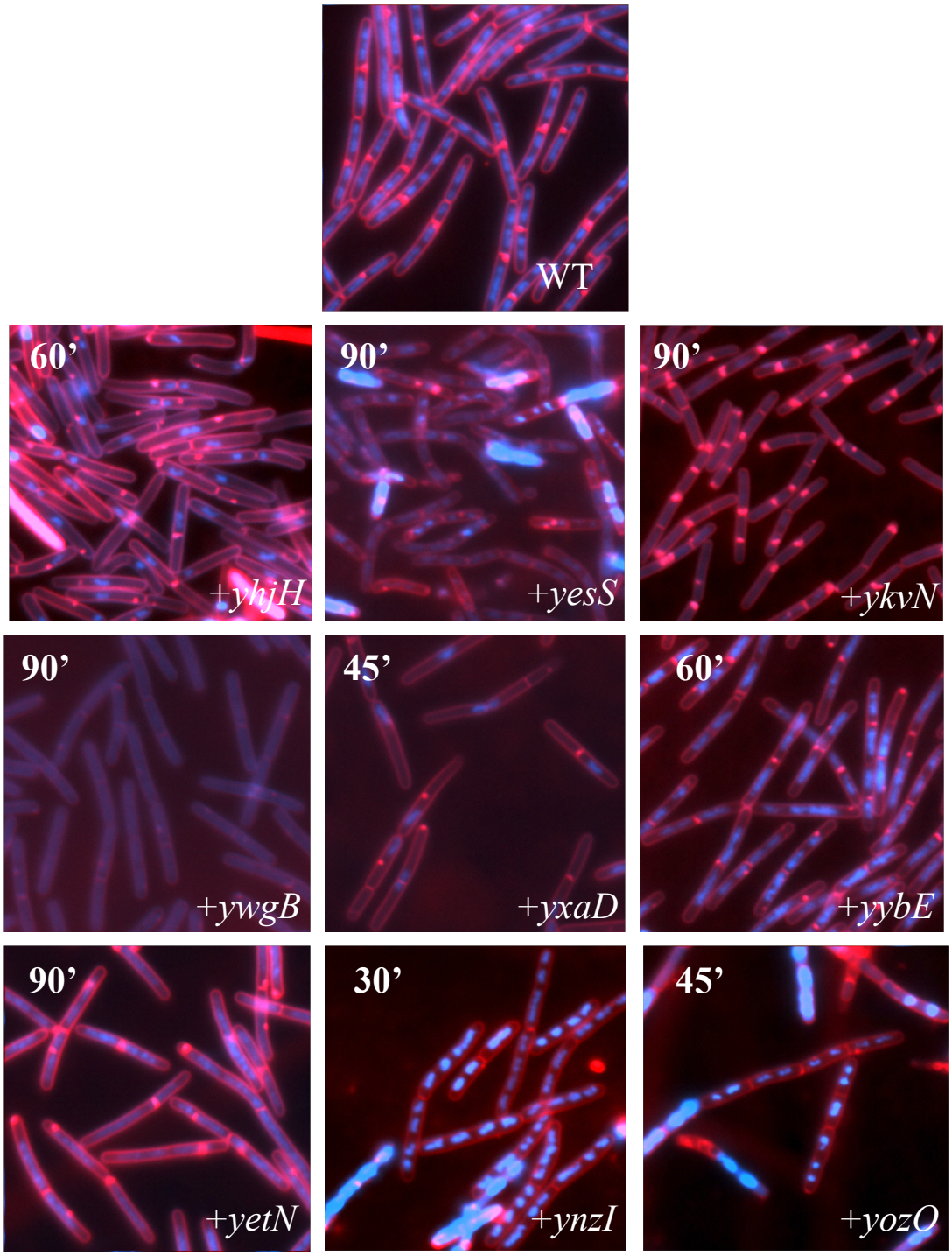
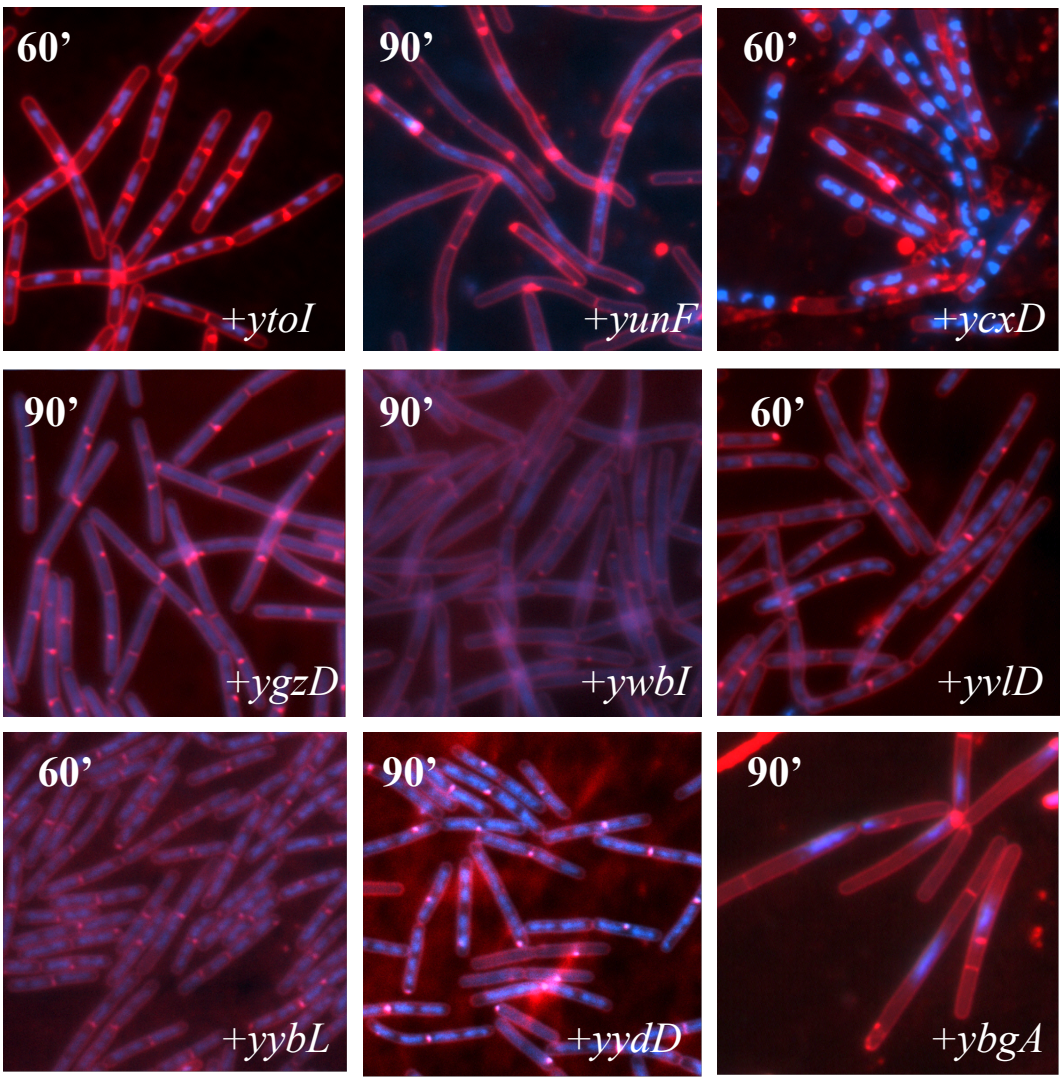



**Figure 2.3 Morphological phenotypes associated with DNA.**
Fluorescence microscopy of *B. subtilis* 168 cells grown in LB medium and overexpressing the indicated gene for the indicated time. Membranes are stained with FM4-64 (red) and DNA is stained with DAPI (blue) for all images.




Not pictured: *ypoP, ysnB, yesM, yvcD*

**Figure 2.3 Continued.**

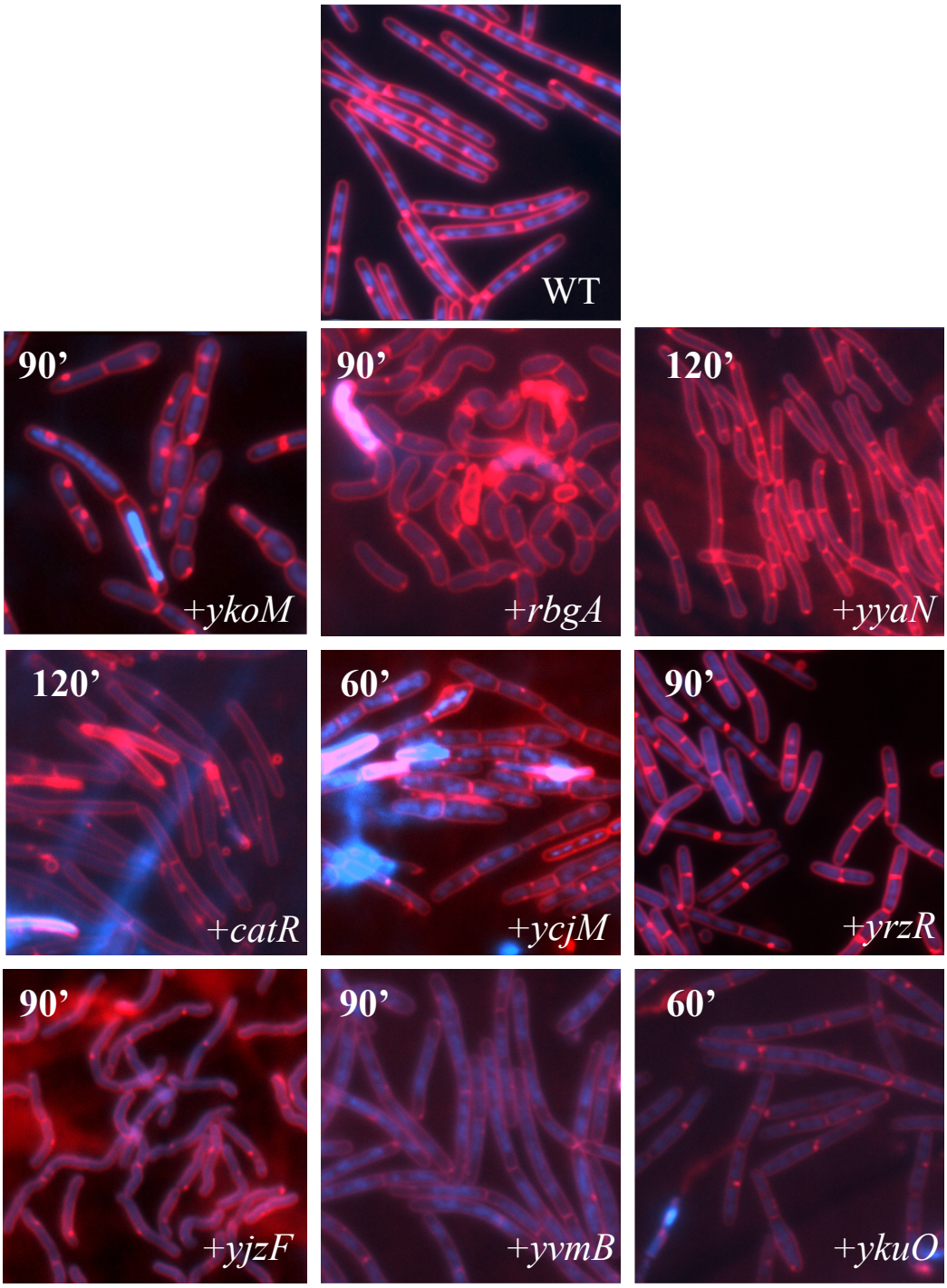



**Figure 2.4 Morphological phenotypes associated with cell shape.**
Fluorescence microscopy performed as in Fig. 2.3.

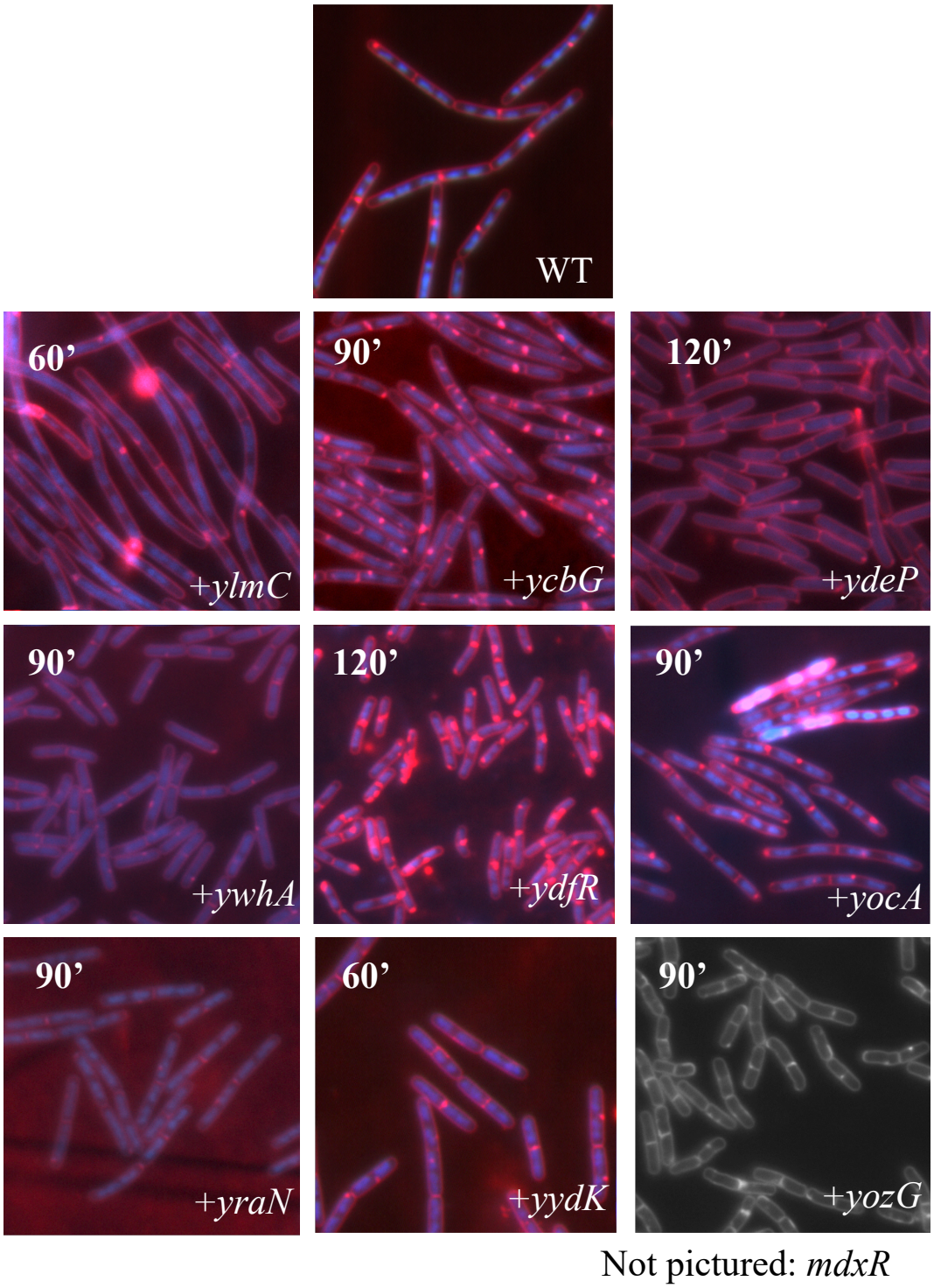



**Figure 2.5 Morphological phenotypes associated with cell division.**
Fluorescence microscopy performed as in Fig. 2.3.

**Table 2.4 Classed morphological phenotypes**

| | |
|---|---|
| **DNA** | *ysnB, ybgA, yesM, yesS* (+cell shape), *yhjH, ykvN, ywgB, yxaD, yybE, yetN, ynzI, yozO, ytoI, yunF, ycxD, ygzD, ypoP* (+cell shape), *ywbI, yvlD, yybL, yvcD, yydD* |
| **Cell Shape** | *ykoM, yyaN, rbgA, yisK, ytxC, catR* (+DNA), *ycjM, yrzR, yjzF, yodL, yvmB, ykuO, yukF* |
| **Cell Division** | *mdxR, ylmC, ycbG, ydeP*(+DNA), *ywhA* (+DNA), *ydfR, yocA, yozG, yraN, yydK* |

### 2.3.4. Suppressor selection analysis to reveal genetic targets

Since some morphological phenotypes are still difficult to link to testable hypotheses to ascertain an uncharacterized gene's function, we next sought to identify genetic targets for a select set of uncharacterized genes by performing suppressor selection analysis. We selected one previously characterized misexpression strains with a known target $P_{hy}$-*refZ* (Wagner-Herman *et al.*, 2012), and three new strains, $P_{hy}$-*mdxR*, $P_{hy}$-*yodL* and $P_{hy}$-*yisK* that had not been experimentally characterized at the time of the study. Expression of *mdxR* and *refZ* during vegetative growth resulted in cell filamentation while expression of *yodL* and *yisK* results in rounded-up cells (Figure 2.6). Several strategies were used to exclude suppressors possessing mutations in the expression constructs. First, a second ($P_{hy}$-*mdxR*) or third ($P_{hy}$-*refZ,* $P_{hy}$-*yodL,* and $P_{hy}$-*yisK*) cassette was introduced into the chromosome to ensure that each strain would produce enough gene product to present a growth defect even if one copy acquired a mutation that was non-functional. Second, each strain expressed $P_{hy}$-*lacZ* as a reporter, allowing us to identify suppressors carrying dominant alleles of *lacI* that were unable to derepress the *Phy* promoter in the presence of IPTG. Lastly, after suppressors were isolated, we confirmed that expression cassettes were still present and functional by

moving the cassette into a clean genetic background. We reduced the likelihood of obtaining clones by growing suppressors in multiple independent cultures and by classing each suppressor by growth on plates and morphological phenotypes by microscopy.

Targeted and whole-genome sequencing were used to identify mutations that conferred resistance to *refZ*, *mdxR*, *yodL* and *yisK* misexpression. Since FtsZ suppressors were found to confer resistance to RefZ misexpression, we first sequenced the coding regions for *ftsZ*. All suppressors obtained for *refZ* were in the previously identified *ftsZ* target. Since the targets of MdxR, YodL and YisK were not known, we performed whole-genome sequencing to identify mutation(s) in suppressors. The YodL and YisK suppressors were found in gene products that regulate cell elongation, *mbl* and *mreB* (Duan *et al.*, 2016b). These data suggested a possible role for YodL and YisK in regulating Mbl and MreB activity. MdxR suppressors were found in *ftsZ* and were G227D and D199V, a result that is consistent with the fact that MdxR induced filamentation.

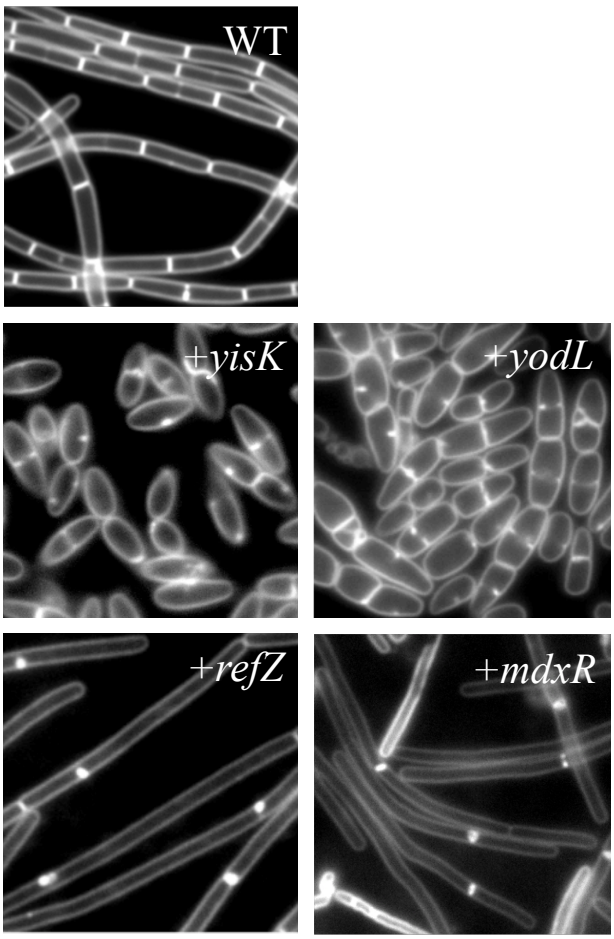



**Figure 2.6 Misexpression phenotypes for suppressors.**
Fluorescence microscopy *of B. subtilis* 168 grown in CH medium and overexpressing the indicated gene for 90 minutes. Cell membranes are stained with TMA-DPH.

### 2.3.5. RNA-seq to identify transcriptional regulators

Bacteria, for the most part, lack subcellular compartmentalization and therefore rely on other organizational cues to temporally and spatially localize molecules within the cell. The nucleoid is a major structure in the cell and possesses positional cues that help direct important functions such as where to initiate and terminate DNA replication and where to divide. We hypothesized that among the uncharacterized DNA-binding

proteins, we might identify not only transcription factors, but also proteins that use DNA to localize functions that affect cellular organization. The BEIGEL misexpression strains harboring genes for putative DNA-binding proteins (100/810) were characterized by protein family, growth and morphological phenotypes, and RNA-sequencing (Table 2.5). Of the 100 strains analyzed, 17 strains did not grow, 9 strains had weak growth, and 4 strains exhibited the delayed reduced transparency phenotype. Strains with growth phenotypes were evaluated by microscopy for morphological defects. In total, morphological phenotypes related to DNA structure, morphology and/or cell division were associated with misexpression of 17/17 strains that did not grow, 7/9 strains with weak growth and 2/4 with reduced colony transparency led to defects in DNA structure, morphology and/or cell division. Representative images are shown in Figure 2.7.

**Table 2.5 Putative DNA-binding proteins in BEIGEL.**

| **Putative DNA-binding Proteins** | 100 |
|---|---|
| Growth phenotype when misexpressed | 30/100 |
| Morphological phenotype when misexpressed | 26/30 |
| RNA-seq performed on | 10/26 |
| Transcriptional Regulators identified by RNA-seq | 8/10 |

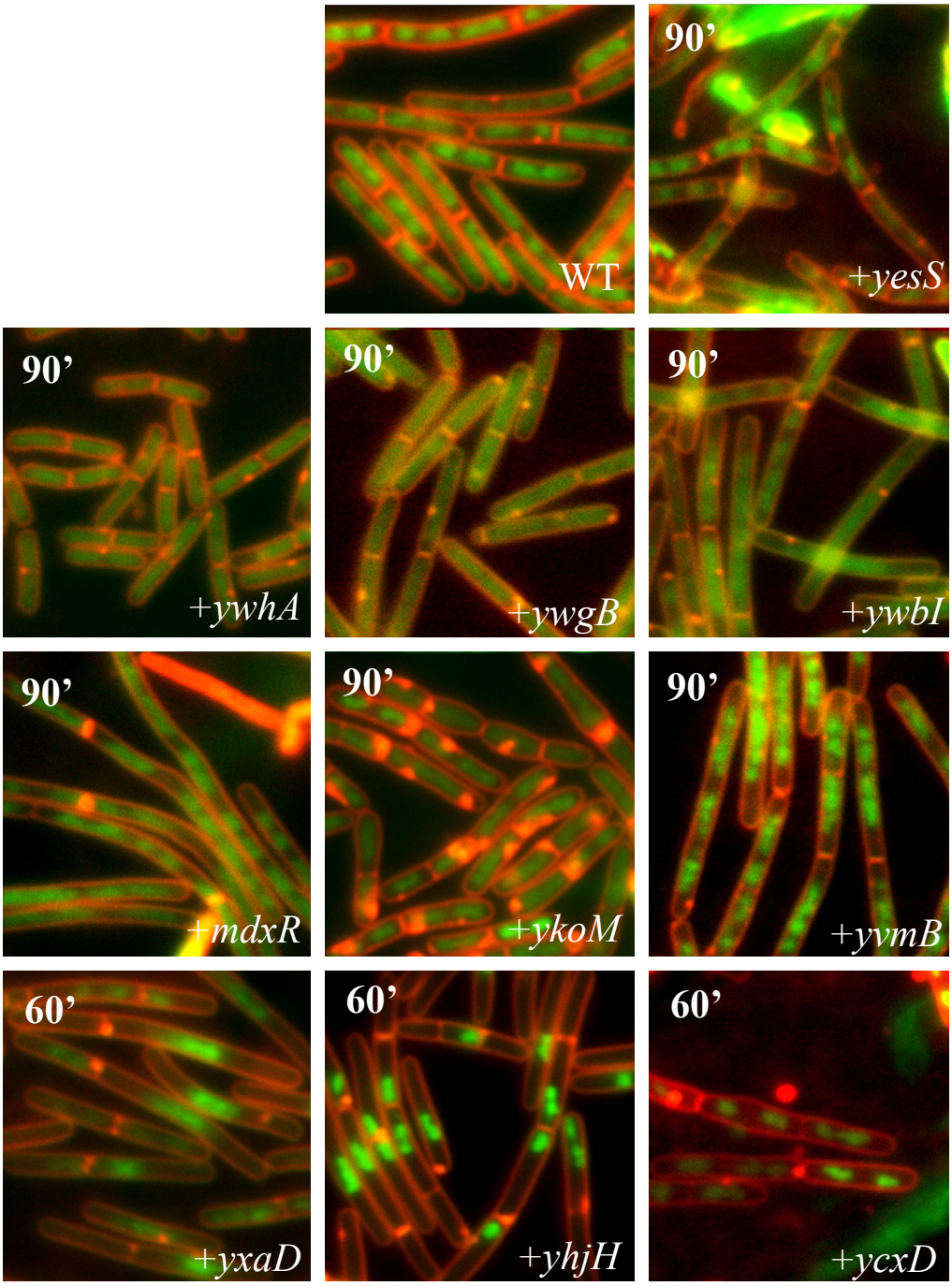



**Figure 2.7 Morphological phenotypes of expressed DNA-binding proteins.** Fluorescence microscopy phenotypes of cells overexpressing DNA-binding proteins analyzed by RNA-seq. Membranes are stained with FM4-64 (red) and DNA is stained with DAPI (green).

We chose to proceed with a subset of 10 putative DNA-binding proteins: *mdxR, ycxD, yesS, yhjH, ykoM, yvmB, ywbI, ywgB, ywhA, yxaD* because misexpression of these 10 gene products sampled the diverse morphological phenotypes we observed and appear to be regulated under similar growth conditions (Nicolas *et al.*, 2012) (Figure 2.7). In order to determine if any of the 10 putative DNA-binding proteins were acting as transcriptional regulators and altering the transcriptional profile, we performed RNA-sequencing experiments on induced cells. In order to measure changes in transcription that are a direct result of the DNA-binding protein and not a regulatory response induced by the cell, samples for RNA processing were collected 15 min post induction. Since we still suspected that 10 min post expression would still result in indirect changes to the transcriptome, we used a global gene expression study to examine growth conditions with maximal expression for each of the 10 DNA-binding proteins (Nicolas *et al.*, 2012). Samples were also collected from markerless deletions of Δ*mdxR*, Δ*ycxD*, Δ*yesS*, Δ*yhjH*, Δ*ykoM*, Δ*yvmB*, Δ*ywbI*, Δ*ywgB*, Δ*ywhA*, Δ*yxaD* strains grown to the maximum expression detected in either rich, LB media or during sporulation. RNA isolated and sequenced from cells expressing $P_{hy}$-*mdxR*, $P_{hy}$-*ycxD*, $P_{hy}$-*yesS*, $P_{hy}$-*yhjH*, $P_{hy}$-*ykoM*, $P_{hy}$-*yvmB*, $P_{hy}$-*ywbI*, $P_{hy}$-*ywgB*, $P_{hy}$-*ywhA*, $P_{hy}$-*yxaD*, was compared to that of a $P_{hy}$-empty control. In addition, RNA isolated and sequenced from the deletion strains grown under the appropriate growth condition were compared against a wild-type control. Expression of genes that decreased or increased by at least 2-fold compared to an empty vector control, and reciprocally, increased or decreased by at least 2-fold in the deletion versus wild type are listed in Table 2.6.

In total, we implicated five genes, *yhjH*, *ywgB*, *yxaD*, *yvmC*, and *ywhA* in the regulation of synonymous gene(s) in both the deletion strain and upon misexpression (Figure 2.9). No synonymous regulation was detected for five genes, *ywbI, mdxR, yesS, ykoM* and *ycxD*. Since transcriptional regulators often regulate expression of nearby genes and can be autoregulatory, we examined the genomic locus of the differentially expressed genes in relation to the putative DNA-binding proteins. This analysis identified three additional genes, *ywbI*, *mdxR*, and *yesS* that when misexpressed, led to increased expression of divergent genes or genes in the same operon, suggesting that YwbI, MdxR and YesS could be transcriptional activators (Figure 2.9). Two genes, *ycxD* and *ykoM*, did not appear to affect synonymous gene expression in both misexpression and deletion samples or the expression of nearby genes in either condition suggesting that YcxD and YkoM are not transcriptional regulators in the conditions we tested (Table 2.6).

*yhjH, ykoM, ywhA, yxaD,* and *yvmB* encode MarR-like DNA-binding proteins, a family of DNA-binding proteins that typically regulate divergently expressed genes which contribute to survival in hostile environments (Deochand & Grove, 2017). Consistent with this conserved role, YhjH, YxaD and YvmB appeared to regulate divergently expressed genes in both tested conditions. Based on expression data, it appears that YhjH represses *yhjG*, encoding a putative aromatic compound monooxygenase/hydrolase, YxaD represses *yxaKC* encoding Cid/Lrg homologs and YvmB represses *yvmC* and *cypX*, genes involved in pulcherriminic acid biosynthesis. The YvmB regulation we observed is consistent with a previous study by Randazzo and collogues (Randazzo *et al.*, 2016); notably, this study also demonstrated that YvmB

negatively regulates *yvmBA* and *yvnB* and positively regulates *yisI* (Randazzo *et al.*, 2016) which we did not observe. Although our results indicate that YvmB affects expression of *sunT*, *bdbA*, and *gcvTPA* in both the misexpression and deletion, these levels were lower than *yvmC* and *cypX* ($log_2$FC of ~2 vs $log_2$FC 3-6) possibly indicating that their activation is indirect or only occurring in a small subpopulation of cells. While most characterized MarR-like proteins act as transcriptional repressors (Deochand & Grove, 2017), *ywhA* appears to activate *glpFK*, encoding a glycerol permease and kinase, and appears to repress *ydaB,* a putative acyl-CoA ligase. Interestingly, even though *glpFK* was observed as being regulated in both conditions, we identified *glpF* activation in *yxaD,* and *glpD* activation in *ywbI* indicating a possible connection between activating glycerol utilization genes, possibly if *ywhA*, *yxaD* and *ywhI* act in similar pathways. In addition, *yxaD* misexpression was found to activate two genes in a nearby operon, *gntRK*, encoding a gluconate operon repressor and a gluconate kinase. Surprisingly, the MarR-like protein YkoM did not appear to directly affect transcription of any genes, suggesting DNA-binding serves a different function, possibly binding DNA to localize a function. It is also possible that the YkoM is autoregulatory, as this could not be inferred from the RNA-seq data.

YwbI belongs to the LysR-type transcriptional regulator (LTTR) family, a highly conserved and characterized family of proteins (Maddocks & Oyston, 2008), and in addition to *glpD*, appeared to activate *ywbHG*. Similar to *yxaKC*, *ywbHG* are annotated as *cid*/*lrg* like genes. Our finding that YwbI appears to activate *ywbHG* is consistent with a recent study that observed activation of *ywbHG* by YwbI in the presence of acetate (Chen *et al.*, 2015).

YesS belongs to the AraC/XylS family of proteins which are typically associated with regulated virulence genes and typically regulate transcription as activators (Santiago *et al.*, 2016, Tobes & Ramos, 2002). Our finding that YesS apparently activates *yesWXYZ*, *lplABCD* and *yetAF* is consistent with previous reports demonstrating YesS activation of these genes in *B. subtilis* (Poncet *et al.*, 2009).

YwgB is a Rrf2-type protein. Rrf2-types are a poorly characterized family of transcriptional regulators that often coordinate with Fe-S clusters (Shepard *et al.*, 2011). YwgB appeared to repress *mmr* encoding a putative toxic compound efflux transporter without experimental validation.

MdxR is part of the LacI family of regulators which have been shown to regulate genes involved in a variety of responses to changes in environment and metabolism (Meinhardt *et al.*, 2012). Consistent with this, we found that MdxR appears to repress the maltodextrin utilization genes encoded by *mdxDEFG* and *yvdJKLM* (Shim *et al.*, 2009).

Lastly, YcxD is part of the GntR-like family of DNA-binding proteins. GntR proteins are a large, well-studied class typically consisting of transcriptional regulators. Interestingly, however, YcxD is annotated as a PLP-dependent regulator, and part of the MocR subfamily of GntR. MocR proteins catalyze a reversible reaction with the help of binding pyridoxal 5'-phosphate (PLP) as a cofactor and transfer an amino group form an amino acid to an $\alpha$-keto acid (Rigali *et al.*, 2002). Consistent with being a member of the PLP-dependent GntR-protein family, we did not find any genes to be regulated by YcxD.

**Table 2.6 Transcriptional profiles of by DNA-binding proteins.**

| Gene | Highest Native Expression Condition | >2X IPTG induced/wt | >2X Δ/wt |
|---|---|---|---|
| *yhjH* | Exponential | - | *yhjG*‡ |
| *ywgB* | Exponential | - | *mmr*‡ |
| *ywbI* | Exponential | *ywbHG*§, *glpD*‡ | - |
| *yxaD* | Stationary | *glpF*‡, *gntRK*** | *yxaK*‡*C*§, *yxnA*§ |
| *yvmB* | Stationary | *sunT*‡, *bdbA*‡, *gcvTPA*‡ | *yvmC*‡, *cypX*‡, |
| *ycxD* | 1 hr into sporulation | - | - |
| *mdxR* | 1 hr into sporulation | *mdxDEFG**, *yvdJKLM** | - |
| *yesS* | 6 hr into sporulation | *yesWXYZ***, *lplABCD**, *yetAF** | - |
| *ywhA* | 6 hr into sporulation | *glpFK*‡ | *ydaB*‡ |
| *ykoM* | 6 hr into sporulation | - | - |

‡ Differential expression in both deletion and misexpression
§ Only detected in deletion
*Only detected in misexpression

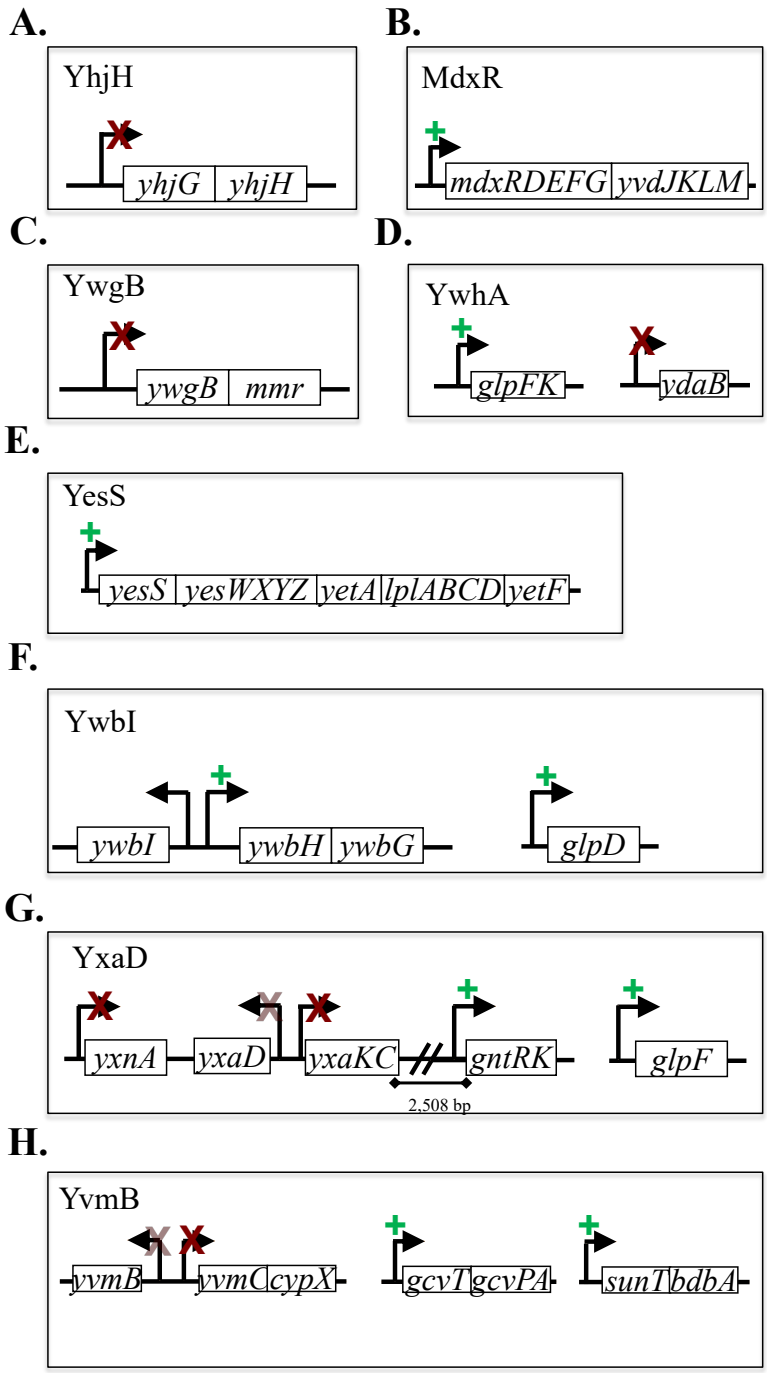



**Figure 2.8 Transcription factor activity of DNA-binding proteins by RNA-seq.** Regulated genes that are near or predicted to be in the same operon as the regulated gene are represented on one strand (line). Genes or operons that appear to be activated or repressed by the DNA-binding protein are represented as green "+" or a red "X", respectively. DNA-binding proteins that are predicted to autoregulate based on protein family are shown as faded symbols.

## 2.4. Discussion

The BEIGEL provides a conduit with which to link uncharacterized gene products with identified functions. Our approach uses artificial gene expression to perturb cells to produce functionally revealing phenotypes. We found that by expressing a gene in a non-native environment, we created a sensitized context more prone to tractable phenotypes. A key characteristic of our approach is that once phenotypes are observed, they can be exploited further to identify genetic targets for our ultimate goal of identifying and characterizing gene functions. The BEIGEL currently holds 810 strains that can be used to generate tractable hypotheses in order to identify gene function. So far, our lab has used this misexpression system to identify key targets and variants for both characterized and putative gene products: RefZ, SirA, YodL and YisK.

We set out to identify DNA-binding proteins that play roles in regulating subcellular organization. Out of 100 putative DNA-binding proteins, 30 had some form of growth inhibition on plates when misexpressed. Surprisingly, 87% of those that exhibited growth defects were found to have morphological phenotypes associated with DNA organization, cell division, or cell morphogenesis. Intriguingly, such growth phenotypes are not only advantageous in forming testable hypotheses or genetic targets but can be exploited to characterize the underlying mechanisms of the proteins as DNA-binders. For example, we found that the DNA-binding proteins RefZ, YxaD and YhjH required DNA-binding to exhibit a growth defect. By employing a selection process similar to that of the suppressor selection detailed early, we obtained mutations in the open reading frames of each gene that confer resistance to misexpression. So far we have identified variants that are unable to bind DNA, bind DNA with higher affinities,

recognize alternative or nonspecific DNA sequences, disrupt the dimerization interface, are dominant negative, and variants that predictably lose the ability to disrupt target proteins.

By using RNA-sequencing analysis we identified two out of 10 putative DNA-binding proteins, *ykoM* and *ycxD*, that do not regulate gene expression suggesting they have other responsibilities in the cell. We analyzed RNA isolated from strains overexpressing the gene of interest, and those harboring a deletion grown under a condition the gene is expressed and considered genomic loci. Although we feel confident that these gene products do not affect transcription, we acknowledge that YkoM and YcxD could be regulating genes under growth conditions we did not test, but we do not expect that to be the case. Four out of the 10 genes appeared to regulate transcription of the same genes under both conditions. Interestingly, we observed regulation of genes involved in glycerol utilization, *glp*, in the deletions of 5 genes, *yxaD*, *ywbI*, *ywhA*, *ywgB*, and *yhjH* suggesting that 1) regulation of *glp* genes may be indirect and/or 2) that these genes may be involved in regulated metabolic pathways. Only two genes, *mdxR* and *yesS* appeared to regulate expression in the misexpression strain only, but this regulation appeared direct since the regulated genes were in the same operon as *mdxR* and *yesS*. Furthermore, our observations were consistent with a previous study characterizing YesS as a transcriptional activator (Poncet *et al.*, 2009).

We found that that misexpression of 75/810 strains exhibited a growth phenotype and 49/75 resulted in morphological phenotypes associated with the DNA, cell division, and/or cell shape under just one growth condition (LB at 37 °C). We believe that perturbing other aspects of growth (i.e. temperature, pH, nutrient limitation) could

identify additional gene products that lead to growth or morphological phenotypes that we did not observe. Additionally, the BEIGEL could be grown in the presence of compounds, peptides, or other metabolites (for example, antibiotics) and screened for effects related to fitness. Lastly, since each strains is selected for by a spectinomycin resistance marker, the BEIGEL (in its entirety or as individuals stains) can be combined with the kanamycin and/or erythromycin resistant deletion libraries in *B. subtilis* (Koo *et al.*, 2017). Combinatory studies of both libraries would create additional sensitized backgrounds that can be screened for perturbations in growth or morphology. Thus, we see many applications for the BEIGEL that can aid in gene discovery and characterization.

# 3. CHARACTERIZATION OF *B. SUBTILIS* YXAD

## 3.1. Introduction

MarR proteins are a family of DNA-binding proteins that regulate expression of genes involved in diverse cellular processes (Deochand & Grove, 2017). The first MarR protein was discovered in *E. coli* and disruptions in *marR* (multiple antibiotic resistance regulator) resulted in increased resistance to many antibiotics (George & Levy, 1983), fluoroquinolones, oxidative stress and organic solvents (Alekshun & Levy, 1997). MarR-like proteins have since been characterized in other bacteria as having roles in regulating virulence factors, drug efflux pumps, antibiotic resistance genes, and metabolism (Deochand & Grove, 2017). MarR-like proteins regulate gene expression as transcriptional activators or, more commonly, repressors. MarR proteins typically regulate their own expression (autoregulatory) and one or more divergent genes by binding specific DNA sequences in gene promoters. The binding of MarR to DNA can be regulated by ligand binding or cysteine modification, usually resulting in a conformation change that releases the protein from DNA.

MarR-like proteins bind DNA as homodimers forming a 'safety triangle'-like shape (Alekshun *et al.*, 2001, Hong *et al.*, 2005, Kumaraswami *et al.*, 2009). Structurally characterized MarR proteins consist of six α-helices and three β-strands with the following topology: α1-α2-β1-α3-α4-β2-β3-α5-α6 (Figure 3.1AB) (Deochand & Grove, 2017). Contacts between the buried hydrophobic residues of helices α1, α5 and α6 (and sometimes α2) and α1', α5' and α6' of the dyadic subunit, form a hydrophobic core and are important for dimerization (Hong *et al.*, 2005, Kumaraswami *et al.*, 2009). Contacts

with the major groove are facilitated by α4, whereas a loop of approximately seven bp formed between β2 and β3 interacts with the minor groove (Hong *et al.*, 2005, Dolan *et al.*, 2011). The loop possesses a conserved DXR motif, with the conserved Arginine making contacts with the DNA minor groove (Hong *et al.*, 2005). The winged helix-turn-helix (wHTH) DNA-binding domain is made up of β1, α3, α4, β2 and β3 (Hong *et al.*, 2005). The overall positive surface potential observed on MarR structures is likely advantageous for DNA-binding (Hong *et al.*, 2005, Kim *et al.*, 2016). MarR-like proteins typically bind 12-18 bp motifs consisting of two inverted repeats separated by 2-5 bp (Wilkinson & Grove, 2006) in promoters regions. For MarR proteins that act as repressors, the motifs commonly overlap -10 promoter elements, positioning the MarR protein to repress transcription by blocking access of RNA polymerase to the promoter (Deochand & Grove, 2017). Although MarR-like proteins bind to diverse sequences, the crystal structure of *B. subtilis* OhrR bound to DNA reveals hydrogen bonding between the NH of R94 and the O2 of thymine (Hong *et al.*, 2005). Because of the nature of this

bond, the conserved Arginine in the loop is exclusive to pyrimidine recognition sequence at this position.

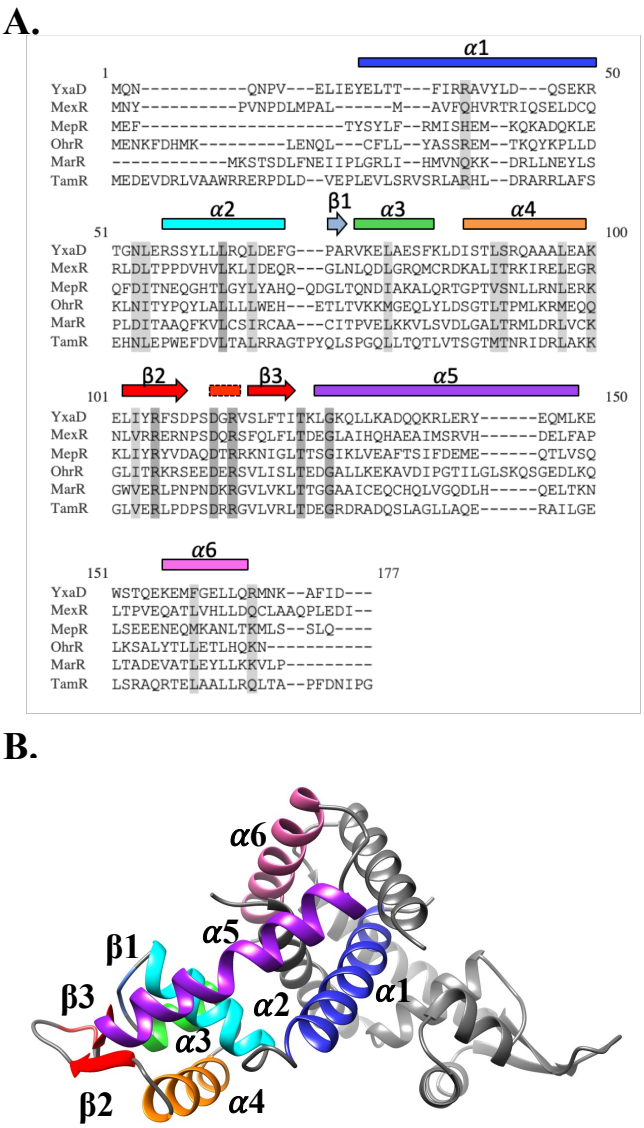



**Figure 3.1 MarR-like protein family.**
(A) Multiple sequence alignment of MarR-like proteins generated by Clustal Omega (Sievers *et al.*, 2011). Secondary structures are represented by colored boxes and arrows for *α* helices and β strands. (B) Predicted structure of YxaD generated from Itasser (Yang *et al.*, 2015, Roy *et al.*, 2010, Zhang, 2008). Secondary structures are separated by color and match the MSA in (A).

MarR DNA-binding activity is typically altered by oxidation or by binding small ligands (Deochand & Grove, 2017). MarR-like proteins have been shown to be allosterically regulated by phenolic ligands such as plant-derived aromatic compounds or antibiotics, carboxylic acids (Huang & Grove, 2013), proteins (Daigle *et al.*, 2007, Jerga & Rock, 2009, Grove, 2017) metals such as zinc (Reyes-Caballero *et al.*, 2010, Pagliai *et al.*, 2018) or by oxidation of at least one cysteine residue (Newberry *et al.*, 2007, Hong *et al.*, 2005). More recently, a pH dependent MarR-like regulator was found to change topology of promoter DNA in response to altered pH (Deochand *et al.*, 2016).

Mutations in *marR* genes have been discovered in drug resistant isolates of many bacteria. These *marR* mutations often result in a regulator that is unable to bind DNA and, therefore, unable to repress genes involved in cell survival in the presence of antibiotic stress (Jalal & Wretlind, 1998, Srikumar *et al.*, 1999, Oh *et al.*, 2003, Brooks *et al.*, 2007, Schindler *et al.*, 2013). Discerning how MarR-like proteins bind and recognize DNA sequences is critical in learning how they regulate virulence, oxidative stress, antibiotic resistance, and metabolic genes. In addition, characterizing the roles of other MarR-like proteins will likely contribute to our understanding on how bacteria are capable of surviving diverse and hostile environments.

The Gram-positive model organism *Bacillus subtilis* possesses 23 annotated MarR-like proteins, 16 of which have not been experimentally characterized. In the present work, we characterize the MarR-like protein YxaD. YxaD was identified in a misexpression screen to uncover factors involved in regulating subcellular organization in *B. subtilis*, we identified the MarR-like protein, YxaD. Here we describe the

phenotypes and possible causes of *yxaD* overexpression, YxaD's function as a transcriptional regulator, and the mechanism YxaD uses to bind DNA.

## 3.2. Materials and Methods

### 3.2.1. General Methods

All *B. subtilis* strains were derived from *B. subtilis* 168. Cloning, bacterial two-hybrids and protein purification were carried out in *E. coli* DH5α, *E.coli* DHP1, and *E.coli* BL21, respectively. All *B. subtilis* and *E. coli* strains used in this study are listed in Table 3.1. Plasmids are listed in Table 3.2. Oligonucleotide primers are listed in Table 3.3. All strains containing $P_{hy}$- misexpression constructs were streaked on LB plates containing 0.2% (w/v) glucose. For all growth experiments, strains were grown in 25 ml LB or 25 ml CH () in 250 ml baffled flasks at 37 °C shaking in a water bath set to 280 rpm. For transcriptional fusions experiments, strains were grown in LB supplemented with 0.3% (w/v) glucose. The following antibiotic concentrations were used for *E. coli*: chloramphenicol (25 µg/ml), kanamycin (25 µg/ml), ampicillin (100 µg/ml or when used in combination with other antibiotics, 50 µg/ml). For *B. subtilis*, antibiotics were used in the following concentrations: spectinomycin (100 µg/ml), chloramphenicol (7.5 µg/ml), kanamycin (10 µg/ml), phleomycin (0.8 µg/ml), and for erythromycin resistance, 1 µg/ml erythromycin (erm) and 25 µg/ml lincomycin. Unless otherwise indicated, isopropyl-β-D-thiogalactopyranoside (IPTG) was added to 1 mM and xylose was added to 1% (w/v).

**Table 3.1 Strains used in Chapter 3**

| Strain | Genotype | Reference |
|---|---|---|
| **Parental** | | |
| *B. subtilis* 168 | | |
| *E. coli* DH5α | F- endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG Φ80dlacZΔM15 Δ(lacZYA-argF)U169, hsdR17(rK- mK+), λ– | |
| *E. coli* DHP1 | *F-, cya-99, araD139, galE15, galK16, rpsL1 (Strr), hsdR2, mcrA1,mcrB1;* | Thomas Bernhardt |
| *E. coli* BL21 (DE3) Codon Plus RIL | *(F- ompT hsdS- dcm+ TetR gal λ(DE3) endA Hte [argU ileY leuW CamR])* | Stratagene |
| ***B. subtilis* 168** | | |
| BSH050 | *amyE::Phyperspank-yxaD (GTG start) (spec)* | This study |
| BSH52 | *amyE::Phyperspank-yxaD (GTG start) (spec), dnaX-yfp (phleo)* | This study |
| BSH43 | *ycgO::PftsW tetR-cfp (phleo), yycR::tetO48 (cat)* | This study |
| BSH53 | *ycgO::PftsW tetR-cfp (phleo) , amyE::Phyperspank-yxaD (spec), yycR::tetO48 (cat)* | This study |
| BSH144 | *PyxaD yxaD-gfp (spec)* | This study |
| BSH190 | *pSH041- PyxaD (erm)(cat), PyxaD yxaD-gfp (spec)* | This study |
| BSH189 | *pSH041 (erm)(cat)* | This study |
| BSH041 | *yxaD::erm* | This study |
| BSH094 | *Δ(soj-spo0J)::cat, yxaD::erm* | This study |
| BSH085 | *Δ(soj-spo0J)::cat* | This study |
| BSH089 | *amyE::Pxyl-scpA (spec)* | This study |
| BSH097 | *yycR::Phyperspank-yxaD (cat), amyE::Pxyl-scpA (spec)* | This study |
| BSH059 | *yycR::Phyperspank-yxaD (cat)* | This study |
| BSH107 | *scpA::erm (erm)* | This study |
| BSH143 | *scpA::erm (erm), amyE::Phyperspank-yxaD (spec)* | This study |
| BSH141 | *amyE::Phyperspank-yxaD R82A(spec)* | This study |
| BSH139 | *amyE::Phyperspank-yxaD R90A* | This study |
| BSH305 | *amyE::spoVG-lacZ (cat), ΔyxaD* | This study |
| BSH306 | *amyE::PyxaD-spoVG-lacZ (cat), ΔyxaD* | This study |
| BSH307 | *amyE::PyxaKC-spoVG-lacZ (cat), ΔyxaD* | This study |
| BSH522 | *yycR::Phy-yxaD (gtg start) (kanR) , amyE::PyxaKC-spoVG-lacZ (cat), ΔyxaD* | This study |
| BSH523 | *yycR::Phy-yxaD* (kanR) R90A , amyE::PyxaKC-spoVG-lacZ (cat), ΔyxaD* | This study |
| BSH524 | *yycR::Phy-yxaD* (kanR) D88A , amyE::PyxaKC-spoVG-lacZ (cat), ΔyxaD* | This study |
| BSH525 | *yycR::Phy-yxaD* (kanR) R82A , amyE::PyxaKC-spoVG-lacZ (cat), ΔyxaD* | This study |

**Table 3.1 Continued**

| Strain | Genotype | Reference |
|---|---|---|
| BSH526 | *yycR::Phy-yxaD* (kanR) K29E , amyE::PyxaKC-spoVG-lacZ (cat), ΔyxaD* | This study |
| BSH527 | *yycR::Phy-yxaD* (kanR) G49E , amyE::PyxaKC-spoVG-lacZ (cat), ΔyxaD* | This study |
| BSH528 | *yycR::Phy-yxaD* (kanR) G49R , amyE::PyxaKC-spoVG-lacZ (cat), ΔyxaD* | This study |
| BSH529 | *yycR::Phy-yxaD* (kanR) R52C , amyE::PyxaKC-spoVG-lacZ (cat), ΔyxaD* | This study |
| BSH530 | *yycR::Phy-yxaD* (kanR) R52L , amyE::PyxaKC-spoVG-lacZ (cat), ΔyxaD* | This study |
| BSH531 | *yycR::Phy-yxaD* (kanR) K54Q , amyE::PyxaKC-spoVG-lacZ (cat), ΔyxaD* | This study |
| BSH532 | *yycR::Phy-yxaD* (kanR) L79P , amyE::PyxaKC-spoVG-lacZ (cat), ΔyxaD* | This study |
| BSH533 | *yycR::Phy-yxaD* (kanR) D85E , amyE::PyxaKC-spoVG-lacZ (cat), ΔyxaD* | This study |
| BSH534 | *yycR::Phy-yxaD* (kanR) G89E , amyE::PyxaKC-spoVG-lacZ (cat), ΔyxaD* | This study |
| BSH535 | *yycR::Phy-yxaD* (kanR) R90G , amyE::PyxaKC-spoVG-lacZ (cat), ΔyxaD* | This study |
| BSH536 | *yycR::Phy-yxaD* (kanR) R90K , amyE::PyxaKC-spoVG-lacZ (cat), ΔyxaD* | This study |
| BSH537 | *yycR::Phy-yxaD* (kanR) L103P, amyE::PyxaKC-spoVG-lacZ (cat), ΔyxaD* | This study |
| BSH538 | *yycR::Phy-yxaD* (kanR) L133M, amyE::PyxaKC-spoVG-lacZ (cat), ΔyxaD* | This study |
| BSH539 | *yycR::Phy-yxaD* (kanR) N138 , amyE::PyxaKC-spoVG-lacZ (cat), ΔyxaD* | This study |
| BSH540 | *yycR::Phy-yxaD* (kanR) N138Y, amyE::PyxaKC-spoVG-lacZ (cat), ΔyxaD* | This study |
| BSH290 | *amyE::spoVG-lacZ (cat)* | This study |
| BSH291 | *amyE::PyxaD-spoVG-lacZ (cat)* | This study |
| BSH292 | *amyE::PyxaKC-spoVG-lacZ (cat)* | This study |
| ***E. coli* DHP1** | | |
| cSH025 | *yxaD*-T18 + *scpA*-T25 | This study |
| cSH027 | *yxaD*-T18 + empty-T25 | This study |
| cSH028 | empty-T18 + *scpA*-T25 | This study |

**Table 3.2 Plasmids used in Chapter 3**

| Plasmid | Description | Reference |
|---|---|---|
| pSH041 | *PyxaD* (erm) (cat) | This study |
| pSH052 | 6His-*yxaD* (kan) | This study |

**Table 3.3 Oligonucleotides used in Chapter 3**

| Oligo | Sequence 5'-3' |
|---|---|
| OSH108 | AAAGTCGTCAGCTCATATTCAATTA |
| OSH109 | CATATAACGACTTGTATTTATTCAGTTAATGT |
| OSH110 | CCAATCTCCGGCATTGAC |
| OSH111 | GGTTCTGAGGAATCGCTTTACTT |
| OSH112 | AATAAATACAAGTCGTTATATGACTAAATCAA |
| OSH113 | ATCAAGTGCTGCTCTCCAATC |

### 3.2.2. Microscopy

For microscopy experiments in LB, all samples were grown in LB media for 2-3 hours to mid-exponential, then back-diluted to $OD_{600}$ = 0.00625 in 25 ml LB. For growth in CH, samples were grown overnight in CH media to mid-exponential, then back-diluted to $OD_{600}$=0.00625 in 25 ml CH. All cultures were grown at 37 °C shaking in a water bath set to 280 rpm. When indicated, cells were grown to $OD_{600}$ = 0.05, and induced with 1 mM IPTG. Time refers to time post induction. Fluorescence microscopy was performed with a Nikon Ti-E microscope equipped with a CFI Plan Apo lambda DM 100× objective, Prior Scientific Lumen 200 illumination system, C-FL UV-2E/C 4′,6-diamidino-2-phenylindole, C-FL green fluorescent protein (GFP) HC HISN Zero Shift and C-FL YFP HC HISN Zero Shift filter cubes, and a CoolSNAP HQ2 monochrome camera. All images were captured with NIS Elements Advanced Research (version 4.10), and processed with NIS Elements and ImageJ64 (W., 1997-2015). For image capture, 1 ml of cells were pellet at 6,010 x g for 1 min in a tabletop microfuge at room temperature, supernatants were removed by aspiration, and pellets were

resuspended in ~7 µl 1X PBS containing DAPI DNA stain (2 µg/ml) (Molecular Probes) and FM4-64 membrane stain (3 µg/ml)(Molecular Probes). Cells were mounted on glass slides with polylysine-treated coverslips.

### 3.2.3. Bacterial two-hybrid

Bacterial two-hybrids were performed as previously described (9576956), except cloning was performed in the presence of 0.2% glucose (w/v) in addition to antibiotics. *E. coli* DHP1 harboring the applicable pT25 and pT18 interaction plasmids were grown in LB containing ampicillin (50 µg/ml) and kanamycin (25 µg/ml) to mid-exponential phase and normalized by $OD_{600}$. Five µl normalized cells were spotted on M9-glucose minimal plates with 250 µM isopropyl-β-D-thiogalactopyranoside (IPTG), 40 µg/ml 5-bromo-4-chloro-3-indolyl-β- D-galactopyranoside (X-gal), ampicillin (50 µg/ml), and kanamycin (25 µg/ml). Plates were incubated at room temperature for 72 hours before imaged.

### 3.2.4. Screen for *yxaD* variants

Single colonies of BSH310 were used to inoculate 5 ml LB independently. Three independent cultures were grown for each strain. Following 5 hr growth at 37 °C, cultures were spun at 6,010 x g for 2 min. Supernatants were removed by aspiration and resulting pellets were resuspended with 150 µl 1X PBS. 1:1, 1:2 and 1:5 dilutions of each resuspension were made in 1X PBS. Fifty µl of each dilution was plated on LB plates containing 100 µg/ml X-gal and either 50 µM, 250 µM, or 500 µM IPTG. Plates were incubated at room temperature until colony growth was observed and color was detected (~3 days). Only white colonies were selected for further analysis; this screen

eliminated YxaD variants unable to bind DNA and LacI mutants unable to derepress *yxaD* expression. Fifty individual colonies obtained from the selection were streaked for isolation on LB plates supplemented with 0.2% glucose (w/v). To select for dominant negative *yxaD* variants, single colonies of BSH082 or BSH098 were used to inoculate independent 5 ml LB cultures. Four independent cultures of each strain were then grown for 5 hr at 37 °C. Cells were treated as before and resuspended in 150 µl 1 XPBS, but no further dilutions were made after resuspension. Fifty µl were plated on LB plates containing 50 µM IPTG. Individual colonies were streaked for isolation and patched on LB plates containing 40 µg/ml X-gal and 50 µM IPTG to check for *lacI* mutants. Only blue colonies were proceeded with as these passed the screen for functional LacI.

### 3.2.5. *In vivo* YxaD DNA-binding assay

The *yxaD* gene from each suppressor (or variant from site-directed mutagenesis) was PCR amplified with oSH037 and oSH038. Enzymatic assembly was used to ligate each *yxaD* variant (*yxaD**) PCR product to a kanamycin resistant marker, $P_{hy}$-promoter, and region of homology to *yycR* for integration into the *B. subtilis* chromosome. Genomic DNA of the transformants was transformed into the DNA-binding reporter strain, BSH307, to analyze DNA-binding ability. Single colonies were used to inoculate 5 ml LB + 0.2% glucose (w/v). Cultures were grown at 37 °C until $OD_{600}$ reached 0.25-0.6 and were then normalized to a final $OD_{600}$ = 0.25. Two µl of normalized cultures were spotted on plates containing kanamycin, 40 µg/ml X-gal and either 50 µM, 250 µM, 500 µM, or 1.0 mM IPTG. Plates were left at 30 °C until blue color developed, ~48 hours.

### 3.2.6. β-galactosidase assays

Strains for transcriptional fusion assays were generated by cloning the intergenic region between *yxaD* and *yxaKC* from both directions into pDG1661 creating a *lacZ* fusion. The construct was then integrated into the non-essential locus, *amyE*, of *B. subtilis*. Three colonies were inoculated into 3 independent 5 ml cultures of LB supplemented with 0.3% glucose (w/v). Cultures were grown to mid-exponential and back-diluted to $OD_{600}$=0.00625 in 25 ml LB with 0.3% glucose (w/v) in 250 ml baffled flasks at 37 °C shaking in a water bath set to 280 rpm. OD readings and samples for β-galactosidase assay were taken every hour. β-galactosidase assays were performed as described previously (Ababneh & Herman, 2015), except 600 µl was collected for each sample.

### 3.2.7. Purification of YxaD

To obtain purified 6His-YxaD, *E.coli* BL21(λDE3) Codon Plus RIL cells were transformed with pSH038. Freshly transformed cells were inoculated in 25 mL Teknova Cinnabar High-Yield Protein Expression Media with 0.4% glucose (w/v) to a starting OD of 0.1. Cultures were grown for 28 hours at 30 °C at 280 rpm then centrifuged for 10 minutes at 9,639 x g at 4 °C. Pellets were resuspended in 30 mL of lysis buffer containing 50mM Phosphate, pH 8.0, 300 mM KCl, 20 mM Imidazole, 20% sucrose, 5 mM β-Mercaptoethanol, 1 mM protease inhibitor, and 100 U/µl lysozyme. Harvested cells were spun at 112,000x g for 45 min to remove cell debris. Cell supernatant was loaded onto 1 mL of Ni-NTA Agarose (Qiagen). Column was washed with 10 mL of wash buffer (50mM Phosphate, pH 8.0, 300 mM KCl, 20 mM Imidazole, 20% sucrose).

6His-YxaD was eluted with increasing imidazole concentrations (50 mM Phosphate, pH 8.0, 300 mM KCl, and either 100 mM, 200 mM, 250 mM, or 500 mM Imidazole). Expression and purification of 6His-YxaD were checked by running samples on a 12% SDS-PAGE gel stained with Colloidal Coomassie Blue Stain. Fractions showing a single band representing 6His-YxaD were pooled and dialyzed three times into 300 mL of 50 mM TrisHCl pH 8.0, 300 mM KCl 0.5 mM EDTA, 1 mM DTT and 15% glycerol. Aliquots of purified protein were kept at -80 °C until needed.

### 3.2.8. Electrophoretic mobility shift assays (EMSA)

Putative YxaD-binding sites were identified by entering the intragenic sequence of *yxaD* and *yxaKC* into the palindrome finder EMBOSS Explorer. DNA probes were generated containing either binding site or both sites using the following primer pairs: Binding Site I (BS-I), OSH108 and OSH109; Binding Site II (BS-II), OSH112 and OSH113; Binding Sites I and II (BS-I+II), OSH110 and OSH111. PCR products were purified then subsequently quantified using the Quant-iT PicoGreen dsDNA Assay Kit (ThermoFisher). Prior to reactions, 6His-YxaD aliquots were dialyzed against 300 ml of 50 mM Tris-HCl pH 8.0, 300 mM KCl 0.5 mM EDTA, 1 mM DTT to remove residual glycerol. Binding reactions were performed with 0.1 ng of DNA, 5x Reaction Buffer (100 mM TrisHCl, pH 8.0, 100 mM KCl, and 5 mM EDTA), and varying concentrations of purified 6His-YxaD to a final volume of 10 µl. Once mixed, reactions were incubated at room temperature for 15 min. Following the incubation period, 2 µl of loading reagent (20 mM TrisHCl, pH 8.0 and 15% Ficoll 400) was added to each reaction. Glycerol or dye was not added as either appeared to abolish the protein-DNA interaction. Ten µl of

each reaction were loaded onto 5% native polyacrylamide gels containing 0.5X TBE pre-run at 120 V for 1 hr. Gels were run at 120 V for 30 min then stained with SYBR Green I (ThermoFisher).

## 3.3. Results

### 3.3.1. YxaD misexpression leads to a defect in chromosome segregation and DNA replication

In a screen to discover *B. subtilis* genes involved in DNA replication and subcellular organization, we identified *yxaD*, encoding a putative MarR-like DNA-binding protein. Fluorescence microscopy on cells overexpressing *yxaD* revealed a loss in their bilobed nucleoid structure after 1.5-2 doublings (Figure 3.2A). After 90 min in LB media, most of the chromosomal DNA was either present at midcell or guillotined by cell division. DNA staining suggested there was no qualitative increase in DNA accumulation associated with each DNA mass, suggesting that YxaD misexpression may also lead to a defect in DNA replication. To assess if replisomes were assembled in the cells, we utilized a strain containing the DNA polymerase clamp loader subunit, DnaX, fused to YFP and imaged cells using fluorescence microscopy. DnaX-YFP foci are only observed when replisomes are formed (Lemon & Grossman, 1998, Lemon & Grossman, 2000). Wild type cells growing in LB medium exhibit DnaX-YFP foci throughout the cell, consistent with prior observations (Lemon & Grossman, 1998) (Figure 3.2B). Upon *yxaD* misexpression, the DnaX-YFP foci appeared similar to wildtype after 30 min (Fig 3.2B), when changes in the nucleoid by DAPI staining are already evident. At later

timepoints, the DnaX-YFP foci appear less spread out and brighter relative to wild type, possibly due to pleotropic effects (Figure 3.2B).

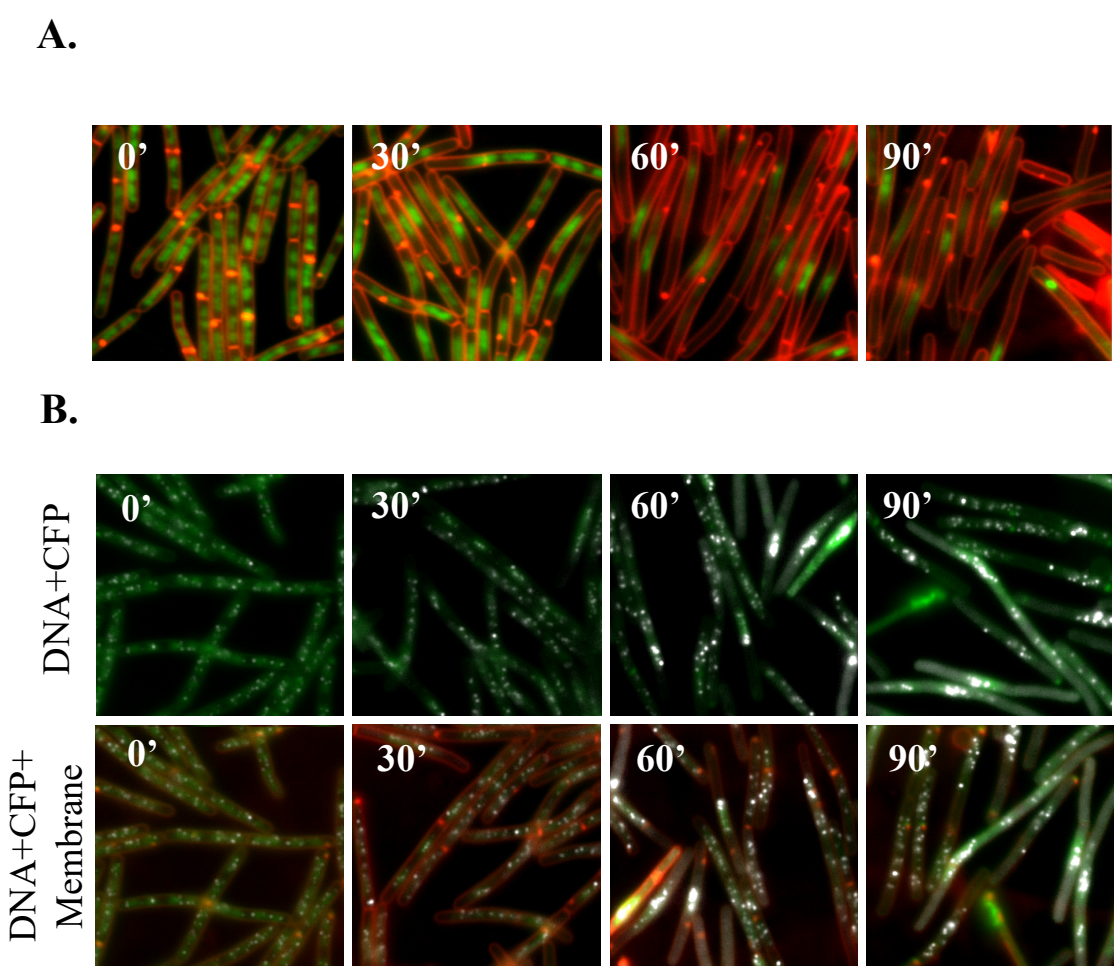



**Figure 3.2 *yxaD* misexpression phenotypes.**
Cells harboring one copy of (A) $P_{hy}$-*yxaD* (BSH050) or (B) $P_{hy}$-*yxaD* and DnaX-YFP (BSH052). Cells were grown in LB media (left) until $OD_{600}$ reached 0.05 when IPTG was added followed by images taken at 30, 60, and 90 min. Cell membranes were stained with FM4-64 (red), DNA was stained with DAPI (pseudocolored green), and DnaX-YFP foci are falsely colored white.

Cells misexpressing *yxaD* exhibited a clear defect in segregation, but continued to divide, leading to the accumulation of anucleate cells. This observation prompted us

to ask whether DNA replication origins were segregated after DNA replication. During growth in LB medium, *B. subtilis* cells double every 18 min and typically possess two to eight origins of replication at one time. To more easily visualize individual origins, we utilized CH medium (Sterlini & Mandelstam, 1969) in which the doubling time is 30 min and there are fewer origins per cell. Importantly, *yxaD* misexpression in CH results in phenotypes similar to that observed in LB (Figure 2.3A).

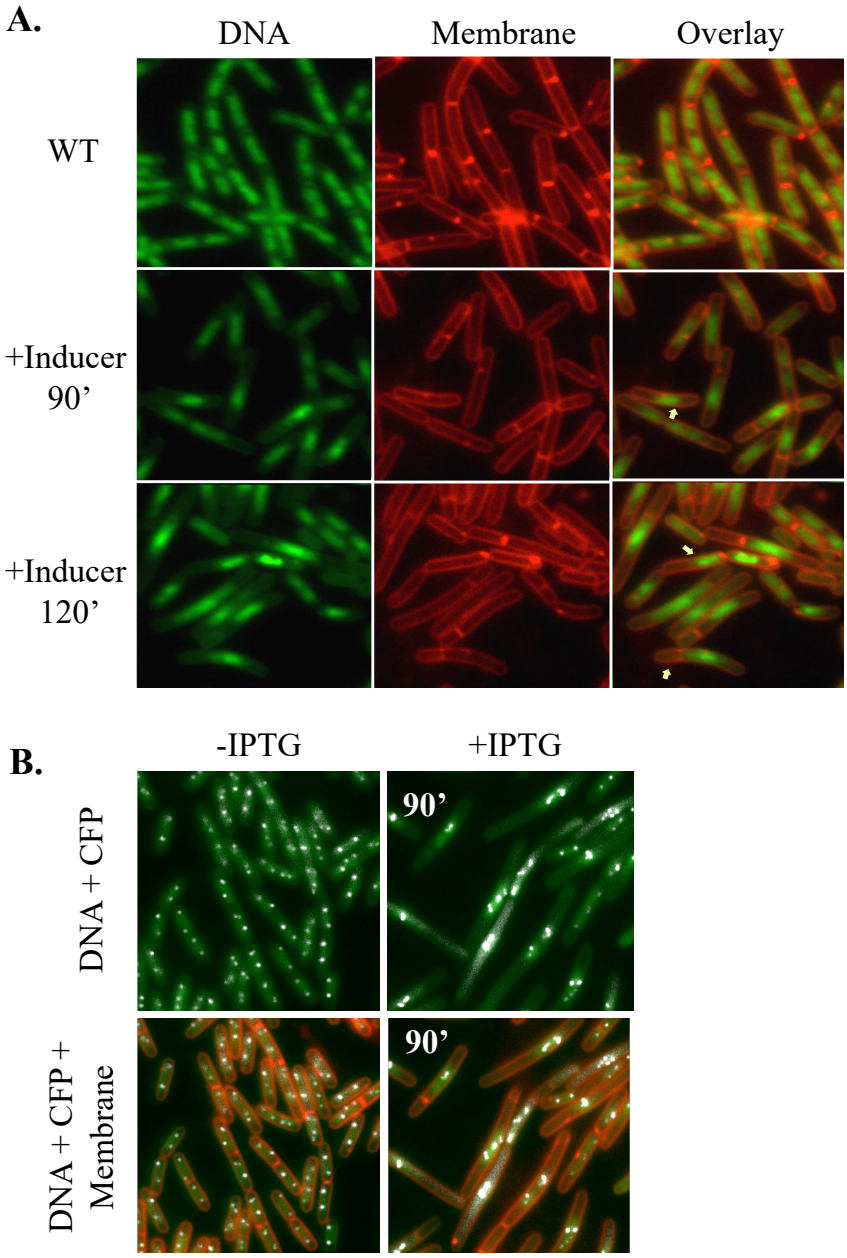



**Figure 3.3 *yxaD* misexpression phenotypes in CH media.**
(A) *Bs* 168 cells harboring one copy of $P_{hy}$-*yxaD* (BSH050) grown in CH liquid media for 1.5 hr and 2 hr. Yellow arrows indicate anucleate cells or cells containing bulk of nucleoid at midcell. (B) Cells grown in CH expressing *tetR-cfp* and *tetO* arrays at -7° only (TOP) or with $P_{hy}$-*yxaD* (BOTTOM). 1 ng/µl aTC was added to inhibit TetR bound tetO arrays from forming roadblock (20807205). Prior to imaging, collected samples were washed twice with 1X PBS to remove aTC. TetR-CFP foci are falsely colored white. Membranes are stained with FM4-64 (red) and DNA is stained with DAPI (psuedocolored green).

To monitor origins of replication, an array of 48 TetR operator sites (*tetO48*) was integrated near *oriC* (at 353°) in cells constitutively expressing TetR-CFP. In this assay, one TetR-CFP focus is observed for each resolved *oriC*, thus providing a proxy for the number of chromosomes associated with each nucleoid mass. Using this assay, we observed that cells expressing *yxaD* for 90 min exhibited similar numbers of foci (2-6 per cell) compared to wild-type cells, although the foci were qualitatively brighter (Figure 3.3B). In contrast to wild type, the YxaD-expressing strain possessed fewer nucleoid masses after 90 min of *yxaD* expression (~3 cell doublings in CH). At the same time, there was no qualitative increase in signal from stained DNA. These results suggest that artificial expression of YxaD results in both replication and segregation defects.

Next, we examined the localization of YxaD-GFP fusion (native promoter) and observed YxaD-GFP in a similar pattern to that of *oriC*. This observation is consistent with YxaD binding to the *yxaD* and *yxaKC* promoters at 351° to regulate transcription (Figure 3.4AB) (Chapter 2). Introducing a plasmid harboring a copy of the *yxaD/KC* intergenic region onto a self-replicating medium copy plasmid in a *yxaD-gfp* expressing strain resulted in the presence of additional foci (Fig 3.4C). The signal of YxaD-GFP was also brighter, which would be expected if YxaD were titrated away from repressing it's own promoter.

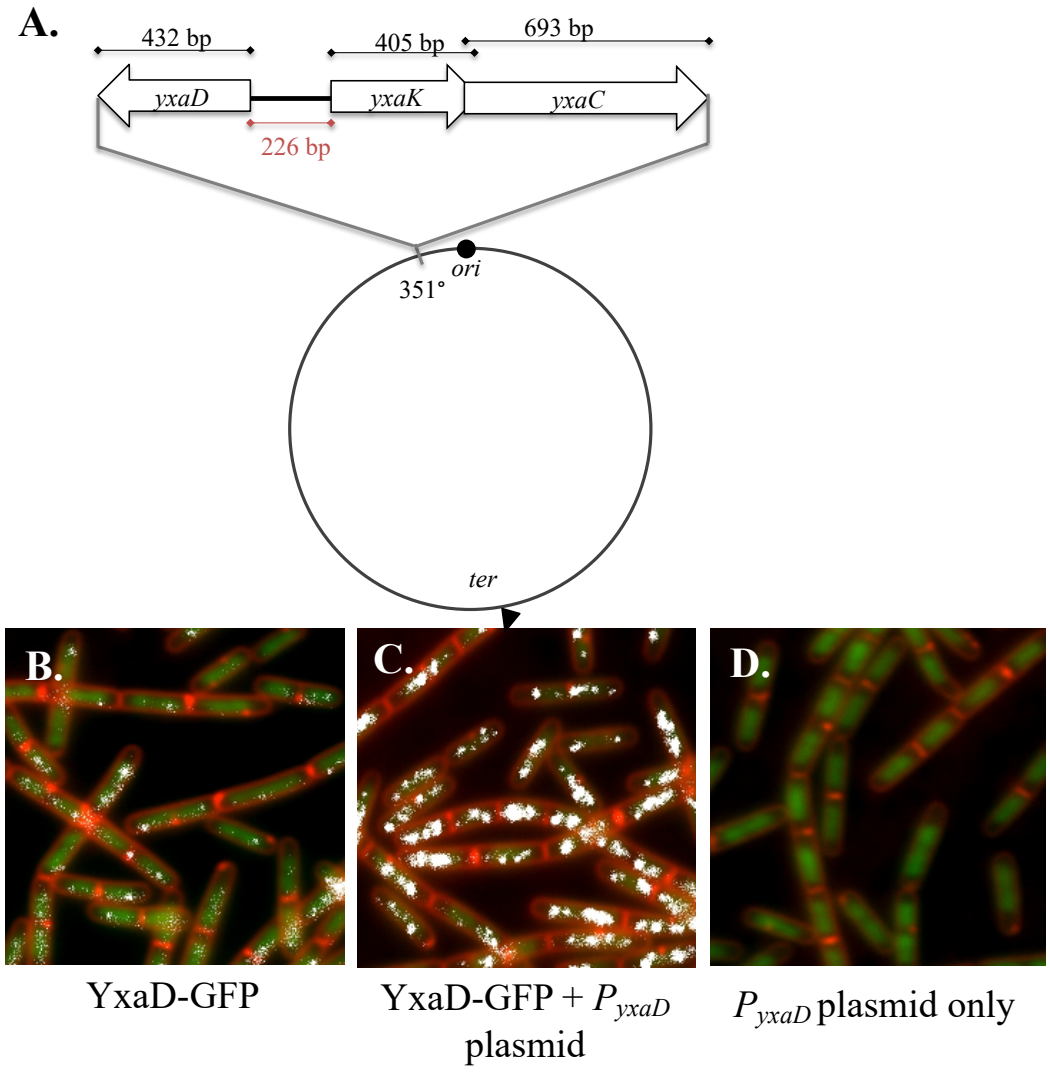



**Figure 3.4 YxaD-GFP localization.**
(A) Map of *yxaD* and *yxaKC* with respect to *oriC*. Intergenic region used in pSH041 (C and D) shown in red. Cells harboring (B) YxaD-GFP only (BSH144), (C) YxaD-GFP and pSH041 (BSH190) or (D) pSH041 only (BSH189). Membranes are stained with FM4-64 (red), DNA is stained with DAPI (pseudocolored green) and YxaD-GFP foci are falsely colored white.

### 3.3.2. Potential targets of YxaD

To determine how *yxaD* misexpression leads to a defect in chromosome segregation, DNA replication and elongation, we sought to identify genetic targets

associated with the phenotypes of YxaD misexpression. A search of the IntAct Molecular Interaction Database (Hermjakob *et al.*, 2004, Kerrien *et al.*, 2012) revealed yeast two-hybrid (Y2H) interactions between YxaD and three *B. subtilis* proteins: ScpA, a subunit of the Structural Maintenance of Chromosome (SMC) condensin system, HolA, the delta subunit of DNA polymerase, and TkmA, modulator of PtkA protein tyrosine kinase activity (Marchadier *et al.*, 2011, Shi *et al.*, 2014). Since ScpA and HolA have known roles in chromosome segregation and DNA replication, respectively, we attempted to recapitulate the YxaD-ScpA and YxaD-HolA interactions by bacterial two-hybrid (B2H). However, we were only successful in recapitulating the interaction between YxaD and ScpA (Figure 3.5A).

ScpA is a component of the SMC complex, or condensin, and is responsible for proper segregation of chromosomes (Gruber *et al.*, 2014). SMC recruitment to the origin-proximal region of the chromosome is partially mediated by an interaction with the DNA-binding protein ParB (Gruber & Errington, 2009), yet *parB* is not required for growth. Therefore, we explored the possibility that *yxaD* and *parB* had redundant roles in recruiting ScpA and/or SMC complexes to origin-proximal DNA. To determine if ParB and YxaD were redundant, we created a Δ*parB* Δ*yxaD* double mutant. If *parB* and *yxaD* were redundant and required for SMC recruitment, we would expect a Δ*parB* Δ*yxaD* strain to have similar or the same growth defects as a strain lacking SMC. Since the Δ*parB* Δ*yxaD* strain construction was successful and grew under conditions that require SMC (growth at 37 °C), we deduced that their functions are not redundant (Figure 3.5B). In order to further probe the potential interaction between YxaD and ScpA, we asked whether artificial expression of *scpA* could rescue *yxaD* misexpression.

Therefore, we placed *scpA* under a xylose-inducible promoter at a nonessential locus (*amyE*) and introduced this construct in *yxaD* overexpressing cells. Overexpression of *scpA* did not detectably rescue cells from the effects of *yxaD* misexpression (Figure 3.5C). However, this observation could be expected if YxaD only targets ScpA when in complex with SMC. Accordingly, the amount of "targetable ScpA" (i.e. in SMC complexes) might not be altered when *scpA* is overexpressed if ScpB and Smc are present at same levels as the uninduced sample. Therefore, to determine if ScpA was required for YxaD's misexpression phenotype, we introduced Δ*scpA* into the *yxaD* overexpression strain. When grown at permissive temperature (25 °C), killing was observed, demonstrating that *scpA* is not required for the *yxaD* misexpression phenotype (Figure 3.5D). These data suggest that whether or not YxaD targets ScpA, there exists at least one additional target for YxaD. In support of this idea, we attempted to obtain mutants resistant to *yxaD* misexpression by preforming a suppressor selection, but only obtained mutants in the expression construct (LacI dominant) (see below).

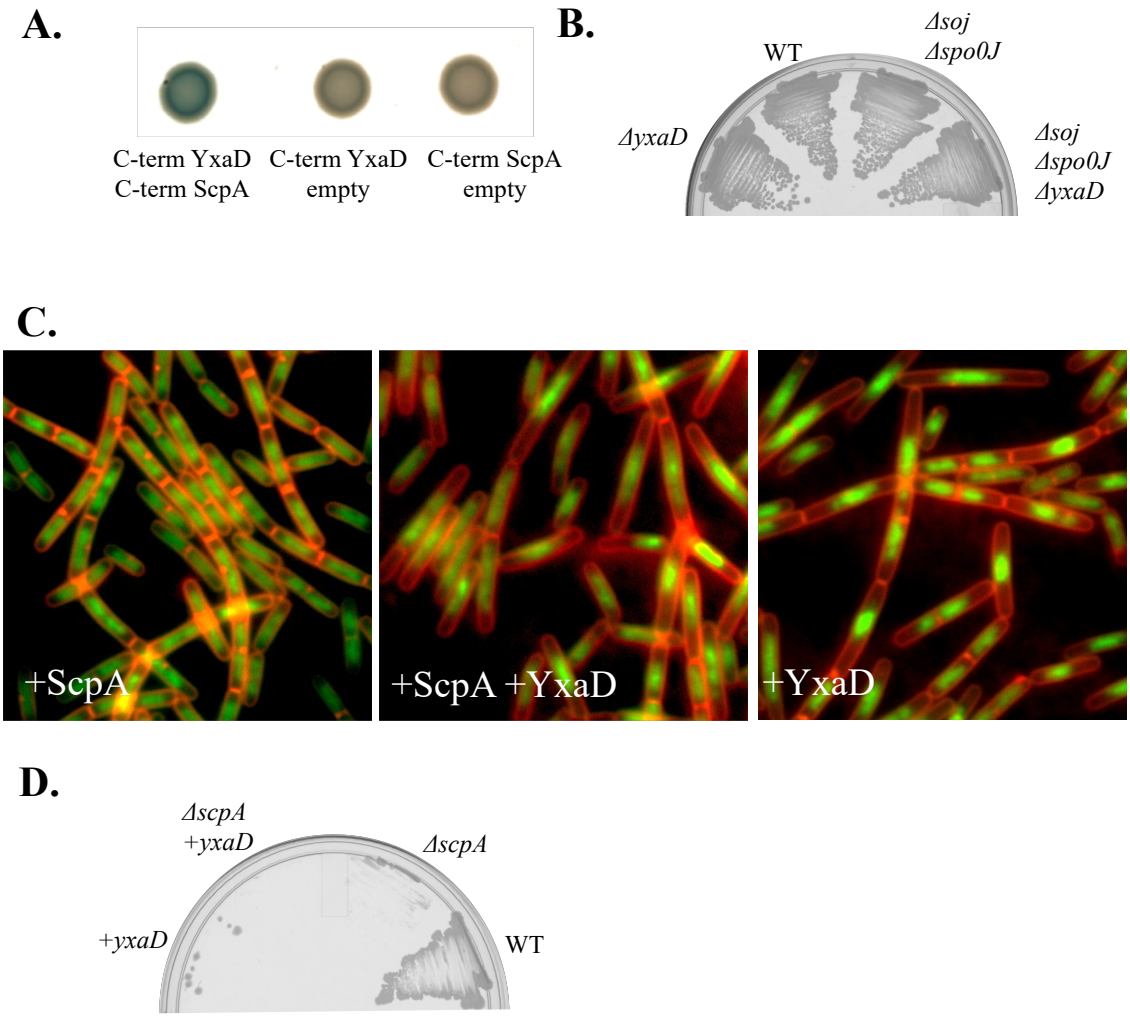



**Figure 3.5 ScpA is not required for phenotypes associated with artificial induction of YxaD.**
(A) B2H of YxaD-T18 and ScpA-T25 (cSH025), YxaD-T18 and empty-T25 (cSH027), and empty-T18 and ScpA-T25 (cSH028). (B) Single colonies of *Bs* 168 or strains harboring Δ*yxaD* (BSH041), Δ*soj* Δ*spo0J* Δ*yxaD* (BSH094), or Δ*soj* Δ*spo0J* (BSH085) were streaked on LB agar plates. (C) Fluorescence microscopy images of cells overexpressing ScpA only (BSH089), ScpA and YxaD (BSH097), or YxaD only (BSH059). Membranes are stained with FM4-64 (red) and DNA is stained with DAPI (pseudocolored green). (D) Single colonies of Bs 168, $P_{hy}$-*yxaD* (BSH050), Δ*scpA* (BSH107), or $P_{hy}$-*yxaD* and Δ*scpA* (BSH143).

### 3.3.3. DNA-binding is required for YxaD misexpression phenotype

Since YxaD belongs to the MarR family and appears to bind DNA, we wondered whether DNA-binding was required for the morphological defect observed from *yxaD* misexpression. DNA-binding in MarR-like proteins is facilitated by the conserved wHTH motif composed of β1, α3, α4, β2 and β3 (Deochand & Grove, 2017). The loop formed between β2 and β3 contains a conserved DXR motif, and the arginine in DXR is required for DNA-binding (Alekshun *et al.*, 2001, Saito *et al.*, 2003). To identify the DXR motif and to distinguish any additional residues that might be important for DNA-binding, we created a multiple sequence alignment (MSA) of MarR-like proteins and generated a predicted structure of YxaD using ITASSER (Yang *et al.*, 2015, Roy *et al.*, 2010, Zhang, 2008)(Figure 3.1AB). Using the MSA and predicted structure, we identified R90 as the arginine residue and D88 as the aspartate residue from the DXR motif, as well as R82, which all appear to be conserved in other MarR-like proteins and possibly located on β2. Point mutations were made in the *yxaD* misexpression construct for the overexpression of YxaD R82A or R90A by site-directed mutagenesis. Overexpression of either R82A or R90A variant did not lead to a growth or morphological defect (Figure 3.6B), suggesting that DNA-binding is required for YxaD's misexpression phenotype. In addition, we mutated the aspartate residue in DXR to alanine, D88A, but misexpression of this variant still killed cells in the presence of 1.0 mM IPTG. The observation that D88A still reduced growth was surprising considering D88 is located in the loop that makes contacts with DNA and is conserved in MarR-like proteins.

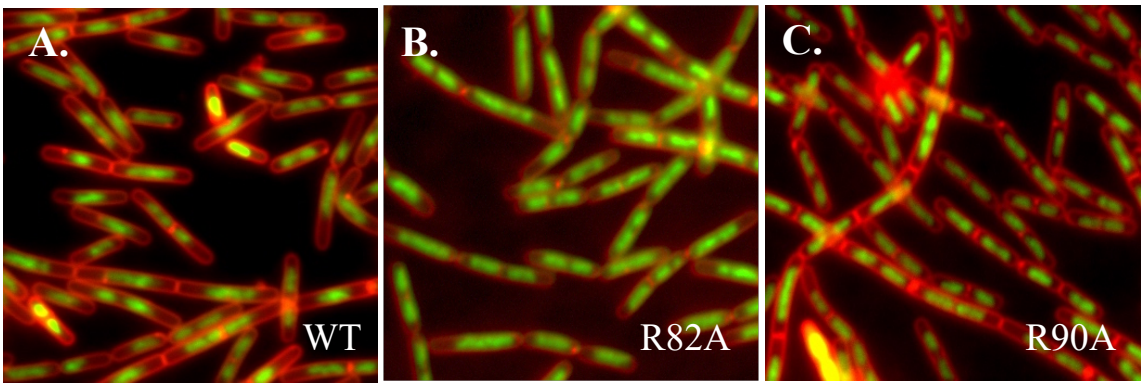



**Figure 3.6 DNA-binding is required for the YxaD misexpression phenotype.** Fluorescence microscopy of cells grown in CH media harboring one copy of (A) $P_{hy}$-*yxaD* (BSH050), (B) $P_{hy}$-*yxaD* R82A (BSH141), or (C) $P_{hy}$-*yxaD* R90A (BSH139). Images were taken 90 min after induction with IPTG. Membranes are stained with FM4-64 (red) and DNA is stained with DAPI (pseudocolored green).

### 3.3.4. YxaD suppressor mutants

We next asked whether we could obtain YxaD variants that are able to bind DNA that no longer exhibited a growth defect upon misexpression. We hypothesized that this class of variants would shed more light on the mechanism of DNA-binding by MarR-like proteins and more specifically, on the YxaD misexpression phenotype. To identify such variants, we created a loss-of-function (LOF) selection that also screens for YxaD binding to DNA using a *lacZ* reporter (Figure 3.7A). Our original findings indicated that YxaD was likely repressing *yxaKC* by binding to the intergenic region between *yxaD* and *yxaKC* (Figure 3.4C). In order to determine if YxaD represses *yxaKC* on agar plates, we constructed a $P_{yxaKC}$-*lacZ* reporter strain in wild-type and Δ*yxaD* backgrounds and spotted each on LB plates containing X-gal. In the presence of wild-type *yxaD*, colonies appeared white (Figure 3.7B). In contrast, Δ*yxaD* colonies turned blue, consistent with the RNA-seq data suggesting YxaD represses *yxaKC*. To select for mutants that were

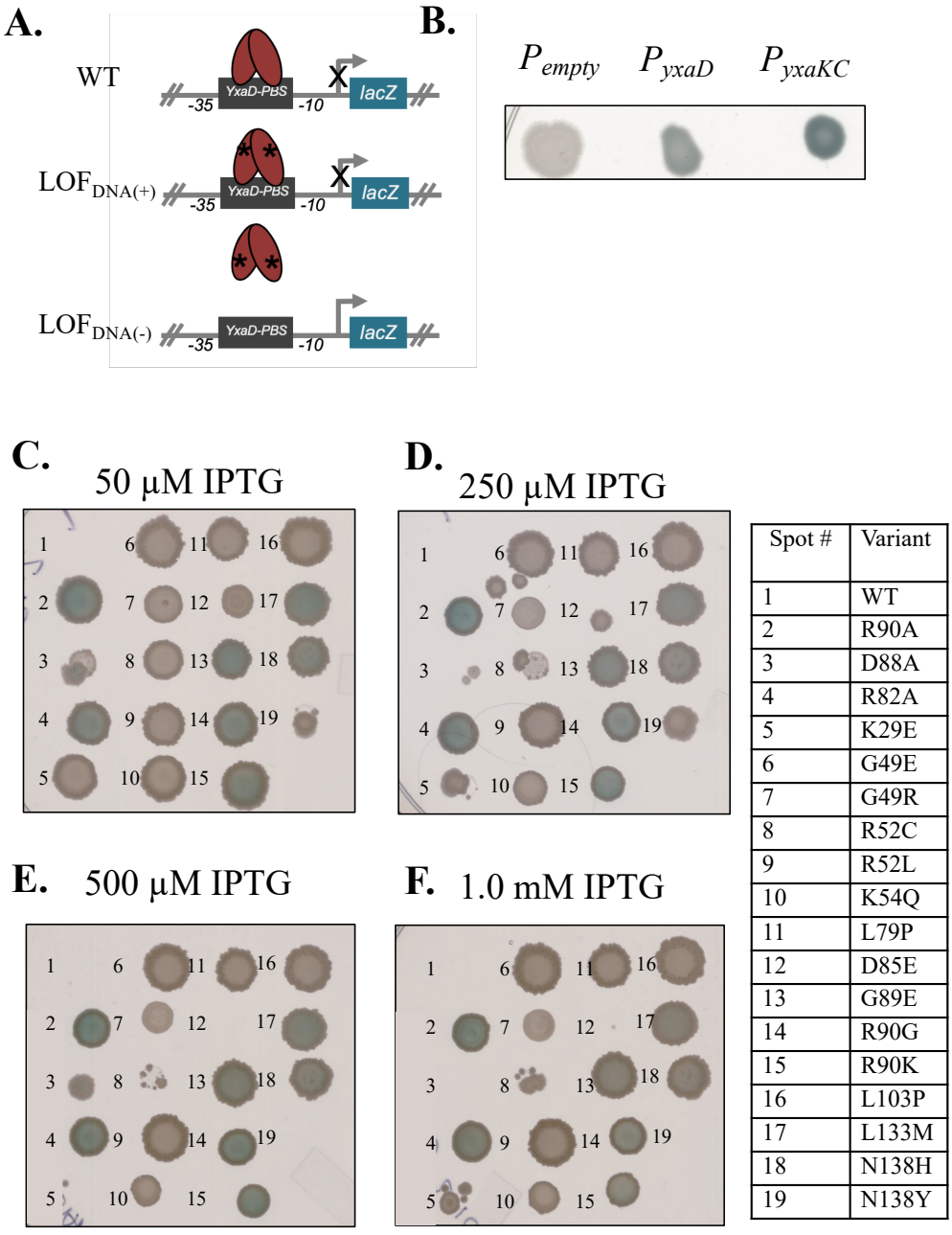
resistant to *yxaD* misexpression but still maintained YxaD's DNA-binding activity, we inserted $P_{hy}$-*yxaD* into the $P_{yxaKC}$-*lacZ* Δ*yxaD* reporter strain and plated on medium containing both inducer and X-gal. Suppressors that grew up were screened for YxaD's ability to bind DNA by identifying white colonies. In addition, the $P_{yxaKC}$-*lacZ*, Δ*yxaD* reporter eliminates LacI mutants that are unable to derepress in the presence of inducer. In the initial testing of this screen, we noticed that few suppressors were obtained at the levels of inducer used for microscopy (1 mM IPTG). Therefore, we attempted to isolate variants with different toxicities by performing the suppressor selection on IPTG concentrations ranging from 50 μM-500 μM. From this screen we isolated 30 total suppressors: 9 on 50 μM, 15 on 250 μM and 6 on 500 μM IPTG. Sequencing of the *yxaD* open reading frame revealed these mutations to be in G49E, D85E, R90G, L103P, N138H, and N138Y. In addition, some suppressors had mutations in *lacO* which presumably led to decreased *yxaD* expression.

In addition to isolating YxaD variants that retain DNA-binding activity, we also wanted to see if we could isolate dominant negative variants as well. To select for dominant negative variants, we performed the suppressor selection using strains harboring two copies of $P_{hy}$-*yxaD* in the nonessential loci *amyE* and *yhdG*. Since only one copy of $P_{hy}$-*yxaD* is required for the no growth defect, only dominant negative variants or suppressors in both constructs will grow in the presence of inducer. In total, we obtained five different *yxaD* suppressor mutants coding for R52L, K54Q, G89E, R90G, and R90K substitutions. We noticed that all suppressors we obtained from this screen were found in the *yxaD* misexpression construct in the *amyE* locus. We attributed this to the result of higher expression levels from the *amyE* locus over the *yhdG* locus.

Thus, we constructed another strain carrying two copies of $P_{hy}$-*yxaD*, this time in the nonessential loci *yycR* and *yhdG*. Using this strain, we observed a more equal distribution of suppressors isolated from either *yycR* or *yhdG*. In total, we identified six unique mutations: G49R, R52C, R52L, L79P, R90G, and L133M.

We next sought to verify if LOF variants we obtained retained DNA binding *in vivo*. The misexpression constructs were introduced into isogenic backgrounds carrying a $P_{yxaKC}$-*lacZ* reporter and cultured on solid medium containing inducer and X-gal. In addition, since the LOF variants were selected on concentrations varying from 50 μM-500 μM IPTG, we examined growth defects in addition to DNA-binding by culturing on 50 μM, 250 μM, 500 μM, and 1.0 mM IPTG (Figure 3.7C). The results (summarized on Table 3.4), reveal that 10 out of the 18 variants, K29E, G49E, G49R, R52C, R52L, K54Q, L79P, D85E, L103P, and N138Y maintained DNA-binding activity. We refer to this class of LOF variants with retained (+) DNA-binding activity as $LOF_{DNA(+)}$. In contrast, five variants, G89E, R90G, R90K, L133M, and N138H were not able to bind DNA, in addition to and unsurprisingly, R90A and R82A. This class of loss of function variants is referred to as $LOF_{DNA(-)}$ for not binding DNA. D88A was able to grow on very low levels of inducer, suggesting that it was not fully functional with regard to the artificial expression phenotypes. In addition, we observed some blue coloration in the D88A expressing colonies; however, this blue appeared to spread out from a region of the colony suggesting a second site mutation developed. Interestingly, the N138Y and N138H variants both grew in the presence of low levels of IPTG (50 μM-250 μM) but only N138H grew in the presence of 1.0 mM IPTG. Furthermore, N138Y retained DNA-binding activity and N138H was unable to bind DNA suggesting an importance for this

residue on DNA-binding activity. Another very interesting finding was that while both retained DNA-binding activity, misexpression of G49E and G49R showed different colony morphologies. The YxaD G49R expressing colony appeared glossy and did not grow out as the others did. It is curious if the G49R variant could be targeting a different protein than wild type YxaD or lost the ability to target one or more of its potential protein targets.




| Spot # | Variant |
|---|---|
| 1 | WT |
| 2 | R90A |
| 3 | D88A |
| 4 | R82A |
| 5 | K29E |
| 6 | G49E |
| 7 | G49R |
| 8 | R52C |
| 9 | R52L |
| 10 | K54Q |
| 11 | L79P |
| 12 | D85E |
| 13 | G89E |
| 14 | R90G |
| 15 | R90K |
| 16 | L103P |
| 17 | L133M |
| 18 | N138H |
| 19 | N138Y |

**Figure 3.7 YxaD DNA-binding activity for LOF variants.**
(A) Schematic of screen designed to identify mutants that bind (white colonies) or do not bind (blue colonies) DNA. (B) Strains harboring Δ*yxaD* expressing either $P_{empty}$-*lacZ* (BSH305), $P_{yxaD}$-*lacZ* (BSH306), or $P_{yxaKC}$-*lacZ* (BSH307) spotted on LB agar containing 40 µg/ml X-gal. For screen for DNA-binding activity, cells were grown in LB liquid cultures, normalized to $OD_{600}$ = 0.025 and 2 µl were spotted on LB plates containing X-gal and indicated IPTG concentrations. Numbers in figure correspond to variants in Table 3.4 and 3.5.

**Table 3.4 *In vivo* DNA-binding activity of YxaD* $LOF_{DNA(+)}$ Variants**

| Variant | IPTG Concentration | | | |
|---|---|---|---|---|
| | 50 µM | 250 µM | 500 µM | 1 mM |
| WT | - | - | - | - |
| K29E | +++ | ++ | + | + |
| G49E | +++ | +++ | +++ | +++ |
| G49R | +++ | ++ | ++ | ++ |
| R52C | +++ | ++ | + | + |
| R52L | +++ | +++ | +++ | +++ |
| K54Q | +++ | ++ | ++ | ++ |
| L79P | +++ | +++ | +++ | +++ |
| D85E | ++ | + | - | - |
| L103P | +++ | +++ | +++ | +++ |
| N138Y | ++ | ++ | - | - |

**Table 3.5 *In vivo* DNA-binding activity of YxaD* $LOF_{DNA(-)}$ Variants**

| Variant | IPTG Concentration | | | |
|---|---|---|---|---|
| | 50 µM | 250 µM | 500 µM | 1 mM |
| WT | - | - | - | - |
| R82A | +++ | +++ | +++ | +++ |
| G89E | +++ | +++ | +++ | +++ |
| R90A | +++ | +++ | +++ | +++ |
| R90G | +++ | +++ | +++ | +++ |
| R90K | +++ | +++ | +++ | +++ |
| L133M | +++ | +++ | +++ | +++ |
| N138H | +++ | +++ | +++ | +++ |
| D88A[a] | + | + | ++ | - |

The suppressor mutations were mapped onto the predicted YxaD structure and revealed that most of the variants unable to bind DNA reside near the DNA-binding region (Figure 3.8). Additional LOF variants unable to bind DNA map to the dimerization domain of YxaD. Surprisingly, many of the residues that retain DNA-binding activity also map near the predicted DNA-binding motif. K29E, L103P and somewhat L79P map to the conserved ligand pocket of MarR proteins.

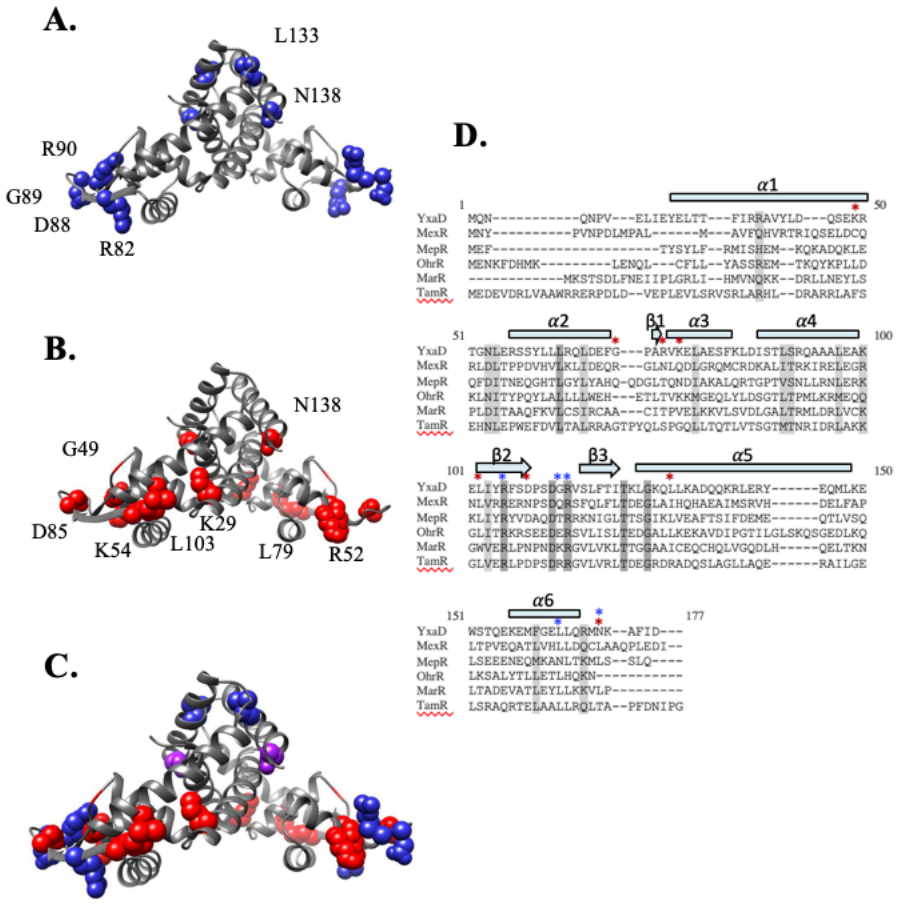



**Figure 3.8 YxaD LOF variants mapped to predicted structure.**
(A) $LOF_{DNA(-)}$ and (B) $LOF_{DNA(-)}$ variants are shown as blue and red spheres, respectively. (C) Both variant classes as in (A) and (B) with the exception of N138 variants (purple) that belong to both classes.

### 3.3.5. YxaD regulates *yxaKC* and *yxaD*

Although RNA-sequencing revealed that YxaD appeared to repress *yxaKC* activity (Chapter 2), we wanted to test this idea further and also address the physiological context of this regulation. In addition, since the RNA samples for RNA-seq were collected during *yxaD* overexpression or from Δ*yxaD*, it was not possible to determine if YxaD regulates its own promoter. Thus, we constructed fusions of the *yxaKC* and *yxaD* promoters, $P_{yxaKC}$ and $P_{yxaD}$ respectively, to the reporter gene, *lacZ*, and performed β-galactosidase assays in LB + 0.3% glucose (w/v) (LB-G). Expression from both $P_{yxaKC}$ and $P_{yxaD}$ increased over time from exponential to stationary, with maximal expression in stationary near $OD_{600}$~2.0, consistent with previously published transcription data (Nicolas *et al.*, 2012) (Figure 3.9A). Lastly, to determine if *yxaD* and *yxaKC* expression was a result of YxaD derepression, we examined transcription from $P_{yxaKC}$ and $P_{yxaD}$ in Δ*yxaD*. In the absence of *yxaD*, we observed an increase of expression by nearly two orders of magnitude from both promoters (Figure 3.9B). These data suggest that YxaD regulates *yxaKC* and *yxaD*, consistent with what has been observed for MarR-like proteins (Deochand & Grove, 2017) and in support of our RNA-seq analysis. Notably, expression from $P_{yxaKC}$ decreased with time even in the absence of *yxaD* suggesting this promoter is subject additional regulation. In order to determine if regulation of *yxaKC* had any effect on YxaD's misexpression phenotype, we generated strains harboring $P_{hy}$-*yxaD* and either $P_{hy}$-*yxaKC* or Δ*yxaKC*. The *yxaD* misexpression still occurred in the Δ*yxaKC* background. Moreover, overexpression *yxaKC* did not

affect cell viability or nucleoid morphology. These results suggest that YxaD's effects do not occur because of misregulation of *yxaKC* (data not shown).

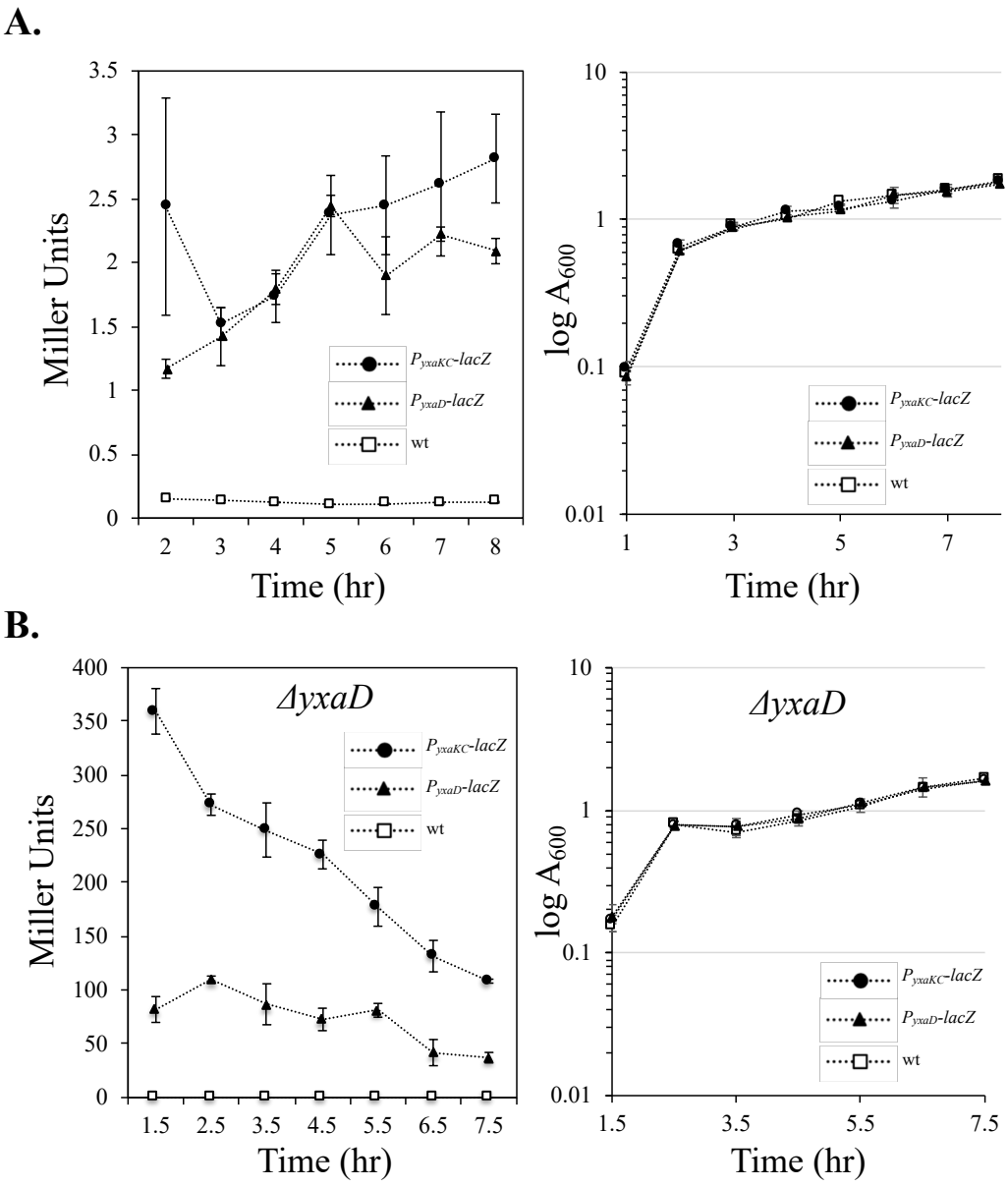



**Figure 3.9 Expression levels of *yxaD* and *yxaKC* promoters in LBG.**
Expression from putative *yxaD* and *yxaKC* promoter regions with (A) wild type *yxaD* (BSH290, BSH291, BSH292) or (B) Δ*yxaD* (BSH305, BSH306, BSH307) was monitored in LBG at 37 °C over the time course. The production of beta-galactosidase (left) or $OD_{600}$ (right) were monitored at 1 hr intervals.

### 3.3.6. YxaD binds two sites in the *yxaDKC* promoter

MarR-like proteins generally bind to palindrome sequences or palindromic repeats within their promoters to regulate transcription (Deochand & Grove, 2017). More specifically, MarR-like proteins generally bind sequences that overlap the -10 promoter region, thereby blocking RNA-polymerase. Therefore, we sought to determine if YxaD bound a similar motif in intergenic region of *yxaD* and *yxaKC*. Thus, we analyzed the 221 bp region of the *yxaDKC* operon using the EMBOSS palindrome sequence finder (Rice *et al.*, 2000) and identified two sets of potential 18 bp binding sites [Binding Site I (BS-I) and Binding Site II (BS-II)] (Figure 3.10A). Each potential binding site overlaps the predicted -10 of both $P_{yxaD}$ and $P_{yxaKC}$, consistent with characterized binding sites of MarR-like proteins (Deochand & Grove, 2017). The 18 bp palindromic repeats, TTGTAC/TTATACAAGTATA, were nearly identical, with the exception of a single base pair (Figure 3.10B). No other motifs were identified from these putative motifs (allowing for up to two mismatches) within the *B. subtilis* genome. Interestingly, the putative YxaD binding sites were not exact palindromes, but instead consisted of two sets of 5 bp inverted repeats (IR) that overlapped; this is contrast to

most characterized MarR-like binding sites, which consist of only one IR.

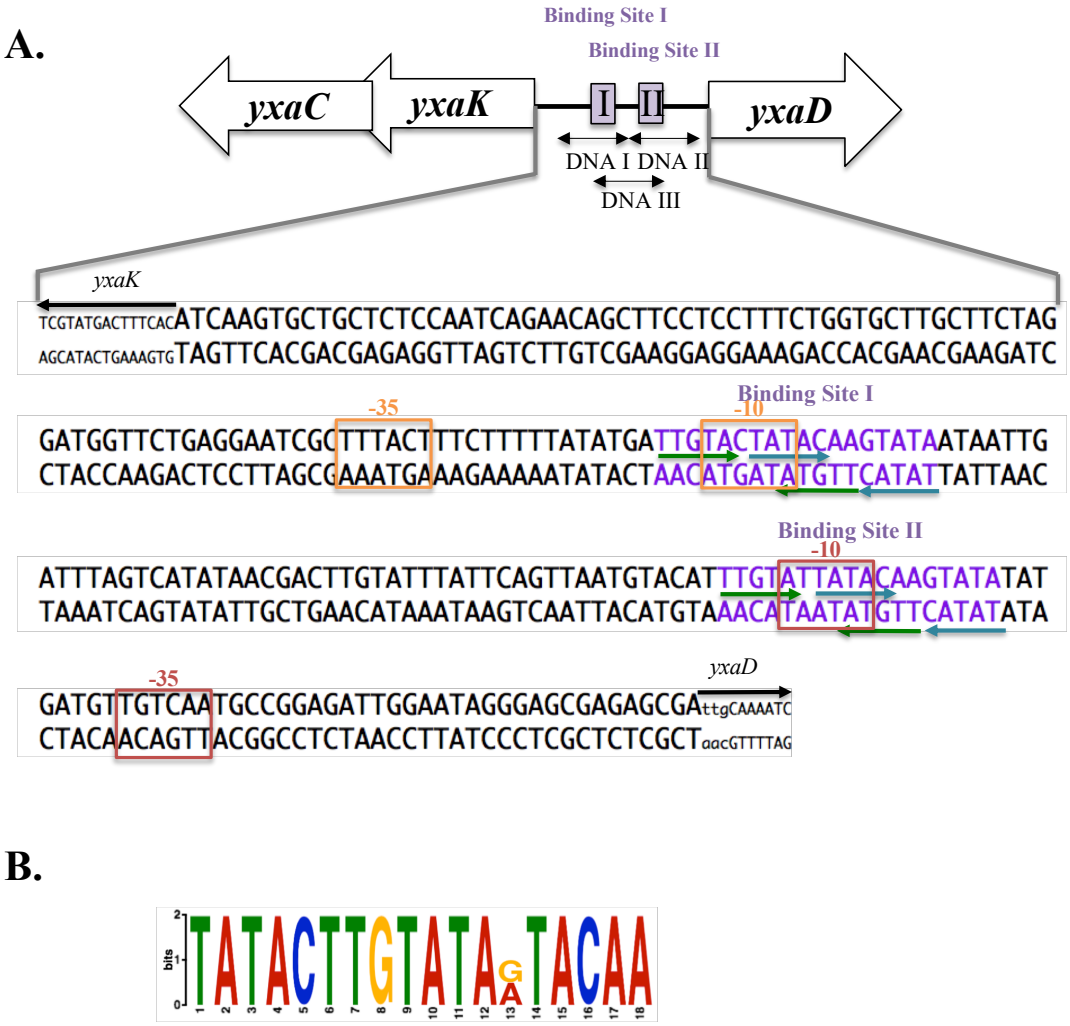



**Figure 3.10 Predicted binding sites of YxaD.**
(A) Intergenic region between *yxaD* and *yxaKC*. Predicted -10 and -35 promoter elements for *yxaD* and *yxaKC* are shown in orange and red, respectively. Predicted binding sites are shown as purple letters and as purple boxes. IR elements are shown as blue or green arrows in the direction of the repeat. Regions of DNA used EMSA (Figure 3.11) are shown as double-sided arrows. (B) Predicted YxaD DNA-binding site generated by MEME Suite (Bailey *et al.*, 2009).

To determine if YxaD interacts with the *yxaD yxaKC* promoters directly, we purified a His-tagged form of YxaD. We determined the functionality of this fusion by

checking for the *yxaD* overexpression phenotype and determined 6His-YxaD was functional (data not shown). A 150 bp DNA probe was amplified from the intergenic region of *yxaDKC,* encompassing both BS I +II (Figure 3.10A), and used in an electrophoretic mobility shift assay (EMSA). YxaD produced a single shift at 2.5 nM protein and two shifts at 25 nM, indicating that YxaD binds to the predicted promoter region of *yxaDKC* (Figure 3.11A). The two shifts observed could correspond to: 1) YxaD initially bound as a monomer and then as dimer (or other combination of multimers, i.e. dimer then tetramer), or 2) binding of YxaD dimers at two sites on the DNA probe. Because MarR-like proteins typically bind DNA as dimers and since we identified two nearly identical DNA-binding motifs, we hypothesized the latter was more likely. To distinguish between 6His-YxaD higher-order binding with increasing protein concentration or binding to both sites simultaneously, we designed two additional 150 bp DNA probes containing either BS-I or BS-II and performed the EMSA with each probe. In agreement with 6His-YxaD binding to two sites, we observed only a single shift in the presence of either BS-I or BS-II (Figure 3.11BC). 6His-YxaD binding to BS-I or BS-II occurred at an apparent $K_d$ of 50 nM and 100 nM, respectively. Taken together, these data suggest that YxaD binds two sites of the intergenic region between *yxaD* and *yxaKC* to regulate transcription of both operons.

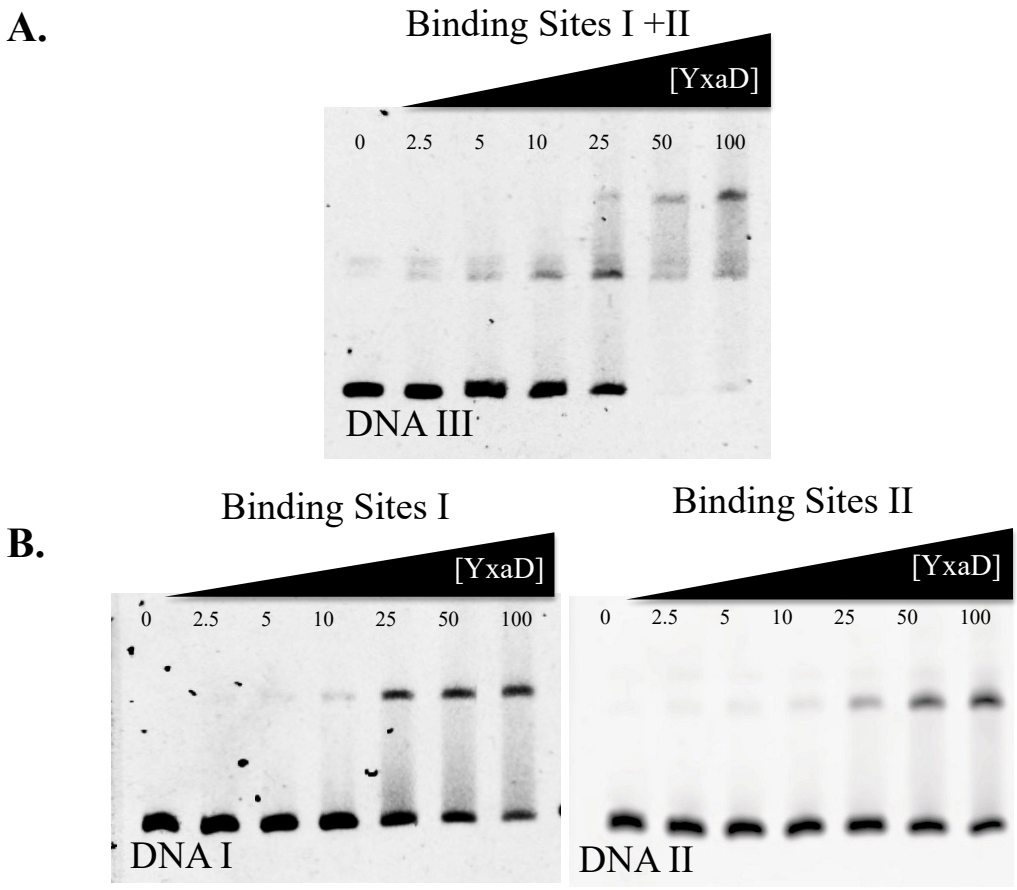



**Figure 3.11 YxaD binds two sites in the *yxaD* and *yxaKC* promoter region.** EMSAs of increase amounts of purified 6His-YxaD and 0.1 ng of the following DNA pieces (A) DNA III (binding sites I and II), (B) DNA I (binding site I only), or (C) DNA II (binding site II only).

### 3.4. Discussion

Here, we show that the predicted DNA-binding protein, YxaD, is a transcriptional regulator, repressing expression of both *yxaD* and *yxaKC*. We found that

YxaD binds to two sites in the *yxaD*/*yxaKC* promoter regions consistent with other characterized MarR-like proteins. Cells grown in LB or CH overexpressing *yxaD* have clear defects in chromosome segregation and appear to have defects in DNA replication. ScpA was identified as a possible target for YxaD by Y2H (Marchadier *et al.*, 2011), a finding we were able to recapitulate by B2H. However, the YxaD misexpression phenotype did not depend on the presence of *scpA*, suggesting perturbations due to YxaD overexpression are not simply a function of perturbation of the condensing complex. We explored the possibility that YxaD could be redundant with ParB. Since we successfully constructed a Δ*parB* Δ*yxaD* mutant, we deduced that ParB and YxaD are likely not redundant in loading SMC complexes. This is consistent with recent data indicating ParB and *parS* sites are required for the juxtaposition and long-range interactions of chromosome arms (Wang *et al.*, 2015). Still, we do not exclude the possibility that YxaD and ScpA may interact in vivo. In support of this idea, ScpA has been shown to interact with proteins involved in recombination and repair, two-component sensor kinases, and at least two other putative transcriptional regulators, YdeL and YlbO (Dervyn *et al.*, 2004). Moreover, *scpA* mutants were isolated that resulted in defects in DNA repair and gene regulation (Dervyn *et al.*, 2004) and further, a recent study that tracked ScpA-YFP foci relative to SMC found that 40% of ScpA is not associated with condensin, suggesting it may have roles outside the condensin complex (Schibany *et al.*, 2018). The nucleoid phenotype associated with artificial induction of YxaD is non-physiological. However, the fact that YxaD interacts with two independent proteins involved in DNA replication (HolA and ScpA) in Y2H and Y2H and B2H assays respectively and that the phenotype appears to affect DNA replication

and segregation suggests there may be a relationship between the DNA replication machineries and YxaD function. At present, we lack hypotheses for why such an interaction, if any, might exist.

We have additionally considered the possibility that the misexpression phenotype is not a result of direct targeting by YxaD, but merely due to elevated levels of DNA-bound YxaD. MarR-like proteins such as MarR and OhrR have been shown to bend DNA when bound (Zhu *et al.*, 2017b, Hong *et al.*, 2005). Therefore, it's possible that YxaD overexpression could lead to DNA-bending, or to the recruitment of large YxaD complexes creating something to the effect of a DNA replication roadblock. Although YxaD misexpression does not result in as dramatic a phenotype as DNA roadblocks introduced on the right arm of the chromosome of *B. subtilis* (Bernard *et al.*, 2010), the effects on DNA replication and chromosome segregation are similar. One reason to think the phenotype is not simply due to non-specific or enhanced binding of YxaD on the chromosome is that we were able to isolate a number of YxaD variants that retain the ability to bind the identified promoter region, but were unable to kill the cell (see below).

Intriguingly, we obtained 10 YxaD $LOF_{DNA(+)}$ variants that lost the ability to kill cells yet still maintained DNA-binding activity, namely K29E, G49E, G49R, R52C, R52L, K54Q, L79P, D85E, L103P, and N138Y. Unexpectedly, most of these variants map to a region near the DNA-binding motif on the predicted structure of YxaD (Figure 3.8, red spheres). An obvious explanation for the $LOF_{DNA(+)}$ variants is that they prevent interaction of YxaD with whatever target(s) lead to the misexpression phenotypes. Another possibility is that the $LOF_{DNA(+)}$ variants could have a higher affinity for DNA. YxaD might use DNA as a positional cue and once released from DNA (either by ligand

binding or on/off rates) is in a form capable of targeting its partner. In this way, more tightly bound variants would be expected to have less of a growth defect as they target less. This is consistent with a super-repressor identified in *E. coli*, that has a nearly 9-fold increase in DNA binding activity compared to wildtype (Alekshun & Levy, 1999). This hypothesis could be tested by purifying the $LOF_{DNA(+)}$ variants and testing their affinity for DNA. Along with this idea, if ligand binding leads to release of YxaD from the DNA, some variants could be disrupting ligand binding. The ligand binding pockets of most MarR-like proteins are made up of $\alpha$1, $\alpha$2, and $\alpha$1' (Kim *et al.*, 2016, Kumarevel *et al.*, 2009). In addition, structural analysis of the MarR-like protein, MexR, in complex with its antirepressor peptide, ArmR, reveals a similar pocket created by $\alpha$1, $\alpha$2, $\alpha$1' and $\alpha$2 (Wilke *et al.*, 2008). $LOF_{DNA(+)}$ variants K29E and G49E/R are predicted to lie on $\alpha$1/ $\alpha$1' and at the C-terminal end of $\alpha$2, respectively. Understanding the nature of these variants and YxaD's interaction with a ligand is difficult to deduce without knowing the ligand.

R52C, R52L and K54Q are charge changes and map to the DNA-binding region. Although these residues do not map to the winged loop or $\alpha$4 which are important for interactions with the DNA minor groove and sequence specificity in the major groove, it is still possible that these residues aid in the overall positive charge of YxaD's DNA-interaction domain. Therefore, mutations in these residues to uncharged or negatively charged could affect DNA-binding. However, since the R52C, R52L and K54Q $LOF_{DNA(+)}$ variants are still able to recognize the binding sites on the *yxaKC* promoter DNA, we speculate these residues could make YxaD binding to this region more specific. A similar conclusion can be made for L79P and L103P located in β1 and $\alpha$5,

respectively. The L79P substitution would likely disrupt the β-strand and $\alpha$-helix secondary structure due to the rigid structure of proline, potentially displacing the wHTH motif and $\alpha$5. The decrease in entropy associated with a proline substitution could also increase the stability of the DNA-binding motif, potentially reducing YxaD's affinity for other sites or increasing YxaD's affinity for *yxaKC* promoter DNA. D85 is conserved in YxaD, MepR and TamR and maps to the wHTH motif on the predicted YxaD structure which suggest a possible role for D85 in DNA-binding. However, the D85E substitution did not affect DNA-binding but lost the ability to kill when misexpressed. Oddly, the D85E substitution conserves the negative charge, suggesting that the change in size from aspartate to glutamate contributes to the reduced toxicity of this variant. This is supported by the susceptibility D85E still has to IPTG concentrations higher than 250 μM, a growth phenotype not exhibited by the other isolated variants. Lastly, N138Y was very interesting as this variant maintained DNA binding while N138H, which was also isolated, did not. In addition, N138Y grew much more poorly on IPTG than N138H. N138 is located near $\alpha$6 so the substitution could affect the dimerization interface. The N138Y is a relatively conservative substitution, with both residues being polar and uncharged. The N138H substitutes the polar, uncharged residue for with a positively charged residue.

In addition to the $LOF_{DNA(+)}$, we isolated five YxaD variants that did not kill cells and also did not bind DNA in our in vivo assay ($LOF_{DNA(-)}$). As expected, G89E, R90G and R90K were classed as $LOF_{DNA(-)}$ which we anticipated since these residues map to the DXR motif in the loop region and are known to be required for DNA-binding for MarR proteins (Alekshun *et al.*, 2001, Saito *et al.*, 2003). Interestingly, even the

R90K substitution maintains the positive charge, it was disrupted in binding DNA *in vivo* (comparable to the R90A substitution). This could be due to role of arginine in DXR in making hydrogen bonds between the guanidinium side chain with the minor groove DNA (Hong *et al.*, 2005). Similar to N138, L133M located on $\alpha$6 could disrupt the dimerization interface, although both are hydrophobic. Another intriguing possibility is that YxaD formers higher order oligomers. Although not frequent, some MarR-like proteins bind as tetramers and other higher order oligomers (Kim *et al.*, 2016, Chang *et al.*, 2014) (MepR-4XRF). Specifically, for HcaR, it appears the $\alpha$6s of each monomer interact to form a tetramer (Kim *et al.*, 2016). Therefore, if YxaD forms tetramers in vivo, the substitutions at L133 and N138 could disrupt this formation. In contrast however, we did not observe banding patterns consistent with tetramer formation by EMSA assays. Nevertheless, further biochemical analysis is required to probe the $LOF_{DNA(+)}$ and $LOF_{DNA(-)}$ variants and their roles in DNA-binding and misexpression. Furthermore, since most identified MarR-like variants are specific to those that abolish DNA-binding, we believe these variants could provide novel insight into the mechanisms MarR-like proteins use to bind DNA, ligands, or other proteins.

# 4. CHARACTERIZATION OF THE CID/LRG HOMOLOG, YXAKC

## 4.1. Introduction

A bacterium's ability to adapt to environmental changes relies on its capacity to sense and respond to stimuli. This type of regulation is often mediated at the level of transcription, where genetic switches or transcriptional regulators impart changes in gene expression in response to stimuli (i.e. metabolites). By binding to ligands utilized or similar to those utilized by their regulated gene products, transcriptional regulators can limit gene expression to only when required. In addition, in order for bacteria to elicit a proper response, gene expression patterns must be tightly linked to nutrient availability, ensuring the requirements for growth are met, and not wasted. Bacteria encode different regulatory mechanisms to enable their survival in their diverse, native environments. For example, in the wild, the soil dwelling bacterium, *Bacillus subtilis*, must adapt to utilize plant materials secreted from roots and induce various developmental programs leading biofilm formation (Lugtenberg & Kamilova, 2009, Allard-Massicotte *et al.*, 2016, Marschner, 2011). Even when considering a seemingly simplistic example such as growth in excess glucose, a variety of regulatory mechanisms still exist to finely tune gene expression to direct metabolic flux, for example, catabolite repression by cyclic AMP (cAMP) and the catabolite control protein A (CcpA) in *E. coli* and *B. subtilis*, respectively (Saier *et al.*, 1995).

*B. subtilis* CcpA is a global regulator of gene expression in response to available carbon and acts as both a transcriptional repressor as well as activator by binding to CRE DNA elements (Fujita, 2009, Warner & Lolkema, 2003). During growth in excess

glucose, or other fermentable sugars, CcpA-mediated transcription factor activity represses the TCA cycle to prevent excess ATP production and activates various genes involve in overflow metabolism. When grown in media containing excess glucose a large amount of imported glucose is metabolized to pyruvate and acetyl-CoA, which are then subsequently converted to compounds such as lactate, acetate and acetoin for excretion as overflow metabolites. Production of these overflow metabolites are needed to maintain homeostatic balance during growth on glucose. For instance, conversion of pyruvate to lactate regenerates the NAD+ required for glycolytic enzymes and in addition, the phosphotransacetylase (PTA) and acetate kinase (AK) conversion of acetyl-CoA to acetate generates additional ATP. Furthermore, pyruvate is converted to acetoin, which mitigates overacidification of the cytoplasm due to accumulation of intracellular acetate and pyruvate. When glucose is depleted, overflow metabolites are imported and subsequently utilized through the citric acid cycle to generate ATP. Due to the complexity and potential for disastrous outcomes, timing and control of gene expression must be tightly coordinated with metabolism.

Cid/Lrg proteins are a family of transmembrane proteins shown to play multiple roles in response to overflow metabolism (Yang *et al.*, 2005, Charbonnier *et al.*, 2017). Cid/Lrg operons are ubiquitous in bacteria and generally consist of two genes encoding transmembrane proteins. Expression of both *cidAB* and *lrgAB* operons has shown to be regulated by two overlapping pathways (Yang *et al.*, 2005). In some organisms, the *cidAB* operon is positively regulated by the LysR-like regulator, CidR, which activates expression in response to acetic acid (Yang *et al.*, 2006, Yang *et al.*, 2005). *lrgAB* have been shown to be regulated by a two-component system, encoded by *lytST* (Ahn *et al.*,

2010). *cidAB* and *lrgAB* are differentially expressed and regulated by catabolite repression, and are therefore, dependent on glucose concentration and growth phase (Groicher *et al.*, 2000, Ahn & Rice, 2016, Ahn *et al.*, 2010, Kim *et al.*, 2019). At low glucose concentrations (at the onset of stationary phase), *lrgAB* expression is high, whereas *cidAB* expression is low. In contrast, during growth in excess glucose, *cidAB* expression is high and *lrgAB* expression is low. Both *cidAB* and *lrgAB* operons were shown to be directly regulated by CcpA; more specifically, CcpA activates *cidAB* and represses *lrgAB* (Kim *et al.*, 2019).

Most of the studies on Cid/Lrg proteins have been primarily conducted in the human pathogen, *Staphylococcus aureus*, and the predominant bacterium found in human dental caries, *Streptococcus mutans*. In *S. mutans*, Δ*lrgAB* mutants are more susceptible to environmental stressors including oxidative stress, heat, and vancomycin (Ahn *et al.*, 2010, Ahn & Rice, 2016, Rice *et al.*, 2017, Ahn *et al.*, 2017). In *S. aureus*, Cid/Lrg proteins are thought to regulate the coordination and timing of cell lysis in glucose rich media (Bayles, 2007, Rice & Bayles, 2008). It was proposed that Cid/Lrg proteins regulated lysis in a subpopulation of cells within a community to help form and stabilize biofilms by providing materials such as eDNA to the rest of the community (Bayles, 2007, Rice & Bayles, 2008). Recently, a study in *S. mutans* found the *lrgAB* genes to be the highest upregulated genes in thicker biofilms indicating a role for biofilm regulation across organisms (Shemesh *et al.*, 2008).

Although not to the same extent as *S. mutans* and *S. aureus*, Cid/Lrg proteins have also been studied in *Bacillus cereus* and *Bacillus anthracis*. Both *B. cereus* and *B. anthracis* encode four Cid/Lrg-like operons, termed CidAB and LrgAB, and in addition,

ClhAB$_1$ and ClhAB$_2$ standing for *cid*/*lrg* homologues AB-1 and AB-2 (Chandramohan *et al.*, 2009, Huillet *et al.*, 2017). Similar to *S. aureus* and *S. mutans*, *cidAB* and *lrgAB* from *B. cereus* and *B. anthracis* are regulated by a LysR-like regulator, CidR, and a two-component response regulator. In contrast, *clhAB$_1$* is expressed in an operon encoding a putative GntR-like transcriptional regulator and a hypothetical protein and *clhAB$_2$* is positively regulated by CodY. CodY is a global regulator that positively and negatively regulates genes to elicit adaption in response to nutrient availability by sensing the available pool of GTP and branched-chain amino acids (BCAAs) isoleucine, leucine and valine (ILV) (Brinsmade & Sonenshein, 2011, Levdikov *et al.*, 2009, Levdikov *et al.*, 2006, Ratnayake-Lecamwasam *et al.*, 2001, Shivers & Sonenshein, 2004). The CodY-dependent *clhAB$_2$* transcription observed in *B. cereus* was independent of glucose. Interestingly however, *clhAB$_2$* were shown to be required for glucose-dependent cell chaining, and that therefore, *clhAB$_2$* leads to down-regulation of cell separation during growth in glucose (Huillet *et al.*, 2017).

Intriguingly however, although Cid/Lrg proteins are implicated in programmed cell lysis (PCD), no direct mechanism for how this could be mediated has been elucidated (Bayles, 2007, Rice & Bayles, 2008). Recent analyses on the *S. mutans* transcriptome and proteome during oxidative, heat or vancomycin-related stress in Δ*lrgAB* revealed large shifts in metabolic genes (Rice *et al.*, 2017, Ahn *et al.*, 2017). In addition, the expression pattern observed for *lrgAB* closely mimicked that of the pyruvate dehydrogenase complex (PDH). These data and others more recently suggest a role for Cid/Lrg proteins in maintaining cellular homeostasis during growth in conditions of stress (Kim *et al.*, 2019). In support of this idea, the *B. subtilis* encoded LrgAB-

homologs, PftAB, were recently discovered to form a pyruvate transporter and are required for efficient growth in media where pyruvate is the sole carbon source (van den Esker *et al.*, 2017, Charbonnier *et al.*, 2017). These findings presented the first crucial link between overflow metabolism and Cid/Lrg function.

*B. subtilis* encodes three Cid/Lrg homologs, *ywbHG*, *pftAB*, and *yxaKC*, where YwbHG and PftAB are more similar to CidAB and LrgAB, respectively, and *yxaKC* are uniquely expressed in some *Bacillus* species (Figure 4.1). Similar to what was previously described in *S. aureus*, *ywbHG* (*cidAB*) expression was found to be induced by YwbI (CidR-like) in the presence of acetate (Yang *et al.*, 2005, Chen *et al.*, 2015) and *pftAB* (*lrgAB*) expression occurred during late exponential/stationary expression and was repressed by CcpA (van den Esker *et al.*, 2017, Charbonnier *et al.*, 2017). In addition to CcpA-mediated repression of *pftAB*, a CcpA-independent regulatory mechanism was identified involving pyruvate. High levels of extracellular pyruvate are sensed by LytST which induces expression of *pftAB* (Charbonnier *et al.*, 2017). In contrast, intracellular pyruvate (or other metabolic intermediate) represses *pftAB* expression (Charbonnier *et al.*, 2017).

Despite the wealth of data on Cid/Lrg proteins and their potential roles in biofilm formation, cell lysis, and adaption to environmental stress, mechanistic details surrounding how Cid/Lrg-like proteins may be regulating these processes is lacking. The recent discovery that PftAB transports pyruvate reveals a possible link between regulating cellular homeostasis and overflow metabolism (Charbonnier *et al.*, 2017). Because of their diversely described roles in regulated cell lysis (Rice & Bayles, 2008) and pyruvate transport (Charbonnier *et al.*, 2017), characterizing the Cid/Lrg proteins is

crucial in understanding how bacteria have evolved to survive and adapt to their environments. In addition, Cid/Lrg homologs might encode novel transport systems we have yet to discover. In support of this idea, a recently identified plant CidAB/LrgAB, PLGG1 (or AtLrgB), encoded by *Arabidopsis thaliana* was discovered to function as a glycolate/glycerate transporter (Yang *et al.*, 2012, Pick *et al.*, 2013). PLGG1 encodes a LrgA-LrgB fusion that likely originated from bacteria (Yang *et al.*, 2012). Interestingly, little is known about *B. subtilis* encoded Cid/Lrg-like proteins, YxaKC, and if they have similar roles as PftAB. Here, we focus on the initial characterization of the *yxaKC* operon, how they are regulated and their potential role in regulating overflow metabolite(s) transport. We found that *yxaKC* expression is dependent on glucose depletion in addition to the MarR-like regulator, YxaD. In addition, we found cells overexpressing *yxaKC* accumulated higher levels of extracellular 2-acetolactate. Lastly, we examined context-specific expression of *yxaKC* in biofilms and observed *yxaKC* expression in a concentric ring in the center of the biofilm. Taken together, we hypothesize that cells in biofilm centers, which are more nutrient deprived than the outer rings, may induce YxaKC to scavenge extracellular 2-acetolactate.

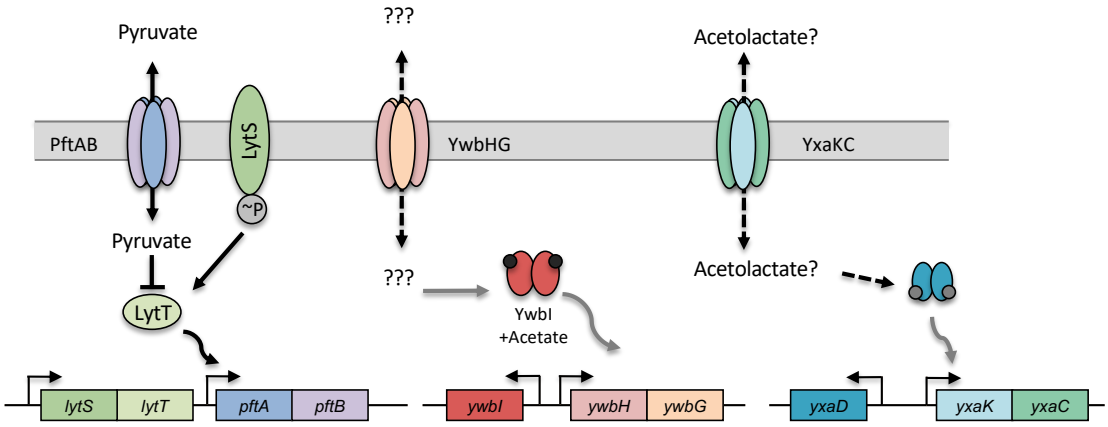



**Figure 4.1 Cid/Lrg homologs in *B. subtilis*.**

## 4.2. Materials and Methods

### 4.2.1. General Methods

All *B. subtilis* strains were derived from *B. subtilis* 168 or 3610. Cloning was carried out in *E. coli* DH5α. All *B. subtilis* and *E. coli* strains used in this study are listed in Table 4.1. Plasmids are listed in Table 4.2. Oligonucleotide primers are listed in Table 4.3. M9 minimal media was prepared as described (Kleijn *et al.*, 2010, Harwood, 1990). For M9-minimal plates, single carbon sources were added as follow: 3 g/L D-glucose, 6 g/L sodium pyruvate, 6 g/L glycerol or 6 (potassium or sodium, 4 or 6 g/L) g/L potassium gluconate. Plates were solidified with bacto-agar to a final of 1.5 (w/v) %. For growth in liquid culture, M9-minimal was supplemented with either 3 g/L D-glucose (M9G) or 2 g/L D-glucose with 4 g/L L-malic acid (M9M). Before addition, malic acid was brought to pH 7.0 with NaOH. For transformation of *E. coli*, 100 µg/ml ampicillin was used. For transformation and selection of *B. subtilis*, antibiotics were used at the following concentrations where indicated: 7.5 µg/ml chloramphenicol, 1 µg/ml

erythromycin with 25 μg/ml lincomycin, 10 μg/ml kanamycin, and 100 μg/ml spectinomycin.

For growth on plates, frozen glycerol stocks were streaked out onto LB plates. Single colonies were then streaked out onto M9-minimal plates with the designated carbon sources. Overexpression phenotypes were examined by inoculating single colonies into 5 mL LB media until cultures reached mid-exponential. Cultures were normalized to $OD_{600}$ = 0.25 and 5 μl was spotted onto plates containing 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) or 1 mM IPTG plus 0.3% glucose (w/v).

**Table 4.1 Strains used in Chapter 4**

| **Strain Parental** | **Description** | **Reference** |
|---|---|---|
| *B. subtilis 168* | *Bacillus subtilis* laboratory strain 168 *trpC2* | BGSC (1A866) |
| *B. subtilis* PY79 | *Bacillus subtilis* laboratory strain | (Youngman *et al.*, 1983) |
| *BSH450* | *Bacillus subtilis* laboratory strain 168 with PY79 *trpC2+* | |
| *B. subtilis* 3610 | *Bacillus subtilis* wild type strain 3610 Δ*comI* | Dan Kearns |
| *E. coli* DH5α | *F- endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG Φ80dlacZΔM15 Δ(lacZYA-argF)U169, hsdR17($r_K$- $m_K$+), λ–* | |
| ***B. subtilis 168*** | | |
| BSH388 | *ysbAB(pftAB)::cat (cat)* | This study |
| BSH226 | *ywbHG::erm (erm)* | This study |
| BSH420 | *yxaKC::kan (kan)* | This study |
| BSH425 | *ysbAB(pftAB)::cat (cat), yxaKC::kan (kan)* | This study |
| BSH426 | *ywbHG::erm (erm) (erm), yxaKC::kan (kan)* | This study |
| BSH428 | *ywbHG::erm (erm), ysbAB(pftAB)::cat (cat)* | This study |
| BSH427 | *ywbHG::erm (erm), ysbAB(pftAB)::cat (cat), yxaKC::kan (kan)* | This study |
| BEA911 | *amyE::$P_{hy}$-ywbH (spec)* | This study |
| BEA912 | *amyE::$P_{hy}$-ywbG (spec)* | This study |

**Table 4.1 Continued**

| Strain Parental | Description | Reference |
|---|---|---|
| BEA909 | *amyE::$P_{hy}$-yxaK (spec)* | This study |
| BEA910 | *amyE::$P_{hy}$-yxaC (spec)* | This study |
| BEA917 | *amyE::$P_{hy}$-ysbA (pftA) (spec)* | This study |
| BEA918 | *amyE::$P_{hy}$-ysbB (pftB) (spec)* | This study |
| **BSH450** | | |
| BSH461 | *amyE::spoVG-lacZ (cat)* | This study |
| BSH462 | *amyE::$P_{yxaD}$-spoVG-lacZ (cat)* | This study |
| BSH463 | *amyE::$P_{yxaKC}$-spoVG-lacZ (cat)* | This study |
| BSH451 | *yxaKC::kan (kan)* | This study |
| BSH457 | *yxaD::kan (kan)* | This study |
| ***B. subtilis* 3610** | | |
| BSH293 | *amyE::spoVG-lacZ (cat) , ΔcomI* | This study |
| BSH294 | *amyE::$P_{yxaD}$-spoVG-lacZ (cat) , ΔcomI* | This study |
| BSH295 | *amyE::$P_{yxaKC}$-spoVG-lacZ (cat) , ΔcomI* | This study |
| BSH315 | *yxaD::spec, amyE::spoVG-lacZ (cat) , ΔcomI* | This study |
| BSH316 | *yxaD::spec, amyE::$P_{yxaD}$-spoVG-lacZ (cat) , ΔcomI* | This study |
| BSH317 | *yxaD::spec, amyE::$P_{yxaKC}$-spoVG-lacZ (cat) , ΔcomI* | This study |

**Table 4.2 Plasmids used in Chapter 4**

| Plasmid | Description | Reference |
|---|---|---|
| pDG1661 | *amyE-lacZ (cat)(spec)(amp)* | |
| pSH042 | *amyE-$P_{yxaD}$-lacZ (cat)(spec)(amp)* | This study |
| pSH043 | *amyE-$P_{yxaKC}$-lacZ (cat)(spec)(amp)* | This study |

**Table 4.3 Oligos used in Chapter 4**

| Oligo | Sequence 5' to 3' |
|---|---|
| OSH101 | TAACTTCGTATAATGTATGCTATACGAACGGTAGAATTCATCAAGT GCTGCTCTCCAATC |
| OSH114 | aaaaaGGATCCTCGCTCTCGCTCCCTATT |
| OSH115 | aaaaagaattcTCGCTCTCGCTCCCTATT |
| OSH116 | aaaaaGGATCCATCAAGTGCTGCTCTCCAATC |

### 4.2.2. β-galactosidase assays

Strains for transcriptional fusion assays were generated by cloning the intergenic region between *yxaD* and *yxaKC* from both directions into pDG1661 creating a *lacZ*

fusion. The construct was then integrated into the non-essential locus, *amyE*, of *B. subtilis*. Three colonies were inoculated into three independent 5 ml cultures of LB and grown for 4-5 hours at 37 °C. Cultures were then subcultured into M9 supplemented with the indicated carbon source to dilutions of 1/500, 1/1250 and 1/2000 and grown overnight (8-10 hr) at 37 °C. Cultures that were between $OD_{600}$ of 0.4-0.7 were then diluted to an $OD_{600}$ of 0.03 in 30 mL of M9 with the appropriate carbon source. Cultures were grown in 250 ml baffled flasks at 37 °C shaking in a water bath set to 300 rpm. OD readings and samples for β-galactosidase assay were taken every hour. β-galactosidase assays were performed as described previously (Ababneh & Herman, 2015), except 600 µl was collected for each sample.

### 4.2.3. $^{1}$H NMR

NMR experiments were conducted similar to what was previously described (Hochgrafe *et al.*, 2008). Briefly, cells were grown in M9G for 7 hr to an $OD_{600}$ of 3. One ml of cell culture was filter sterilized and kept at -20 °C until further analysis. For NMR experiments, 400 µl cell supernatant was mixed with 48 µl of 1 mM sodium hydrogen phosphate buffer, pH 7.0, 60 µl of 10 mM sodium 3-trimethylsilyl-[2,2,3,3-$D_4$]-1-propionic acid (TMSP) made up with 100% $D_2O$ and 92 µl $ddH_2O$ to final concentrations of 0.08 mM sodium hydrogen phosphate buffer, 1 mM TMSP and 10% $D_2O$. Spectra were obtained at 500 MHz at 298.5 K with Bruker AVANCE III 500 MHz spectrometer operating TOPSPIN. In total, 128 free induction decays (FID scans) were obtained in 64k data points.

### 4.2.4. Biofilm formation

MSgg plates were made as previously described (Branda *et al.*, 2001) with the following exceptions: 15 ml was poured in 60 x 15 mm plates and 100 μg/ml 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal) was added. *Bacillus subtilis* 3610 strains were streaked from frozen glycerol stocks. Three individual colonies were used to inoculate 3 ml of LB media. Three independent cultures were grown for 3-4 hours at 37 °C until culture density was between $OD_{600}$ = 0.6-0.7. Individual cultures were then normalized to $OD_{600}$ = 0.6 and 5 μl of the normalized dilution was spotted in the center of an MSgg plate.

## 4.3. Results

### 4.3.1. *yxaKC* are not required for growth on pyruvate, glucose, gluconate, or glycerol

YxaKC are members of the Cid/Lrg protein family, encoded by *B. subtilis* which also encodes PftAB and YwbHG (Figure 4.1). PftAB has been shown to function as a facilitated pyruvate transporter and a Δ*pftAB* strain grew poorly when provided pyruvate as a sole carbon source (Charbonnier *et al.*, 2017, van den Esker *et al.*, 2017). In addition, the *ywbHG* operon was found to be induced by acetate (Chen *et al.*, 2015). These recent findings that *B. subtilis* PftAB and YwbHG may play roles in central metabolism prompted us to ask whether YxaKC was regulated by or performed a similar function. Charbonnier and colleagues found that Δ*pftAB* had reduced growth in pyruvate compared to wild type (Charbonnier *et al.*, 2017). Consistent with this finding, we grew Δ*pftAB* on plates containing pyruvate as the sole carbon source (M9P) and observed that

PftAB were required for efficient growth, though they were not essential (Figure 4.2B). Therefore, we asked whether YxaKC and/or YwbHG might be acting redundantly with PftAB as a pyruvate transporter. In order to test if Δ*yxaKC* and/or Δ*ywbHG* had any growth effects on pyruvate we grew combinations of Δ*pftAB*, Δ*yxaKC*, and Δ*ywbHG* on M9P. Combinatory mutations of Δ*yxaKC* or Δ*ywbHG* did not alter the growth defects associated with Δ*pftAB* indicating that *yxaKC* and *ywbHG* likely do not play a direct role in pyruvate utilization (Figure 4.2B). In addition, to ask if *yxaKC*, *ywbHG* or *pftAB* were required for growth on additional carbon sources, we grew each deletion strain on M9-glucose (M9G), M9-glycerol (M9-Gly), and M9-gluconate (M9-Glu). Neither Δ*yxaKC*, Δ*ywbHG*, Δ*pftAB*, or combinations of all three mutants resulted in a growth defect on M9G, M9-Gly or M9-Glu indicating that these Cid/Lrg homologs likely do play a role in utilization of glucose, glycerol or gluconate (Figure 4.2C-E).

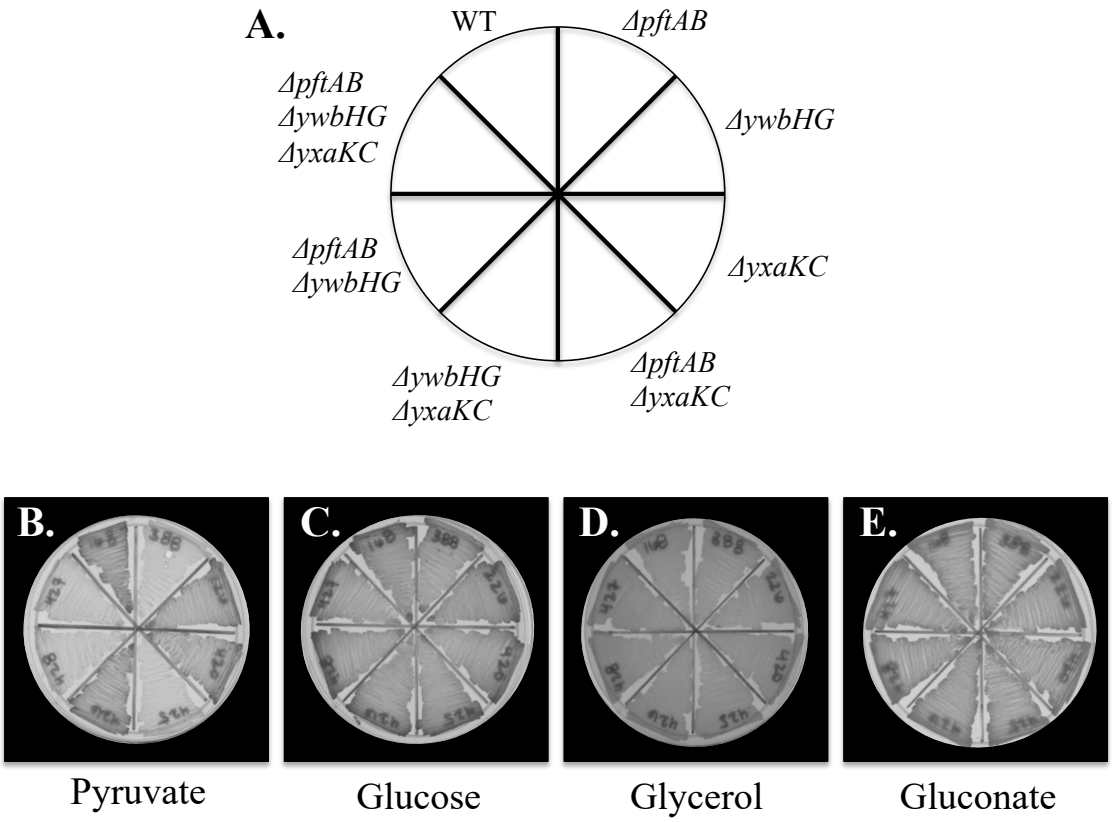



**Figure 4.2 Growth of *B. subtilis cid/lrg* mutants on various carbon sources.** Single colonies of strains harboring Δ*pftAB,* Δ*ywbHG,* Δ*yxaKC,* or combinations of the three were streaked on M9 minimal plates supplemented with either (B) pyruvate, (C) glucose, (D) glycerol, or (E) gluconate.

### 4.3.2. *yxaC* or *yxaK* overexpression foes not lead to cell lysis

A recent study conducted in *S. aureus* found that a deletion in the two-component response regulator, *srrAB*, resulted in increased *cidAB* expression and subsequent cell lysis during growth in a glucose-rich media. This phenotype appeared to be a result of SrrAB-mediated derepression of *cidB*, and consequent *cidB* overexpression (Windham *et al.*, 2016) and could be observed on agar plates (Ahn *et al.*, 2010). We wondered whether misexpression of the *cidB/lrgB* homologs encoded by *B. subtilis*, *pftB*, *ywbG*, and *yxaC*, acted in this way leading to slow growth or cell lysis. To test this,

we generated strains carrying an inducible copy of *pftB*, *ywbG*, or *yxaC* and spotted these strains on LB or glucose-rich LB agar plates in the presence of inducer. Overexpression of *pftB*, *ywbG*, or *yxaC* did not appear to result in a growth defect (Figure 4.3A). In addition, we constructed similar strains carrying *pftA*, *ywbH*, or *yxaK* under inducible promoters. However, overexpression of *pftA*, *ywbH*, or *yxaK* did not lead to a growth defect, consistent with what was previously reported for *pftA* and *ywbH* (Figure 4.3B) (van den Esker *et al.*, 2017). Taken together, these data suggest that *pftAB*, *ywbHG*, and *yxaKC* do not affect cell lysis *in B. subtilis*.

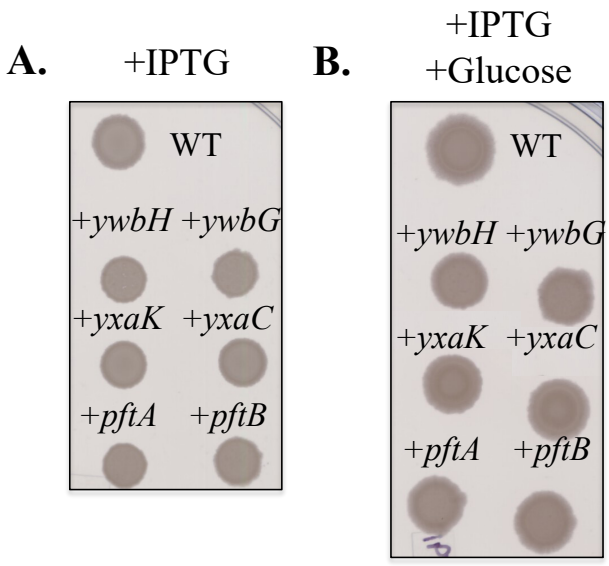



**Figure 4.3 Misexpression of *pftAB*, *ywbHG*, and *yxaKC*.**

### 4.3.3. *yxaD* and *yxaKC* are expressed in stationary phase during glucose depletion

Expression of *cid/lrg* operons in *S. aureus* and *S. mutans* is dependent on glucose concentration and growth phase (Groicher *et al.*, 2000, Ahn & Rice, 2016, Ahn *et al.*, 2010, Kim *et al.*, 2019). Growth in excess glucose leads to the accumulation of acetate which induces expression of *cidAB(C)* in *S. aureus* (Chaudhari *et al.*, 2016) and *B.*

*subtilis* (Chen *et al.*, 2015). In contrast, expression of *lrgAB* and *pftAB* is repressed by CcpA (Charbonnier *et al.*, 2017, van den Esker *et al.*, 2017) and therefore, repressed until stationary phase when glucose is depleted and *pftAB* is activated by extracellular pyruvate. Thus, *pftAB* expression does not occur until not only glucose depletes, but extracellular pyruvate accumulates (Charbonnier *et al.*, 2017). Based on these data, we tested if the third *B. subtilis* Cid/Lrg-like operon *yxaKC,* were regulated and if this regulation was similar to that of *ywbHG* or *pftAB*. We showed that *yxaKC* is repressed by the MarR-like protein, YxaD (Chapter 3). To our knowledge, this is the first Cid/Lrg-like operon shown to be regulated by a MarR-like protein. Some MarR-like proteins alter gene expression in response to ligands and have regulatory roles in metabolic genes (Deochand & Grove, 2017, Grove, 2017). In Chapter 3, we showed that *yxaKC* were expressed during stationary phase in LB + 0.3% glucose (w/v) (LB-G) and repressed by YxaD but did not examine glucose-specific gene expression. To determine if *yxaKC* were expressed during growth in glucose, we constructed fusions of $P_{yxaKC}$ to the reporter gene, *lacZ*, and integrated this construct in the *B. subtilis* chromosome at a non-essential locus. Since MarR-like proteins are generally autoregulatory, we also constructed a *lacZ* transcriptional fusion to the promoter of *yxaD*, $P_{yxaD}$. Cells harboring both transcriptional fusions were grown in M9-minimal media with glucose (M9G) and samples were collected every hour for β-galactosidase assays. When grown in M9G, $P_{yxaD}$ and $P_{yxaKC}$ transcription increases with growth for about 8 hr until cells lyse as a result of glucose depletion (Figure 4.4A) (Hadjipetrou & Stouthamer, 1963). Interestingly, unlike growth in LB-G, expression from $P_{yxaD}$ was significantly lower than $P_{yxaKC}$ when grown in M9G. This observation could indicate there is an additional

unknown repressor for *yxaD* or activator for *yxaKC.* Interestingly, *yxaKC* expression increases gradually until hr 6 when there is a sharp increase in expression, which is consistent with timing of glucose depletion. We wondered whether the gene expression patterns for *yxaD* and *yxaKC* we observed was a result of being grown in glucose or due to growth-phase (i.e. stationary). *B. subtilis* has been shown to use malate at similar levels as glucose (Kleijn *et al.*, 2010). Thus, we performed the transcriptional fusion assay on cells grown in M9 minimal media supplemented with malate and a lower amount of glucose than M9G (M9MG). The expression patterns of *yxaD* and *yxaKC* grown in M9MG were similar to that in M9G, increasing until lysis occurred around hr 6 (Figure 4.4B). However, in contrast to M9G, expression in M9MG steadily increased throughout growth without the inflection point observed for M9G. These data reveal that *yxaKC* expression is most similar to *pftAB*, expressed in stationary phase when glucose is depleted.

The *pftAB* operon was found to be repressed by catabolite repression by CcpA (Charbonnier *et al.*, 2017) and in addition, a genome-wide search for CcpA-regulated genes in *B. subtilis* revealed *ywbHG* activation by CcpA (Moreno *et al.*, 2001). However, we or others have not observed CcpA binding sites (CRE) or other known regulator binding sites at or near the promoters of *yxaD* and *yxaKC*. Therefore, we speculated that expression of *yxaKC* could be the result of YxaD derepression as we have shown that *yxaKC* are regulated by YxaD. DNA-binding by MarR-like proteins is often attenuated by binding ligands (Deochand & Grove, 2017). This prompted us to ask whether the increase in *yxaKC* expression was the result of induction by YxaD in response to a ligand and further, if YxaKC could play a role in regulating metabolic flux

of that ligand or one similar. Such transcriptional regulation could act as a way to control metabolic flux in response to metabolites. This idea is consistent with activation of *pftAB* in response to extracellular pyruvate allowing cells to utilize excreted pyruvate in the absence of glucose. In this way, we hypothesized that YxaKC could form a similar transport system that allows for the uptake of an overflow metabolite(s).

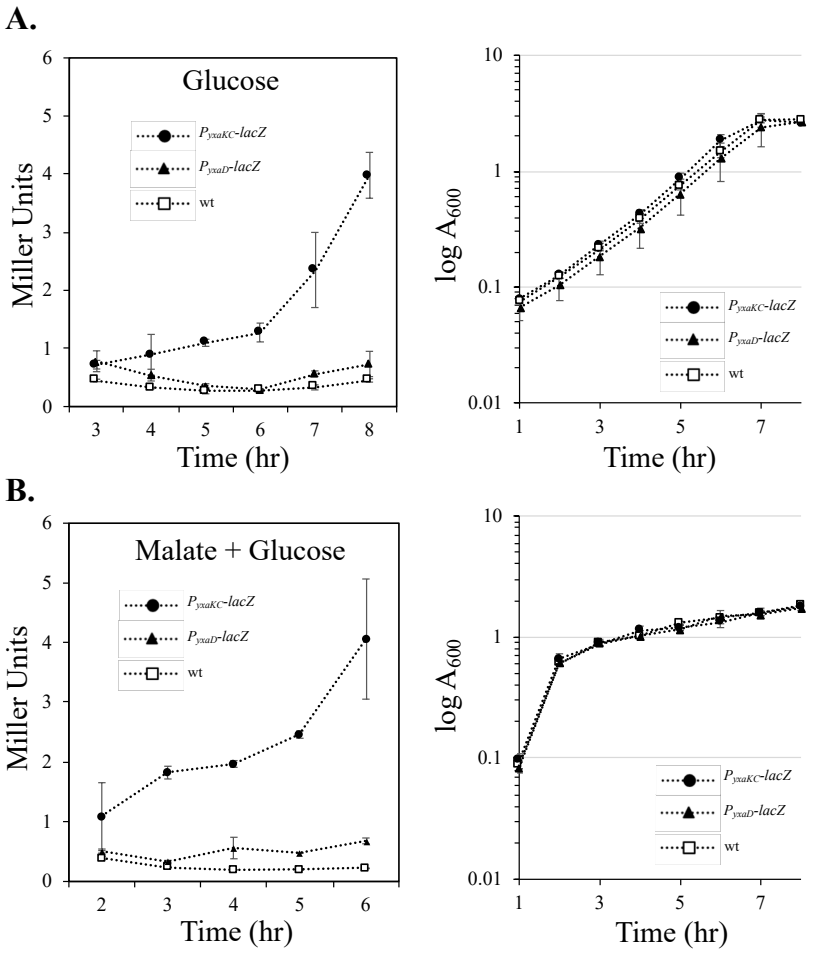



**Figure 4.4 Expression levels of *yxaKC* and *yxaD* in M9 minimal.**
Expression levels of *yxaD* and *yxaKC* promoters during a time course in M9 minimal media supplemented with (A) glucose or (B) glucose and malate. Expression from an empty control (BSH461), *yxaD* (BSH462) and *yxaKC* (BSH463) promoter regions. The production of beta-galactosidase (left) or $OD_{600}$ (right) were monitored at 1 hr intervals.

### 4.3.4. Δ*yxaD* (+*yxaKC*) affects excreted metabolites

We explored the possibility that YxaKC were acting as a transporter using NMR-based metabolomics and analyzed the exometabolome of *B. subtilis*. If YxaKC were acting as a transporter, we would expect the transport of specific compound(s) to correlate with *yxaKC* expression. Therefore, we hypothesized that deletions in *yxaKC* would prevent import or export of the compound(s) and that a strain overexpressing *yxaKC* would increase the import or export of the target metabolite(s). Thus, we grew wild type, Δ*yxaKC*, and Δ*yxaD* cultures in M9G and examined supernatants sampled at timepoints using $^{1}$H NMR. Samples were collected at 4 hr and 7 hr to examine excreted metabolites before the inflection point corresponding to increased *yxaKC* expression and during maximal *yxaKC* expression, but before lysis of the culture.

Spectra obtained from secreted metabolites isolated from wild type, Δ*yxaKC*, and Δ*yxaD* were analyzed and compared. No obvious difference in secreted compounds between wild type and Δ*yxaKC* were observed. Conversely, supernatant from Δ*yxaD* (+*yxaKC*) had ~40% less acetoin than wild type and Δ*yxaKC*. Additionally, Δ*yxaD* (+*yxaKC*) supernatant possessed three additional peaks; a triplet resonance at 0.83 ppm, and two singlets at 1.46 ppm and 2.26 ppm (Figure 4.5). The singlet resonances at 1.46 and 2.26 are consistent with 2-acetolactate (2-ACL) (Nemeria *et al.*, 2005). The triplet at 0.83 appears to be consistent with 2-hydroxybutyrate, but further analysis will be required to confirm the identify if this molecule. Although we did not detect a difference in 2-ACL between wild type and Δ*yxaKC* mutants, as would be expected, we attribute this to the low levels of 2-ACL detected. Taken together, these data suggest that YxaKC affect the transport and/or production of metabolites, particularly acetoin and 2-ACL.

These results are consistent with hypothesized roles of Cid/Lrg proteins in regulated cellular homeostasis. Additional studies will be required to determine if YxaKC are acting in the same way as PftAB, and directly transporting acetoin and/or 2-ACL or if YxaKC are affecting metabolite excretion by some other mechanism (Charbonnier *et al.*, 2017).

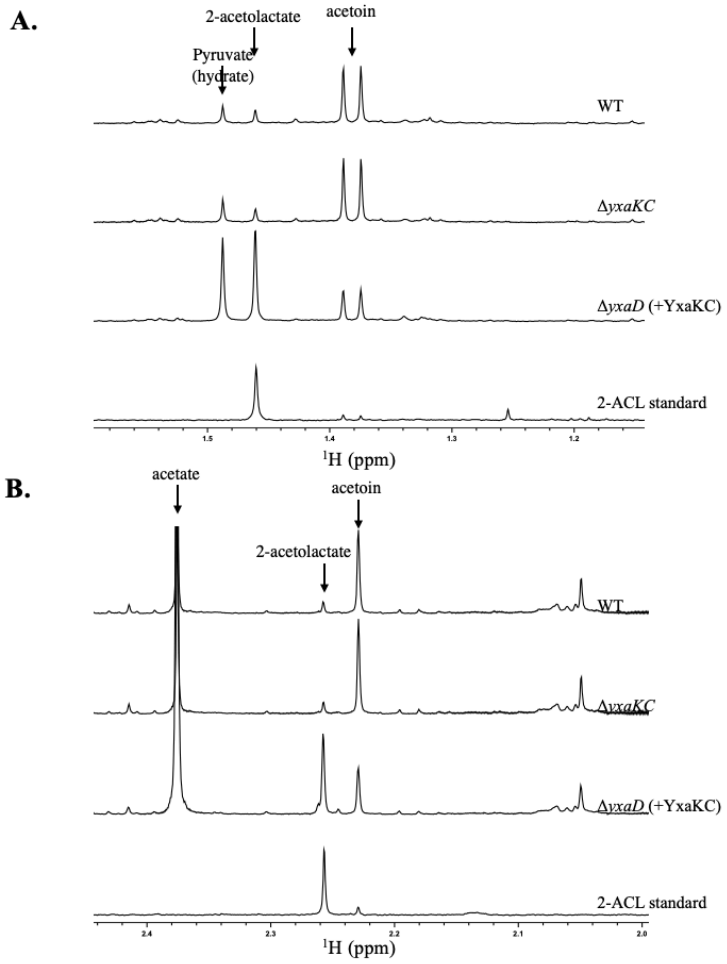



**Figure 4.5 1H NMR spectra of excreted metabolites.**
Representative $^{1}$H NMR spectra from supernatants collected from wild type, ΔyxaKC, and Δ*yxaD* grown for 7 hr in M9G, and the 2-ACL standard. Ranges of spectra shown are between (A) 1.2-1.6 ppm (B) 2.0-2.4 ppm.

### 4.3.5. *yxaD* and *yxaKC* are expressed in the center of biofilms

Recently, it was found that the *B. subtilis*-encoded Cid/Lrg homologs, *ywbHG*, *pftAB* and *yxaKC*, were induced in the presence of acetic acid, which lead to the induction of biofilm formation, and deletions in all three genes resulted in decreased pellicle robustness (Chen *et al.*, 2015). These observations led the authors to speculate a role for YwbHG, PftAB, and YxaKC in regulating proper timing of biofilm formation (Chen *et al.*, 2015). Yet, a large-scale gene expression study that analyzed gene expression from *B. subtilis* grown in over 100 growth conditions, did not detect *yxaKC* expression during biofilm formation (Nicolas *et al.*, 2012). This observation could be consistent with *yxaKC* functioning early in biofilm formation, before the 24 hr sample collection time from the study. Another possible reason for not detecting *yxaKC* expression in this study would be if *yxaKC* appears to only be expressed a subset of the cells in the biofilm population; this could result in it being below the detection limit.

To examine *yxaKC* and *yxaD* expression during biofilm formation, we constructed the same $P_{yxaKC}$-*lacZ* and $P_{yxaD}$-*lacZ* reporters as before in a strain capable of forming biofilms, *B. subtilis* 3610, and grew the resulting strain on the biofilm-inducing medium, MSgg containing X-gal. Within two days, $P_{yxaD}$-*lacZ* and more so, $P_{yxaKC}$-*lacZ*, is expressed in a concentric ring around the center of the biofilm (Figure 4.6). Moreover, after 72 hr, *yxaD* and *yxaKC* expression spreads outward with biofilm growth. Expression of $P_{yxaD}$-*lacZ* and $P_{yxaKC}$-*lacZ* in a strain containing Δ*yxaD* reveals expression throughout the entire biofilm indicating that cells outside of the biofilm are capable of *yxaD* and *yxaKC* expression. Thus, the pattern of expression appears to be a function of YxaD-dependent regulation. Taken together, these data are consistent with differential

activation of both *yxaD* and *yxaKC* in a subset of cells comprising the biofilm structure. Biofilm centers are much more limited in access to nutrients than the outsides (Liu *et al.*, 2015) coinciding with differential gene expression and developmental programs throughout the biofilm as well (Vlamakis *et al.*, 2008). Therefore, these data suggest that *yxaKC* and *yxaD* expression exclusive to biofilm centers are a result of or due to nutrient limitation.

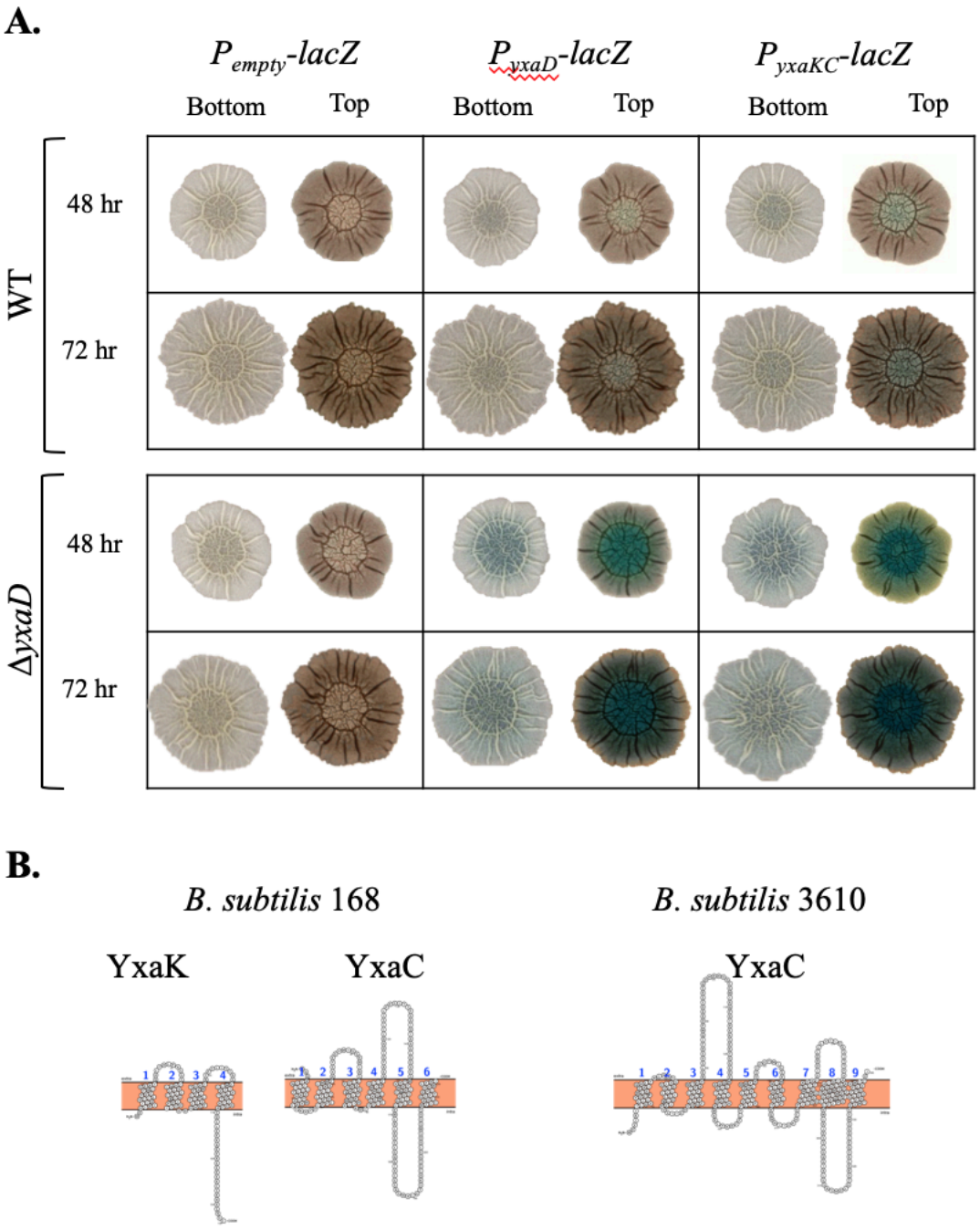



**Figure 4.6 Expression of *yxaC* and *yxaD* during biofilm formation.**
(A) *B. subtilis* 3610 strains capable of biofilm formation expressing $P_{empty}$-*lacZ* (BSH283), $P_{yxaD}$-*lacZ* (BSH294), or $P_{yxaC}$-*lacZ* (BSH295) harboring wild type *yxaD* (TOP) or Δ*yxaD* (BOTTOM) spotted on MSgg media with 100 µg/ml X-gal. Images were taken at 48 hr and 72 hr. (B) Predicted transmembrane regions for *B. subtilis* 168 YxaK and YxaC and *B. subtilis* 3610 encoded YxaC.

### 4.4. Discussion

Cid/Lrg proteins encoded by *S. aureus* and *S. mutans* are thought to control cell lysis in response to environmental changes (Bayles, 2007, Rice & Bayles, 2008). *B. subtilis* PftAB was recently found to transport pyruvate (Charbonnier *et al.*, 2017). Consistent with what was previously published, we observed a defect in growth of *pftAB* lacking cells grown on minimal media agar plates containing pyruvate as the sole carbon source. Since *pftAB* was not essential for growth on pyruvate, we explored the possibility that the other encoded *cid/lrg* homologs, *ywbHG* and *yxaKC*, were needed for growth on pyruvate. We observed that *ywbHG* and/or *yxaKC* were not required for growth. The observation that *ywbHG* did not affect growth on pyruvate is consistent with *ywbHG* and *pftAB* having differential expression patterns in response to glucose, and likely, different roles in the cell (Chen *et al.*, 2015, van den Esker *et al.*, 2017, Charbonnier *et al.*, 2017). In addition, we examined growth of combinations of Δ*pftAB*, Δ*ywbHG*, and Δ*yxaKC* on glucose, glycerol, and gluconate. Expression of *cid/lrg* operons and their lytic behavior in *S. aureus* and *S. mutans* is dependent on glucose (Kim *et al.*, 2019). Since we did not observe any growth phenotype for any combination of *cid/lrg*-like proteins in *B. subtilis*, we concluded that *pftAB*, *ywbHG* and *yxaKC* were not required for growth on glucose.

YwbHG expression increased when grown on glycerol (Chen *et al.*, 2015). Although this was likely due to glycerol eventually being converted into acetate which induces *ywbHG*, we were curious if there was a growth requirement for the *cid/lrg* mutants on glycerol. However, we did not observe any growth defects on glycerol, suggesting that *pftAB*, *ywbHG* and *yxaKC* do not play a direct role in glycerol uptake and utilization. *yxaD* and *yxaKC* genes are located near the gluconate utilization operon,

*gntRKPZ* and genes conserved in synteny have been shown to affect similar processes. Therefore, we examined growth of the Δ*pftAB*, Δ*ywbHG*, and Δ*yxaKC* on minimal media with gluconate as a sole carbon source and found *B. subtilis cid/lrg* genes were not required for gluconate. Although we did not observe a growth requirement for *pftAB*, *ywbHG*, or *yxaKC* on various carbon sources, we attribute this to the hypothesis that the *cid/lrg* proteins in *B. subtilis* affect metabolic flux under the harsh conditions cells may encounter outside of standard laboratory growth.

We observed *yxaKC*, and to a lesser extent, *yxaD* expression in M9G. Interestingly, expression of *yxaD* was much higher when grown in LB-G than in M9G. In contrast, expression of *yxaKC* was lower in LB-G than in M9-G. Therefore, we hypothesize that *yxaD* and/or *yxaKC* are regulated by an additional factor that is currently unknown. Although the expression pattern of *yxaKC* is consistent with catabolite repression by CcpA, we did not identify a canonical CcpA binding site (CRE) at or near the *yxaKC* promoter. In addition, genome-wide studies on CcpA binding sites have been conducted in *B. subtilis* and no CcpA regulation on *yxaKC* or *yxaD* has been observed (Moreno *et al.*, 2001). Our $^{1}$H NMR data indicated higher levels of 2-ACL in Δ*yxaD* (+*yxaKC*) cells, revealing a possible role of YxaKC as a transporter. We did not detect a substantial difference in 2-ACL between wild type and Δ*yxaKC* which could be explained if another 2-ACL transporter exists or if YxaKC do not transport 2-ACL directly, but influence its transport. We also suspect that the premature expression of *yxaKC* resulting from Δ*yxaD*, could subsequently lead to 2-ACL secretion concomitant with its synthesis. This would be expected to lead to enhanced accumulation of

extracellular 2-ACL at 7 hr. It is also possible *B. subtilis yxaKC*-overexpressing cells are excreting 2-ACL as soon as it is made allowing it to accumulate.

When grown in excess glucose, *B. subtilis* produces potentially detrimental concentrations of acetate. To reduce intra-and extracellular acidification, some of the pyruvate produced from glycolysis is irreversibly converted into 2-ACL which is subsequently metabolized into acetoin and $CO_2$ or diacetyl and $CO_2$. 2-ACL can spontaneously decarboxylate into diacetyl the rate of which is increased as pH decreases (below pH 6) temperature increases (Suomalainen & Ronkainen, 1968). Intracellular diacetyl can lead to reactive oxygen species (Kovacic & Cooksy, 2005, Wondrak *et al.*, 2002). Due to the detrimental effects that can result from increased intracellular diacetyl, 2-ACL transporters might exist to export 2-ACL before being spontaneously decarboxyated to toxic diacetyl. Additional experimentation is required to confirm that 2-ACL is transported and to determine if YxaKC is responsible for its transport.

# 5. CONCLUSIONS

## 5.1. Misexpression of YxaD

YxaD was identified in a misexpression screen for uncharacterized DNA-binding proteins involved in subcellular organization in *B. subtilis*. When YxaD is misexpressed, cells lose their bilobed nucleoid structure and divide, producing anucleate daughter cells. Growth continues without segregation of replication origins or a concomitant increase in nucleoids, suggesting that YxaD misexpression leads to a defect in both DNA replication and chromosome segregation.

Interestingly, YxaD was found to interact with ScpA by Y2H (Marchadier *et al.*, 2011), but *scpA* is not required for the YxaD misexpression phenotype, as a Δ*scpA* strain is still sensitive to *yxaD* misexpression. YxaD and HolA (the delta clamp loader of DNA polymerase) also interact by Y2H (Marchadier *et al.*, 2011), but this interaction could not be recapitulated in a B2H assay. Attempts to identify stable, extragenic suppressor mutations to YxaD misexpression were not successful; however, suppressor selections were used to identify intergenic mutations that resulted in loss of function phenotypes.

Suppressor selections of similar design were used to identify extragenic genetic targets of RefZ, SirA, YodL and YisK in *ftsZ*, *dnaA*, *mreB* and *mbl*, respectively (Wagner-Herman *et al.*, 2012, Duan *et al.*, 2016a, Duan *et al.*, 2016b). One reason we may have been unable to obtain YxaD suppressors may be that YxaD has more than one essential target, supported by Y2H data (Marchadier *et al.*, 2011). If YxaD interacts with ScpA and HolA as indicated by Y2H (Marchadier *et al.*, 2011), then multiple mutations may be required to obtain resistance. In addition, if YxaD were to target an essential

gene, it is possible that mutations rendering the target resistant to YxaD misexpression would compromise the target's function, and therefore would not be isolated by our screen. One future consideration would be to perform suppressor selections using a chemical mutagen in order to increase the change of obtaining relatively rare mutations. YxaD misexpression appears to disrupt DNA replication and chromosome segregation, perhaps through ScpA and/or HolA. An alternative to identify the YxaD target would be to perform pull-downs with 6His-YxaD, which is fully functional based on its ability to bind DNA *in vitro* and *in vivo* and the observation that it retains killing capability when misexpressed (Chapter 3).

### 5.2. DNA-binding activity of YxaD

YxaD binds to two sites in the intergenic region of *yxaD* and *yxaKC* to repress expression of both genes. Most of the studied MarR-like proteins bind 12-18 bp motifs consisting of one IR, usually 4-7 bp, separated by 2-5 bp (Wilkinson & Grove, 2006). Interestingly, we identified two putative YxaD-binding sites that contain two overlapping inverted repeats of 5 (or 6) and 5 bp separated by 2 (or 1) and 2 bp (Chapter 2). Such overlapping IRs could be a novel binding preference for MarR-like proteins or potentially the binding site of another regulator, for instance, GntR. GntR is known to regulate the gluconate utilization operon encoded by *B. subtilis*, *gntRKPZ*, by binding to the half site ATACTTGTA (Fujita & Fujita, 1986, Fujita & Miwa, 1989). Surprisingly, YxaD's predicted binding sites contain a similar GntR half site. Further analysis is required to define the YxaD recognition sequence and to determine if YxaD recognizes both half sites. One future direction is to mutate each IR in the binding sites of YxaDand

perform EMSAs as in Chapter 3 to determine if YxaD recognizes the entire sequene or only a half site. In addition, one could use ChIP-seq on wild type and cells overexpressing YxaD in order to more accurately define the DNA sequences YxaD binds when overexpressed.

### 5.3. YxaKC in overflow metabolism

YxaD regulates the expression of *yxaD* and the divergent operon, *yxaKC*, encoding a Cid/Lrg-like operon (Chapter 4). *B. subtilis* encodes three Cid/Lrg homologs: *pftAB*, *ywbHG* and *yxaKC*. Recently, PftAB were found to transport pyruvate (Charbonnier *et al.*, 2017) suggesting a possible role for other Cid/Lrg homologs in transport. In order to test if YxaKC were functioning as a transporter, we analyzed supernatants from wild type, Δ*yxaKC* and Δ*yxaD* by $^{1}$H NMR. Since a strain lacking *yxaD* results in derepressed *yxaKC*, Δ*yxaD* was used to examine overexpression of YxaKC (+YxaKC). While wild type and Δ*yxaKC* revealed no differences in extracellular metabolites, Δ*yxaD* (+YxaKC) secreted less acetoin than wild type and Δ*yxaKC*. In addition, Δ*yxaD (*+YxaKC) had three additional peaks, a triplet resonance at 0.83 ppm, and two singlets at 1.46 ppm and 2.26 ppm, that were higher in Δ*yxaD (*+YxaKC) than wild type and Δ*yxaKC*. The singlets at 1.46 ppm and 2.26 match peaks for 2-acetolactate (2-ACL) while the triplet at 0.83 ppm could not be identified.

During growth in excess glucose, *B. subtilis* produces an abundant amount of acetate, which can be secreted. Intracellular acetate is converted to pyruvate to reduce intracellular acidification. Pyruvate can be secreted, or to reduce intracellular acidification further, pyruvate is irreversibly converted into 2-ACL by $\alpha$-acetolactate

synthetase (AlsS). Subsequently, 2-ACL is metabolized into acetoin and $CO_2$ or diacetyl and $CO_2$ by acetolactate decarboxylase (AlsD)(Loken & Stormer, 1970) (Figure 5.1). In addition to overflow metabolism, pyruvate is converted to 2-ACL by IlvBH for synthesis of branched chain amino acids (BCAA). In total, *B. subtilis* encodes two acetolactate synthase enzymes: AlsS (Holtzclaw & Chapman, 1975)and IlvB (large subunit) and IlvH (small subunit). Interestingly, in a *B. subtilis* expression study that examined gene expression in over 100 growth conditions, *alsS* and *yxaKC* were both highly expressed at stationary and transition in LB supplemented with glucose (Nicolas *et al.*, 2012).

The additional triplet resonance at 0.83 ppm was difficult to assign as the low concentration prevented the determination of how many additional peaks corresponded to the compound of interest. However, we were able to deduce that a triplet peak at 0.83 ppm likely corresponded to a methyl group in the formation $CH_3$-$CH_2$-R. Interestingly, the 0.83 ppm peak could be consisted with 2-aceto-2-hydroxybutanoate (2ACH) which is synthesized by condensing one molecule of pyruvate and one molecule of $\alpha$-ketobutyrate, performed by IlvBH, for isoleucine biosynthesis (Figure 5.1).

### 5.3.1. Excretion of 2-ACL

The observation that the Δ*yxaD* (+YxaKC) strain excretes more 2-ACL (and perhaps more 2-ACH) (Figure 4.5) suggests YxaKC directly transports 2-ACL or influences its transport. The observation that YxaKC could be a small molecule transporter is in line with the recent finding that *B. subtilis* encoded PftAB transports pyruvate (Charbonnier *et al.*, 2017).

In addition to 2-ACL, strains carrying Δ*yxaD* (+YxaKC) also excreted less acetoin than wild type. If YxaKC were transporting 2-ACL, then YxaKC overexpression could lead to: 1) early excretion of 2-ACL relative to wild type and Δ*yxaKC* and/or 2) higher than normal levels of excreted 2-ACL. Early excretion of 2-ACL would deplete the intracellular pool of 2-ACL, resulting in less acetoin production and thus, less acetoin excretion. Similar findings were observed with pyruvate transport as a strain overexpressing PftAB led to earlier and more abundant excretion of pyruvate (Charbonnier *et al.*, 2017). Alternatively, YxaKC could import acetoin, in which case, excess 2-ACL might be excreted in the media. As similar levels of acetoin were present in the supernatants of wild type and the Δ*yxaKC* mutant (Figure 4.5) suggests YxaKC is not a transporter of acetoin. In addition, we acknowledge that YxaKC could be influencing 2-ACL levels by affecting import/export of another transporter. For example, YxaKC could affect metabolic flux by transporting a compound we did not detect or YxaKC could interact directly with another transporter to inhibit or activate. Based on the recent discoveries implicating Cid/Lrg homologs in pyruvate transport, we hypothesize that YxaKC function as a transporter of 2-ACL (Charbonnier *et al.*, 2017). In *Saccharomyces cerevisiae*, conversion of 2-ACL to 2,3-dihydroxyisovalerate (a precursor for valine synthesis) is the rate-limiting step in valine synthesis (Gibson *et al.*, 2015, Krogerus, 2013). Secretion of 2-ACL has been proposed as a method for *S. cerevisiae* to protect itself from carbonyl stress (van Bergen *et al.*, 2006, Kosmachevskaya *et al.*, 2015)(see below). It is likely that bacteria are susceptible to detrimental effects of carbonyl stress as well. Thus, if YxaKC is a 2-ACL transporter, it could have evolved as a way to minimize such stress.

### 5.3.2. Toxic accumulation of diacetyl

In addition to conversion by enzymatic reactions, 2-ACL can spontaneous decarboxylate at low pH to diacetyl or undergo oxidative decarboxylation at neutral pH forming diacetyl (Park *et al.*, 1995, Suomalainen & Ronkainen, 1968). Diacetyl is an $\alpha$-dicarbonyl, similar to methylglyoxal and glyoxal, a class of reactive carbonyl species (RCS) (Kovacic & Cooksy, 2005, Wondrak *et al.*, 2002). RCS consist of aldehydes and ketones possessing a carbonyl group that provides an electrophilic carbon that are capable of reacting with the nucleophilic nitrogens of amino acids, peptides and guanine bases (Kosmachevskaya *et al.*, 2015). $\alpha$,β-dicarbonyls, are capable of reacting with arginine, lysine, and cysteine residues damaging proteins by forming cross-links. In addition to carbonyl stress, RCS can increase oxidative stress leading to reactive oxygen species (ROS). In humans, RCS generate advanced glycation end products (AGEs) which have been linked to diabetes, cancer, and neurodegenerative diseases (Kosmachevskaya *et al.*, 2015). Although much less information is known about glycation in prokaryotes, carbonyl compounds have been shown to be produced in bacteria and glyoxylase systems that detoxify $\alpha$-ketoaldehydes have been characterized (Sukdeo & Honek, 2008). *B. subtilis* does not encode a system for glyoxylate removal, suggesting additional mechanisms are in place to deal with carbonyl and oxidative stress. Taken together, we hypothesize that, similar to *S. cerevisiae*, *B. subtilis* secretes 2-ACL to reduce the potentially toxic effects from intracellular carbonyl stress (Gibson *et al.*, 2015, Milne *et al.*, 2016).

In addition to the potentially toxic effects of diacetyl formation, the spontaneous decarboxylation of 2-ACL has also been shown to reduce various flavins, nicotinamide

and quinone coenzymes (Park *et al.*, 1995). Intriguingly, one group found that *B. subtilis* 3610 carrying a transposon insertion between *yxaD* and *yxaC* requires menaquinone to form complex colonies on spreading agar plates (Pelchovich *et al.*, 2013). Unfortunately, for unknown reasons, we were unable to observe consistent growth rates in the published medium. Nevertheless, a requirement for menaquinone in development of complex colonies is intriguing given the possible role of YxaC in 2-ACL secretion. In a strain harboring Δ*yxaC*, cells might be unable or less able to secrete 2-ACL into the medium. If so, the accumulated intracellular 2-ACL could undergo spontaneous decarboxylation resulting in reduction of flavins, quinones, and nicotinamide coenzymes, which could explain the requirement for exogenous menaquinone (Pelchovich *et al.*, 2013).

### 5.4. A possible mechanism for YxaD derepression

We observed 2-ACL excretion from *B. subtilis* in a Δ*yxaD* mutant. MarR-like proteins are involved in regulating a variety of processes including antibiotic resistance, metabolism, and responses to oxidative stress. YxaKC could function to regulate oxidative stress by managing intracellular 2-ACL concentration. If so, *yxaKC* expression might be regulated by YxaD responding to 2-ACL levels or some form of oxidative stress. Commonly, MarR-regulated oxidative stress operons are induced by oxidation of redox-active cysteine residues, which subsequently causes release of the DNA-binding protein from DNA (Deochand & Grove, 2017). This cannot be the mechanism for YxaD, however, as YxaD has no cysteine residues. Although less common, another possible oxidation sensor is methionine, which has been shown for the LysR-like regulator in

*E.coli*, HypT (Drazic *et al.*, 2013). YxaD possesses three methionine residues, one that maps to $\alpha$5 and two that map to $\alpha$6 in the predicted structure (Figure 4.1).

MarR-like proteins form a 'safety-triangle' shape where the N- and C-termini of both monomers meet at the top burying $\alpha$1 (Figure 4.1). In the case of OhrR, oxidation of the lone cysteine on $\alpha$1 forms Cys-sulfenic acid (Cys-SOH) which disrupts DNA-binding due to steric clash by two tyrosine residues on $\alpha$1 (Hong *et al.*, 2005). *E.coli* MarR undergoes a completely different mechanism where the oxidation of cysteine positioned on $\alpha$2 near the DNA-binding domain leads to disulfide bond formation followed by tetramerization of two dimers (Hao *et al.*, 2014). Since no structures of YxaD have been solved, we are not able to define the positioning of the methionine residues or any local residues that could facilitate an allosteric effect. Further experimentation will be required to determine if the YxaD DNA-binding activity is affected by oxidation. First, expression levels of *yxaKC* and *yxaD* in response to various oxidative stressors must be analyzed in order to determine if YxaD responds to oxidative stress *in vivo*. In addition, solving the structure of YxaD in the absence and presence of DNA would reveal the residues required for DNA binding and would reveal the position of methionine residues as well as how oxidation may or may not play a role in DNA-binding of YxaD. To our knowledge, no MarR protein has been shown to be regulated by methionine oxidation. However, this does not exclude the possibility of a novel mechanism for MarR-like regulation. Prior to the recent finding that *E. coli* MarR is regulated by cysteine oxidation, it was long thought that DNA-binding of *E. coli* MarR was affected by directly binding phenolic compounds such as salicylate (Zhu *et al.*, 2017b, Hao *et al.*, 2014). Transcriptomic and proteomic analyses of *B. subtilis*

performed in the presence of superoxide or peroxide stress did not significantly affect *yxaD* or *yxaKC* expression (Mostertz *et al.*, 2004). This suggested that the genes are not expressed in response to oxidative stress or that their activation is specific to a particular stressor or compound.

Another possibility is that YxaD is regulated by ligand binding, possibly 2-ACL or a related compound in the same pathway. Although most MarR-like proteinss are thought to be regulated by binding phenolic ligands, YxaD regulation by 2-ACL might be similar to the allosteric regulation of the MarR-like protein TamR of *Streptomyces coelicolor*, by binding *trans*-aconitate, cis-aconitate, citrate and isocitrate (Huang & Grove, 2013).

### 5.5. Possible additional *yxaKC* regulation mechanisms

Expression of *yxaD* and *yxaKC* in a Δ*yxaD* background resulted a gradual decrease in overall expression during growth in LB medium (Figure 3.9) which raises the question of whether *yxaD* and *yxaKC* are regulated by an additional regulator. In addition, we observed that Δ*yxaD* secreted more 2-ACL than wild type or Δ*yxaKC* in M9G medium (Figure 4.5). 2-ACL secretion by *S. cerevisiae* is believed to be the result of the rate-limiting step consisting of 2-ACL to 2,3-dihydroxyisovalerate during valine synthesis (Gibson *et al.*, 2015). BCAA synthesis in *B. subtilis* is under tight control by the global regulators CodY (Shivers & Sonenshein, 2004) and CcpA (Fujita *et al.*, 2014). Therefore, one possibility is that the regulation of *yxaKC* and *yxaD* could be mediated by CodY and/or CcpA (Gibson *et al.*, 2015). In support of this, CcpA exhibits tight control

over the *B. subtilis* pyruvate transporter PftAB which is repressed in the presence of glucose (van den Esker *et al.*, 2017, Charbonnier *et al.*, 2017).

Genome-wide CodY binding studies have been performed on *B. subtilis* and no CodY binding sites have been found at or near *yxaD* and *yxaKC* promoters (Belitsky & Sonenshein, 2013), suggesting that CodY likely does not directly regulate *yxaD* or *yxaKC*. A more recent study examined global gene expression under different concentrations of CodY and in CodY mutants (Brinsmade *et al.*, 2014). Interestingly, *yxaK* and *yxaC* were overexpressed by nearly 2-fold in a Δ*codY* strain (Brinsmade *et al.*, 2014) suggesting *yxaKC* are at least indirectly upregulated in the absence of *codY*. In addition, this study revealed similar regulation amongst *yxaC* and *ilvD*, *ilvA*, and *ilvBH*. This suggests the proteins may function in a common pathway. We also do not exclude the possibility that CodY could regulate *yxaKC* by directly binding to promoter DNA. Such an interaction might not have been observed before because of the conditions used in the *in vitro* assay (Belitsky & Sonenshein, 2013). Recently, it was found that CcpA regulates *cidAB* through a cryptic CcpA binding site in the *cidAB* promoter in *S. mutans* (Kim *et al.*, 2019). In addition, *B. cereus ClhAB*$_2$ was recently shown to require CodY for expression in glucose (Huillet *et al.*, 2017). *B. cereus ClhAB*$_2$ is most closely related to *B. subtilis yxaKC*, but *B. cereus* does not encode a nearby MarR-like regulator, like *yxaD*. Interestingly, since *clhAB*$_2$ and *yxaKC* are functional homologs, both could be regulated by CodY, but *yxaKC* may have evolved YxaD as a more specific regulator in a common ancestor.

How and why might a Δ*codY* strain lead to *yxaKC* overexpression? Interestingly, the only amino acids that have been shown to regulate CodY activity are branched-chain

amino acids, suggesting the pool of BCAA is critical for nutrient availability and subsequent cell response (Shivers & Sonenshein, 2004, Guedon *et al.*, 2001). The altered regulation of *yxaKC* in Δ*codY*, could be result from altered expression of BCAA genes that may affect pathways producing the ligand of YxaD and/or are transported or affect by YxaKC. Ultimately, further studies are needed to determine if *yxaKC* are regulated by CodY directly and to probe a possible role for YxaKC in BCAA pathways or overflow metabolism.

### 5.6. Functions in the real-world: Biofilm formation

The *B. subtilis* encoded *cid*/*lrg* operons were recently shown to be required for efficient pellicle formation in the *B. subtilis* 3610 (Chen *et al.*, 2015). *yxaD* in *B. subtilis* 3610 shares 100% sequence identity with that in *B. subtilis* 168. On the other hand, *yxaKC* in *B. subtilis* 3610 are expressed as a fusion protein in termed *yxaC*. We found that *yxaD* and *yxaC* are expressed as a concentric ring during biofilm formation which grew outward as the biofilm matured (Figure 4.6). Nutrients are more depleted in biofilm centers compared to the biofilm edge. In the Δ*yxaD* mutant, expression of *yxaD* and *yxaC* was observed throughout the entire biofilm, suggesting that YxaD represses *yxaD* and *yxaC* expression in the outer regions of the biofilm but that depression occurs in the biofilm center. These data are consistent with expression during nutrient depletion and repression during higher nutrient availability. Therefore, we hypothesize that if YxaC exhibits transporter activity, then YxaC transport activity preferentially occurs where nutrients are depleted in order to import metabolite(s). Under nutrient-limited conditions, tight regulation of facilitated transporters would be absolutely necessary in a community

environment such as a biofilm to ensure that vital compounds are not excreted and lost to other bacteria. Alternatively, under nutrient-depleted conditions, such transporters might become necessary for maintaining fitness.

### 5.7. Final Remarks

Here, we utilize misexpression to reveal growth and morphological phenotypes and assign preliminary functions to genes that lack deletion phenotypes under standard laboratory growth conditions. We quantified the expression of the *yxaD* and *yxaKC* genes during growth in LB, M9G and M9MG and found that expression increases with growth, consistent with glucose utilization. In addition, we defined the mechanism to which YxaD represses *yxaD* and *yxaKC* including the site YxaD binds as well as residues that likely facilitate contacts with the DNA. Furthermore, we have identified a role for YxaKC in affecting 2-ACL transport possibly to reduce the chances of 2-ACL mediated oxidative stress by spontaneous decarboxylation to diacetyl and/or to prevent reduction of flavins, nicotinamide and quinone coenzymes (Park *et al.*, 1995). Accumulation of intracellular 2-ACL could result from the rate-limiting step of 2-ACL to 2,3-dihydroxyisovalerate or an additional step in valine biosynthesis (Figure 5.1B) (Gibson *et al.*, 2015). Another possible mechanism is to utilize nutrients under starvation conditions, or to possibly to import and export a signal to communicate with nearby cells or communities (Figure 5.1C).

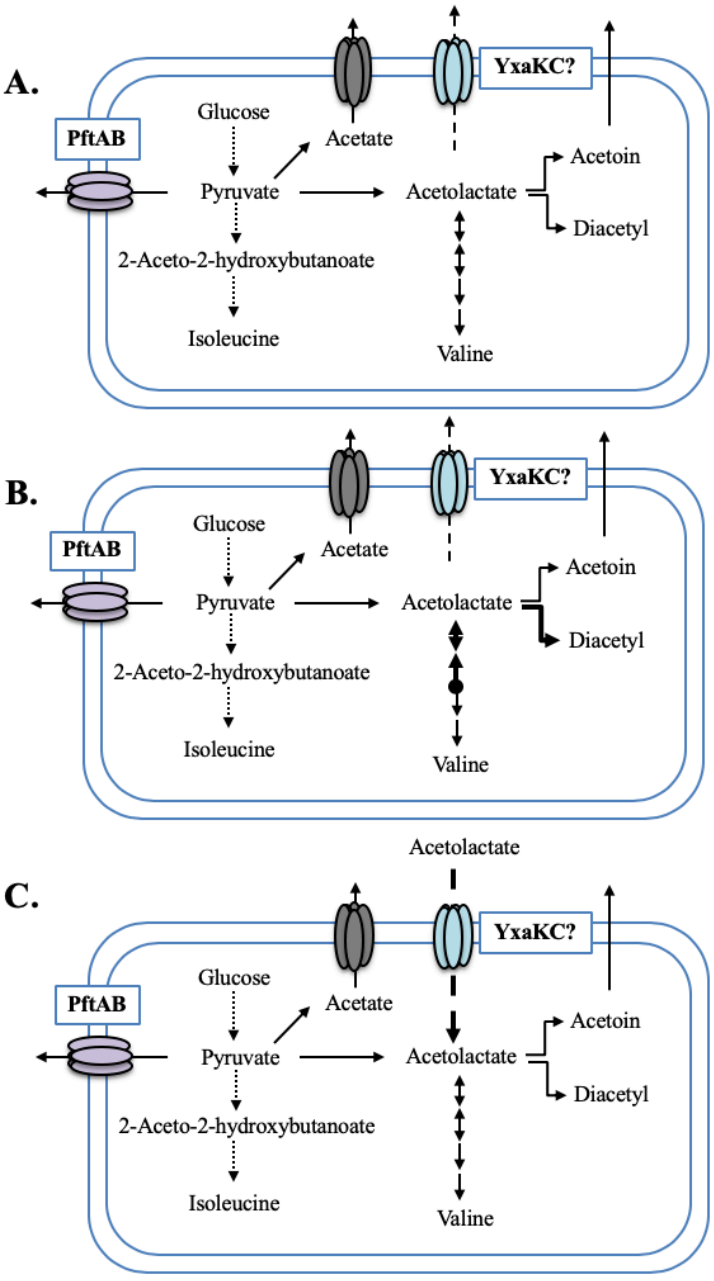



**Figure 5.1 Possible mechanisms for 2-ACL secretion**
(A) During growth in excess glucose, acetate and pyruvate accumulate. In order to reduce intracellular acidification, acetate and pyruvate are exported or converted into overflow metabolites. Similar precursors are required for BCAA synthesis. (B) Rate-limiting reactions from valine synthesis could increase 2-ACL. Accumulated 2-ACL is secreted to prevent ROS and RCS as result of diacetyl formation. (C) Under nutrient-limited conditions, 2-ACL may be imported for BCAA synthesis.

REFERENCES

Ababneh, Q.O., and Herman, J.K. (2015) RelA inhibits Bacillus subtilis motility and chaining. *J Bacteriol* **197**: 128-137.

Ahn, S.J., Gu, T., Koh, J., and Rice, K.C. (2017) Remodeling of the Streptococcus mutans proteome in response to LrgAB and external stresses. *Sci Rep* **7**: 14063.

Ahn, S.J., and Rice, K.C. (2016) Understanding the Streptococcus mutans Cid/Lrg System through CidB Function. *Appl Environ Microbiol* **82**: 6189-6203.

Ahn, S.J., Rice, K.C., Oleas, J., Bayles, K.W., and Burne, R.A. (2010) The Streptococcus mutans Cid and Lrg systems modulate virulence traits in response to multiple environmental signals. *Microbiology* **156**: 3136-3147.

Alekshun, M.N., and Levy, S.B. (1997) Regulation of chromosomally mediated multiple antibiotic resistance: the mar regulon. *Antimicrob Agents Chemother* **41**: 2067-2075.

Alekshun, M.N., and Levy, S.B. (1999) Characterization of MarR superrepressor mutants. *J Bacteriol* **181**: 3303-3306.

Alekshun, M.N., Levy, S.B., Mealy, T.R., Seaton, B.A., and Head, J.F. (2001) The crystal structure of MarR, a regulator of multiple antibiotic resistance, at 2.3 A resolution. *Nat Struct Biol* **8**: 710-714.

Allard-Massicotte, R., Tessier, L., Lecuyer, F., Lakshmanan, V., Lucier, J.F., Garneau, D., Caudwell, L., Vlamakis, H., Bais, H.P., and Beauregard, P.B. (2016) Bacillus subtilis Early Colonization of Arabidopsis thaliana Roots Involves Multiple Chemotaxis Receptors. *MBio* **7**.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990) Basic local alignment search tool. *J Mol Biol* **215**: 403-410.

Alwine, J.C., Kemp, D.J., and Stark, G.R. (1977) Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. *Proc Natl Acad Sci U S A* **74**: 5350-5354.

Ambriz-Avina, V., Contreras-Garduno, J.A., and Pedraza-Reyes, M. (2014) Applications of flow cytometry to characterize bacterial physiological responses. *Biomed Res Int* **2014**: 461941.

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K.A., Tomita, M., Wanner, B.L., and Mori, H. (2006) Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* **2**: 2006 0008.

Babu, M., Diaz-Mejia, J.J., Vlasblom, J., Gagarinova, A., Phanse, S., Graham, C., Yousif, F., Ding, H., Xiong, X., Nazarians-Armavil, A., Alamgir, M., Ali, M., Pogoutse, O., Pe'er, A., Arnold, R., Michaut, M., Parkinson, J., Golshani, A., Whitfield, C., Wodak, S.J., Moreno-Hagelsieb, G., Greenblatt, J.F., and Emili, A. (2011) Genetic interaction maps in Escherichia coli reveal functional crosstalk among cell envelope biogenesis pathways. *PLoS Genet* **7**: e1002377.

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.

Bassett, A.R., Tibbit, C., Ponting, C.P., and Liu, J.L. (2014) Mutagenesis and homologous recombination in Drosophila cell lines using CRISPR/Cas9. *Biol Open* **3**: 42-49.

Bayles, K.W. (2007) The biological role of death and lysis in biofilm development. *Nat Rev Microbiol* **5**: 721-726.

Belitsky, B.R., and Sonenshein, A.L. (2013) Genome-wide identification of Bacillus subtilis CodY-binding sites at single-nucleotide resolution. *Proc Natl Acad Sci U S A* **110**: 7026-7031.

Bernard, R., Marquis, K.A., and Rudner, D.Z. (2010) Nucleoid occlusion prevents cell division during replication fork arrest in Bacillus subtilis. *Mol Microbiol* **78**: 866-882.

Bernhardt, T.G., and de Boer, P.A. (2005) SlmA, a nucleoid-associated, FtsZ binding protein required for blocking septal ring assembly over Chromosomes in E. coli. *Mol Cell* **18**: 555-564.

Bingol, K. (2018) Recent Advances in Targeted and Untargeted Metabolomics by NMR and MS/NMR Methods. *High Throughput* **7**.

Blum, M., De Robertis, E.M., Wallingford, J.B., and Niehrs, C. (2015) Morpholinos: Antisense and Sensibility. *Dev Cell* **35**: 145-149.

Bork, P. (2000) Powers and pitfalls in sequence analysis: the 70% hurdle. *Genome Res* **10**: 398-400.

Branda, S.S., Gonzalez-Pastor, J.E., Ben-Yehuda, S., Losick, R., and Kolter, R. (2001) Fruiting body formation by Bacillus subtilis. *Proc Natl Acad Sci U S A* **98**: 11621-11626.

Brandner, C.J., Maier, R.H., Henderson, D.S., Hintner, H., Bauer, J.W., and Onder, K. (2008) The ORFeome of Staphylococcus aureus v 1.1. *BMC Genomics* **9**: 321.

Bray, N.L., Pimentel, H., Melsted, P., and Pachter, L. (2016) Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* **34**: 525-527.

Brettin, T., Altherr, M.R., Du, Y., Mason, R.M., Friedrich, A., Potter, L., Langford, C., Keller, T.J., Jens, J., Howie, H., Weyand, N.J., Clary, S., Prichard, K., Wachocki, S., Sodergren, E., Dillard, J.P., Weinstock, G., So, M., and Arvidson, C.G. (2005) Expression capable library for studies of Neisseria gonorrhoeae, version 1.0. *BMC Microbiol* **5**: 50.

Brinsmade, S.R., Alexander, E.L., Livny, J., Stettner, A.I., Segre, D., Rhee, K.Y., and Sonenshein, A.L. (2014) Hierarchical expression of genes controlled by the Bacillus subtilis global regulatory protein CodY. *Proc Natl Acad Sci U S A* **111**: 8227-8232.

Brinsmade, S.R., and Sonenshein, A.L. (2011) Dissecting complex metabolic integration provides direct genetic evidence for CodY activation by guanine nucleotides. *J Bacteriol* **193**: 5637-5648.

Brochado, A.R., and Typas, A. (2013) High-throughput approaches to understanding gene function and mapping network architecture in bacteria. *Curr Opin Microbiol* **16**: 199-206.

Brooks, B.E., Piro, K.M., and Brennan, R.G. (2007) Multidrug-binding transcription factor QacR binds the bivalent aromatic diamidines DB75 and DB359 in multiple positions. *J Am Chem Soc* **129**: 8389-8395.

Camara, J.E., Breier, A.M., Brendler, T., Austin, S., Cozzarelli, N.R., and Crooke, E. (2005) Hda inactivation of DnaA is the predominant mechanism preventing hyperinitiation of Escherichia coli DNA replication. *EMBO Rep* **6**: 736-741.

Caspi, R., Altman, T., Billington, R., Dreher, K., Foerster, H., Fulcher, C.A., Holland, T.A., Keseler, I.M., Kothari, A., Kubo, A., Krummenacker, M., Latendresse, M., Mueller, L.A., Ong, Q., Paley, S., Subhraveti, P., Weaver, D.S., Weerasinghe, D., Zhang, P., and Karp, P.D. (2014) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Res* **42**: D459-471.

Chakravorty, S., and Hegde, M. (2018) Inferring the effect of genomic variation in the new era of genomics. *Hum Mutat* **39**: 756-773.

Chandramohan, L., Ahn, J.S., Weaver, K.E., and Bayles, K.W. (2009) An overlap between the control of programmed cell death in Bacillus anthracis and sporulation. *J Bacteriol* **191**: 4103-4110.

Chang, Y.M., Ho, C.H., Chen, C.K., Maestre-Reyna, M., Chang-Chien, M.W., and Wang, A.H. (2014) TcaR-ssDNA complex crystal structure reveals new DNA binding mechanism of the MarR family proteins. *Nucleic Acids Res* **42**: 5314-5321.

Charbonnier, T., Le Coq, D., McGovern, S., Calabre, M., Delumeau, O., Aymerich, S., and Jules, M. (2017) Molecular and Physiological Logics of the Pyruvate-Induced Response of a Novel Transporter in Bacillus subtilis. *MBio* **8**.

Chaudhari, S.S., Thomas, V.C., Sadykov, M.R., Bose, J.L., Ahn, D.J., Zimmerman, M.C., and Bayles, K.W. (2016) The LysR-type transcriptional regulator, CidR,

regulates stationary phase cell death in Staphylococcus aureus. *Mol Microbiol* **101**: 942-953.

Chen, Y., Gozzi, K., Yan, F., and Chai, Y. (2015) Acetic Acid Acts as a Volatile Signal To Stimulate Bacterial Biofilm Formation. *MBio* **6**: e00392.

da Silva, R.R., Dorrestein, P.C., and Quinn, R.A. (2015) Illuminating the dark matter in metabolomics. *Proc Natl Acad Sci U S A* **112**: 12549-12550.

Daigle, D.M., Cao, L., Fraud, S., Wilke, M.S., Pacey, A., Klinoski, R., Strynadka, N.C., Dean, C.R., and Poole, K. (2007) Protein modulator of multidrug efflux gene expression in Pseudomonas aeruginosa. *J Bacteriol* **189**: 5441-5451.

Denome, S.A., Elf, P.K., Henderson, T.A., Nelson, D.E., and Young, K.D. (1999) Escherichia coli mutants lacking all possible combinations of eight penicillin binding proteins: viability, characteristics, and implications for peptidoglycan synthesis. *J Bacteriol* **181**: 3981-3993.

Deochand, D.K., and Grove, A. (2017) MarR family transcription factors: dynamic variations on a common scaffold. *Crit Rev Biochem Mol Biol* **52**: 595-613.

Deochand, D.K., Meariman, J.K., and Grove, A. (2016) pH-Dependent DNA Distortion and Repression of Gene Expression by Pectobacterium atrosepticum PecS. *ACS Chem Biol* **11**: 2049-2056.

Dervyn, E., Noirot-Gros, M.F., Mervelet, P., McGovern, S., Ehrlich, S.D., Polard, P., and Noirot, P. (2004) The bacterial condensin/cohesin-like protein complex acts in DNA repair and regulation of gene expression. *Mol Microbiol* **51**: 1629-1640.

Dolan, K.T., Duguid, E.M., and He, C. (2011) Crystal structures of SlyA protein, a master virulence regulator of Salmonella, in free and DNA-bound states. *J Biol Chem* **286**: 22178-22185.

Drazic, A., Miura, H., Peschek, J., Le, Y., Bach, N.C., Kriehuber, T., and Winter, J. (2013) Methionine oxidation activates a transcription factor in response to oxidative stress. *Proc Natl Acad Sci U S A* **110**: 9493-9498.

Duan, Y., Huey, J.D., and Herman, J.K. (2016a) The DnaA inhibitor SirA acts in the same pathway as Soj (ParA) to facilitate oriC segregation during Bacillus subtilis sporulation. *Mol Microbiol* **102**: 530-544.

Duan, Y., Sperber, A.M., and Herman, J.K. (2016b) YodL and YisK Possess Shape-Modifying Activities That Are Suppressed by Mutations in Bacillus subtilis mreB and mbl. *J Bacteriol* **198**: 2074-2088.

Elbourne, L.D., Tetu, S.G., Hassan, K.A., and Paulsen, I.T. (2017) TransportDB 2.0: a database for exploring membrane transporters in sequenced genomes from all domains of life. *Nucleic Acids Res* **45**: D320-D324.

Ellens, K.W., Christian, N., Singh, C., Satagopam, V.P., May, P., and Linster, C.L. (2017) Confronting the catalytic dark matter encoded by sequenced genomes. *Nucleic Acids Res* **45**: 11495-11514.

Fujita, Y. (2009) Carbon catabolite control of the metabolic network in Bacillus subtilis. *Biosci Biotechnol Biochem* **73**: 245-259.

Fujita, Y., and Fujita, T. (1986) Identification and nucleotide sequence of the promoter region of the Bacillus subtilis gluconate operon. *Nucleic Acids Res* **14**: 1237-1252.

Fujita, Y., and Miwa, Y. (1989) Identification of an operator sequence for the Bacillus subtilis gnt operon. *J Biol Chem* **264**: 4201-4206.

Fujita, Y., Satomura, T., Tojo, S., and Hirooka, K. (2014) CcpA-mediated catabolite activation of the Bacillus subtilis ilv-leu operon and its negation by either CodY- or TnrA-mediated negative regulation. *J Bacteriol* **196**: 3793-3806.

Gelperin, D.M., White, M.A., Wilkinson, M.L., Kon, Y., Kung, L.A., Wise, K.J., Lopez-Hoyo, N., Jiang, L., Piccirillo, S., Yu, H., Gerstein, M., Dumont, M.E., Phizicky, E.M., Snyder, M., and Grayhack, E.J. (2005) Biochemical and genetic analysis of the yeast proteome with a movable ORF collection. *Genes Dev* **19**: 2816-2826.

George, A.M., and Levy, S.B. (1983) Amplifiable resistance to tetracycline, chloramphenicol, and other antibiotics in Escherichia coli: involvement of a non-plasmid-determined efflux of tetracycline. *J Bacteriol* **155**: 531-540.

Gerdes, S., El Yacoubi, B., Bailly, M., Blaby, I.K., Blaby-Haas, C.E., Jeanguenin, L., Lara-Nunez, A., Pribat, A., Waller, J.C., Wilke, A., Overbeek, R., Hanson, A.D., and de Crecy-Lagard, V. (2011) Synergistic use of plant-prokaryote comparative genomics for functional annotations. *BMC Genomics* **12 Suppl 1**: S2.

Gershon, D. (2002) Microarray technology: an array of opportunities. *Nature* **416**: 885-891.

Giaever, G., Chu, A.M., Ni, L., Connelly, C., Riles, L., Veronneau, S., Dow, S., Lucau-Danila, A., Anderson, K., Andre, B., Arkin, A.P., Astromoff, A., El-Bakkoury, M., Bangham, R., Benito, R., Brachat, S., Campanaro, S., Curtiss, M., Davis, K., Deutschbauer, A., Entian, K.D., Flaherty, P., Foury, F., Garfinkel, D.J., Gerstein, M., Gotte, D., Guldener, U., Hegemann, J.H., Hempel, S., Herman, Z., Jaramillo,

D.F., Kelly, D.E., Kelly, S.L., Kotter, P., LaBonte, D., Lamb, D.C., Lan, N., Liang, H., Liao, H., Liu, L., Luo, C., Lussier, M., Mao, R., Menard, P., Ooi, S.L., Revuelta, J.L., Roberts, C.J., Rose, M., Ross-Macdonald, P., Scherens, B., Schimmack, G., Shafer, B., Shoemaker, D.D., Sookhai-Mahadeo, S., Storms, R.K., Strathern, J.N., Valle, G., Voet, M., Volckaert, G., Wang, C.Y., Ward, T.R., Wilhelmy, J., Winzeler, E.A., Yang, Y., Yen, G., Youngman, E., Yu, K., Bussey, H., Boeke, J.D., Snyder, M., Philippsen, P., Davis, R.W., and Johnston, M. (2002) Functional profiling of the Saccharomyces cerevisiae genome. *Nature* **418**: 387-391.

Giaever, G., and Nislow, C. (2014) The yeast deletion collection: a decade of functional genomics. *Genetics* **197**: 451-465.

Gibson, B., Krogerus, K., Ekberg, J., Monroux, A., Mattinen, L., Rautio, J., and Vidgren, V. (2015) Variation in alpha-acetolactate production within the hybrid lager yeast group Saccharomyces pastorianus and affirmation of the central role of the ILV6 gene. *Yeast* **32**: 301-316.

Gibson, D.G., Young, L., Chuang, R.Y., Venter, J.C., Hutchison, C.A., 3rd, and Smith, H.O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* **6**: 343-345.

Gorke, B., Foulquier, E., and Galinier, A. (2005) YvcK of Bacillus subtilis is required for a normal cell shape and for growth on Krebs cycle intermediates and substrates of the pentose phosphate pathway. *Microbiology* **151**: 3777-3791.

Green, M.L., and Karp, P.D. (2005) Genome annotation errors in pathway databases due to semantic ambiguity in partial EC numbers. *Nucleic Acids Res* **33**: 4035-4039.

Griesemer, M., Kimbrel, J.A., Zhou, C.E., Navid, A., and D'Haeseleer, P. (2018) Combining multiple functional annotation tools increases coverage of metabolic annotation. *BMC Genomics* **19**: 948.

Groicher, K.H., Firek, B.A., Fujimoto, D.F., and Bayles, K.W. (2000) The Staphylococcus aureus lrgAB operon modulates murein hydrolase activity and penicillin tolerance. *J Bacteriol* **182**: 1794-1801.

Grove, A. (2017) Regulation of Metabolic Pathways by MarR Family Transcription Factors. *Comput Struct Biotechnol J* **15**: 366-371.

Gruber, S., and Errington, J. (2009) Recruitment of condensin to replication origin regions by ParB/SpoOJ promotes chromosome segregation in B. subtilis. *Cell* **137**: 685-696.

Gruber, S., Veening, J.W., Bach, J., Blettinger, M., Bramkamp, M., and Errington, J. (2014) Interlinked sister chromosomes arise in the absence of condensin during fast replication in B. subtilis. *Curr Biol* **24**: 293-298.

Guedon, E., Serror, P., Ehrlich, S.D., Renault, P., and Delorme, C. (2001) Pleiotropic transcriptional repressor CodY senses the intracellular pool of branched-chain amino acids in Lactococcus lactis. *Mol Microbiol* **40**: 1227-1239.

Hadjipetrou, L.P., and Stouthamer, A.H. (1963) Autolysis of Bacillus Subtilis by Glucose Depletion. *Antonie Van Leeuwenhoek* **29**: 256-260.

Hanson, A.D., Pribat, A., Waller, J.C., and de Crecy-Lagard, V. (2009) 'Unknown' proteins and 'orphan' enzymes: the missing half of the engineering parts list--and how to find it. *Biochem J* **425**: 1-11.

Hao, Z., Lou, H., Zhu, R., Zhu, J., Zhang, D., Zhao, B.S., Zeng, S., Chen, X., Chan, J., He, C., and Chen, P.R. (2014) The multiple antibiotic resistance regulator MarR is a copper sensor in Escherichia coli. *Nat Chem Biol* **10**: 21-28.

Harwood, C.R.a.C., S. M., (1990) *Molecular biological methods for Bacillus* Wiley, New York.

Heasman, J. (2002) Morpholino oligos: making sense of antisense? *Dev Biol* **243**: 209-214.

Hensel, M., Shea, J.E., Gleeson, C., Jones, M.D., Dalton, E., and Holden, D.W. (1995) Simultaneous identification of bacterial virulence genes by negative selection. *Science* **269**: 400-403.

Hermjakob, H., Montecchi-Palazzi, L., Lewington, C., Mudali, S., Kerrien, S., Orchard, S., Vingron, M., Roechert, B., Roepstorff, P., Valencia, A., Margalit, H., Armstrong, J., Bairoch, A., Cesareni, G., Sherman, D., and Apweiler, R. (2004) IntAct: an open source molecular interaction database. *Nucleic Acids Res* **32**: D452-455.

Hochgrafe, F., Wolf, C., Fuchs, S., Liebeke, M., Lalk, M., Engelmann, S., and Hecker, M. (2008) Nitric oxide stress induces different responses but mediates comparable protein thiol protection in Bacillus subtilis and Staphylococcus aureus. *J Bacteriol* **190**: 4997-5008.

Holtzclaw, W.D., and Chapman, L.F. (1975) Degradative acetolactate synthase of Bacillus subtilis: purification and properties. *J Bacteriol* **121**: 917-922.

Hong, M., Fuangthong, M., Helmann, J.D., and Brennan, R.G. (2005) Structure of an OhrR-ohrA operator complex reveals the DNA binding mechanism of the MarR family. *Mol Cell* **20**: 131-141.

Hrdlickova, R., Toloue, M., and Tian, B. (2017) RNA-Seq methods for transcriptome analysis. *Wiley Interdiscip Rev RNA* **8**.

Huang, H., and Grove, A. (2013) The transcriptional regulator TamR from Streptomyces coelicolor controls a key step in central metabolism during oxidative stress. *Mol Microbiol* **87**: 1151-1166.

Huillet, E., Bridoux, L., Wanapaisan, P., Rejasse, A., Peng, Q., Panbangred, W., and Lereclus, D. (2017) The CodY-dependent clhAB2 operon is involved in cell shape, chaining and autolysis in Bacillus cereus ATCC 14579. *PLoS One* **12**: e0184975.

Jalal, S., and Wretlind, B. (1998) Mechanisms of quinolone resistance in clinical strains of Pseudomonas aeruginosa. *Microb Drug Resist* **4**: 257-261.

Jameson, K.H., and Wilkinson, A.J. (2017) Control of Initiation of DNA Replication in Bacillus subtilis and Escherichia coli. *Genes (Basel)* **8**.

Jerga, A., and Rock, C.O. (2009) Acyl-Acyl carrier protein regulates transcription of fatty acid biosynthetic genes via the FabT repressor in Streptococcus pneumoniae. *J Biol Chem* **284**: 15364-15368.

Johnson, C.H., Ivanisevic, J., and Siuzdak, G. (2016) Metabolomics: beyond biomarkers and towards mechanisms. *Nat Rev Mol Cell Biol* **17**: 451-459.

Karsch-Mizrachi, I., Takagi, T., Cochrane, G., and International Nucleotide Sequence Database, C. (2018) The international nucleotide sequence database collaboration. *Nucleic Acids Res* **46**: D48-D51.

Kearns, D.B. (2008) Division of labour during Bacillus subtilis biofilm formation. *Mol Microbiol* **67**: 229-231.

Kerrien, S., Aranda, B., Breuza, L., Bridge, A., Broackes-Carter, F., Chen, C., Duesbury, M., Dumousseau, M., Feuermann, M., Hinz, U., Jandrasits, C., Jimenez, R.C., Khadake, J., Mahadevan, U., Masson, P., Pedruzzi, I., Pfeiffenberger, E., Porras, P., Raghunath, A., Roechert, B., Orchard, S., and Hermjakob, H. (2012) The IntAct molecular interaction database in 2012. *Nucleic Acids Res* **40**: D841-846.

Keseler, I.M., Mackie, A., Peralta-Gil, M., Santos-Zavaleta, A., Gama-Castro, S., Bonavides-Martinez, C., Fulcher, C., Huerta, A.M., Kothari, A., Krummenacker, M., Latendresse, M., Muniz-Rascado, L., Ong, Q., Paley, S., Schroder, I., Shearer, A.G., Subhraveti, P., Travers, M., Weerasinghe, D., Weiss, V., Collado-Vides, J., Gunsalus, R.P., Paulsen, I., and Karp, P.D. (2013) EcoCyc: fusing model organism databases with systems biology. *Nucleic Acids Res* **41**: D605-612.

Keseler, I.M., Mackie, A., Santos-Zavaleta, A., Billington, R., Bonavides-Martinez, C., Caspi, R., Fulcher, C., Gama-Castro, S., Kothari, A., Krummenacker, M., Latendresse, M., Muniz-Rascado, L., Ong, Q., Paley, S., Peralta-Gil, M., Subhraveti, P., Velazquez-Ramirez, D.A., Weaver, D., Collado-Vides, J., Paulsen, I., and Karp, P.D. (2017) The EcoCyc database: reflecting new knowledge about Escherichia coli K-12. *Nucleic Acids Res* **45**: D543-D550.

Kim, H.M., Waters, A., Turner, M.E., Rice, K.C., and Ahn, S.J. (2019) Regulation of cid and lrg expression by CcpA in Streptococcus mutans. *Microbiology* **165**: 113-123.

Kim, Y., Joachimiak, G., Bigelow, L., Babnigg, G., and Joachimiak, A. (2016) How Aromatic Compounds Block DNA Binding of HcaR Catabolite Regulator. *J Biol Chem* **291**: 13243-13256.

Kitagawa, M., Ara, T., Arifuzzaman, M., Ioka-Nakamichi, T., Inamoto, E., Toyonaga, H., and Mori, H. (2005) Complete set of ORF clones of Escherichia coli ASKA library (a complete set of E. coli K-12 ORF archive): unique resources for biological research. *DNA Res* **12**: 291-299.

Kleijn, R.J., Buescher, J.M., Le Chat, L., Jules, M., Aymerich, S., and Sauer, U. (2010) Metabolic fluxes during strong carbon catabolite repression by malate in Bacillus subtilis. *J Biol Chem* **285**: 1587-1596.

Kolisnychenko, V., Plunkett, G., 3rd, Herring, C.D., Feher, T., Posfai, J., Blattner, F.R., and Posfai, G. (2002) Engineering a reduced Escherichia coli genome. *Genome Res* **12**: 640-647.

Komor, A.C., Badran, A.H., and Liu, D.R. (2017) CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes. *Cell* **168**: 20-36.

Konig, H., Frank, D., Heil, R., and Coenen, C. (2013) Synthetic genomics and synthetic biology applications between hopes and concerns. *Curr Genomics* **14**: 11-24.

Koo, B.M., Kritikos, G., Farelli, J.D., Todor, H., Tong, K., Kimsey, H., Wapinski, I., Galardini, M., Cabal, A., Peters, J.M., Hachmann, A.B., Rudner, D.Z., Allen,

K.N., Typas, A., and Gross, C.A. (2017) Construction and Analysis of Two Genome-Scale Deletion Libraries for Bacillus subtilis. *Cell Syst* **4**: 291-305 e297.

Kosmachevskaya, O.V., Shumaev, K.B., and Topunov, A.F. (2015) Carbonyl Stress in Bacteria: Causes and Consequences. *Biochemistry (Mosc)* **80**: 1655-1671.

Kovacic, P., and Cooksy, A.L. (2005) Role of diacetyl metabolite in alcohol toxicity and addiction via electron transfer and oxidative stress. *Arch Toxicol* **79**: 123-128.

Krogerus, K.a.G., B. R. (2013) 125th Anniversary Review: Diacetyl and its control during brewery fermentation. *Journal of the Institute of Brewing*: 86-97.

Kumaraswami, M., Schuman, J.T., Seo, S.M., Kaatz, G.W., and Brennan, R.G. (2009) Structural and biochemical characterization of MepR, a multidrug binding transcription regulator of the Staphylococcus aureus multidrug efflux pump MepA. *Nucleic Acids Res* **37**: 1211-1224.

Kumarevel, T., Tanaka, T., Umehara, T., and Yokoyama, S. (2009) ST1710-DNA complex crystal structure reveals the DNA binding mechanism of the MarR family of regulators. *Nucleic Acids Res* **37**: 4723-4735.

Labaer, J., Qiu, Q., Anumanthan, A., Mar, W., Zuo, D., Murthy, T.V., Taycher, H., Halleck, A., Hainsworth, E., Lory, S., and Brizuela, L. (2004) The Pseudomonas aeruginosa PA01 gene collection. *Genome Res* **14**: 2190-2200.

Lemon, K.P., and Grossman, A.D. (1998) Localization of bacterial DNA polymerase: evidence for a factory model of replication. *Science* **282**: 1516-1519.

Lemon, K.P., and Grossman, A.D. (2000) Movement of replicating DNA through a stationary replisome. *Mol Cell* **6**: 1321-1330.

Levdikov, V.M., Blagova, E., Colledge, V.L., Lebedev, A.A., Williamson, D.C., Sonenshein, A.L., and Wilkinson, A.J. (2009) Structural rearrangement accompanying ligand binding in the GAF domain of CodY from Bacillus subtilis. *J Mol Biol* **390**: 1007-1018.

Levdikov, V.M., Blagova, E., Joseph, P., Sonenshein, A.L., and Wilkinson, A.J. (2006) The structure of CodY, a GTP- and isoleucine-responsive regulator of stationary phase and virulence in gram-positive bacteria. *J Biol Chem* **281**: 11366-11373.

Liu, H., Krizek, J., and Bretscher, A. (1992) Construction of a GAL1-regulated yeast cDNA expression library and its application to the identification of genes whose overexpression causes lethality in yeast. *Genetics* **132**: 665-673.

Liu, J., Prindle, A., Humphries, J., Gabalda-Sagarra, M., Asally, M., Lee, D.Y., Ly, S., Garcia-Ojalvo, J., and Suel, G.M. (2015) Metabolic co-dependence gives rise to collective oscillations within biofilms. *Nature* **523**: 550-554.

Loken, J.P., and Stormer, F.C. (1970) Acetolactate decarboxylase from Aerobacter aerogenes. Purification and properties. *Eur J Biochem* **14**: 133-137.

Lopez, D., and Kolter, R. (2010) Extracellular signals that define distinct and coexisting cell fates in Bacillus subtilis. *FEMS Microbiol Rev* **34**: 134-149.

Lugtenberg, B., and Kamilova, F. (2009) Plant-growth-promoting rhizobacteria. *Annu Rev Microbiol* **63**: 541-556.

Maddocks, S.E., and Oyston, P.C. (2008) Structure and function of the LysR-type transcriptional regulator (LTTR) family proteins. *Microbiology* **154**: 3609-3623.

Marchadier, E., Carballido-Lopez, R., Brinster, S., Fabret, C., Mervelet, P., Bessieres, P., Noirot-Gros, M.F., Fromion, V., and Noirot, P. (2011) An expanded protein-

protein interaction network in Bacillus subtilis reveals a group of hubs: Exploration by an integrative approach. *Proteomics* **11**: 2981-2991.

Marschner, H., (2011) *Marschner's Mineral Nutrition of Higher Plants*. Academic Press.

Martins, B.M., and Locke, J.C. (2015) Microbial individuality: how single-cell heterogeneity enables population level strategies. *Curr Opin Microbiol* **24**: 104-112.

Meehan, T.F., Conte, N., West, D.B., Jacobsen, J.O., Mason, J., Warren, J., Chen, C.K., Tudose, I., Relac, M., Matthews, P., Karp, N., Santos, L., Fiegel, T., Ring, N., Westerberg, H., Greenaway, S., Sneddon, D., Morgan, H., Codner, G.F., Stewart, M.E., Brown, J., Horner, N., International Mouse Phenotyping, C., Haendel, M., Washington, N., Mungall, C.J., Reynolds, C.L., Gallegos, J., Gailus-Durner, V., Sorg, T., Pavlovic, G., Bower, L.R., Moore, M., Morse, I., Gao, X., Tocchini-Valentini, G.P., Obata, Y., Cho, S.Y., Seong, J.K., Seavitt, J., Beaudet, A.L., Dickinson, M.E., Herault, Y., Wurst, W., de Angelis, M.H., Lloyd, K.C.K., Flenniken, A.M., Nutter, L.M.J., Newbigging, S., McKerlie, C., Justice, M.J., Murray, S.A., Svenson, K.L., Braun, R.E., White, J.K., Bradley, A., Flicek, P., Wells, S., Skarnes, W.C., Adams, D.J., Parkinson, H., Mallon, A.M., Brown, S.D.M., and Smedley, D. (2017) Disease model discovery from 3,328 gene knockouts by The International Mouse Phenotyping Consortium. *Nat Genet* **49**: 1231-1238.

Meeske, A.J., Rodrigues, C.D., Brady, J., Lim, H.C., Bernhardt, T.G., and Rudner, D.Z. (2016) High-Throughput Genetic Screens Identify a Large and Diverse

Collection of New Sporulation Genes in Bacillus subtilis. *PLoS Biol* **14**: e1002341.

Meeske, A.J., Sham, L.T., Kimsey, H., Koo, B.M., Gross, C.A., Bernhardt, T.G., and Rudner, D.Z. (2015) MurJ and a novel lipid II flippase are required for cell wall biogenesis in Bacillus subtilis. *Proc Natl Acad Sci U S A* **112**: 6437-6442.

Meinhardt, S., Manley, M.W., Jr., Becker, N.A., Hessman, J.A., Maher, L.J., 3rd, and Swint-Kruse, L. (2012) Novel insights from hybrid LacI/GalR proteins: family-wide functional attributes and biologically significant variation in transcription repression. *Nucleic Acids Res* **40**: 11139-11154.

Miller, A.K., Brown, E.E., Mercado, B.T., and Herman, J.K. (2016) A DNA-binding protein defines the precise region of chromosome capture during Bacillus sporulation. *Mol Microbiol* **99**: 111-122.

Milne, N., Wahl, S.A., van Maris, A.J.A., Pronk, J.T., and Daran, J.M. (2016) Excessive by-product formation: A key contributor to low isobutanol yields of engineered Saccharomyces cerevisiae strains. *Metab Eng Commun* **3**: 39-51.

Moreno, M.S., Schneider, B.L., Maile, R.R., Weyler, W., and Saier, M.H., Jr. (2001) Catabolite repression mediated by the CcpA protein in Bacillus subtilis: novel modes of regulation revealed by whole-genome analyses. *Mol Microbiol* **39**: 1366-1381.

Morgens, D.W., Wainberg, M., Boyle, E.A., Ursu, O., Araya, C.L., Tsui, C.K., Haney, M.S., Hess, G.T., Han, K., Jeng, E.E., Li, A., Snyder, M.P., Greenleaf, W.J., Kundaje, A., and Bassik, M.C. (2017) Genome-scale measurement of off-target activity using Cas9 toxicity in high-throughput screens. *Nat Commun* **8**: 15178.

Mostertz, J., Scharf, C., Hecker, M., and Homuth, G. (2004) Transcriptome and proteome analysis of Bacillus subtilis gene expression in response to superoxide and peroxide stress. *Microbiology* **150**: 497-512.

Muller, S., and Nebe-von-Caron, G. (2010) Functional single-cell analyses: flow cytometry and cell sorting of microbial populations and communities. *FEMS Microbiol Rev* **34**: 554-587.

Nemeria, N., Tittmann, K., Joseph, E., Zhou, L., Vazquez-Coll, M.B., Arjunan, P., Hubner, G., Furey, W., and Jordan, F. (2005) Glutamate 636 of the Escherichia coli pyruvate dehydrogenase-E1 participates in active center communication and behaves as an engineered acetolactate synthase with unusual stereoselectivity. *J Biol Chem* **280**: 21473-21482.

Newberry, K.J., Fuangthong, M., Panmanee, W., Mongkolsuk, S., and Brennan, R.G. (2007) Structural mechanism of organic hydroperoxide induction of the transcription regulator OhrR. *Mol Cell* **28**: 652-664.

Nichols, R.J., Sen, S., Choo, Y.J., Beltrao, P., Zietek, M., Chaba, R., Lee, S., Kazmierczak, K.M., Lee, K.J., Wong, A., Shales, M., Lovett, S., Winkler, M.E., Krogan, N.J., Typas, A., and Gross, C.A. (2011) Phenotypic landscape of a bacterial cell. *Cell* **144**: 143-156.

Nicolas, P., Mader, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., Bidnenko, E., Marchadier, E., Hoebeke, M., Aymerich, S., Becher, D., Bisicchia, P., Botella, E., Delumeau, O., Doherty, G., Denham, E.L., Fogg, M.J., Fromion, V., Goelzer, A., Hansen, A., Hartig, E., Harwood, C.R., Homuth, G., Jarmer, H., Jules, M., Klipp, E., Le Chat, L., Lecointe, F., Lewis, P., Liebermeister, W., March, A.,

Mars, R.A., Nannapaneni, P., Noone, D., Pohl, S., Rinn, B., Rugheimer, F., Sappa, P.K., Samson, F., Schaffer, M., Schwikowski, B., Steil, L., Stulke, J., Wiegert, T., Devine, K.M., Wilkinson, A.J., van Dijl, J.M., Hecker, M., Volker, U., Bessieres, P., and Noirot, P. (2012) Condition-dependent transcriptome reveals high-level regulatory architecture in Bacillus subtilis. *Science* **335**: 1103-1106.

Nobre, T., Campos, M.D., Lucic-Mercy, E., and Arnholdt-Schmitt, B. (2016) Misannotation Awareness: A Tale of Two Gene-Groups. *Front Plant Sci* **7**: 868.

Oh, H., Stenhoff, J., Jalal, S., and Wretlind, B. (2003) Role of efflux pumps and mutations in genes for topoisomerases II and IV in fluoroquinolone-resistant Pseudomonas aeruginosa strains. *Microb Drug Resist* **9**: 323-328.

Pagliai, F.A., Pan, L., Silva, D., Gonzalez, C.F., and Lorca, G.L. (2018) Zinc is an inhibitor of the LdtR transcriptional activator. *PLoS One* **13**: e0195746.

Park, H.S., Xing, R., and Whitman, W.B. (1995) Nonenzymatic acetolactate oxidation to diacetyl by flavin, nicotinamide and quinone coenzymes. *Biochim Biophys Acta* **1245**: 366-370.

Passos, G.A., (2014) *Transcriptomics in health and disease*. Switzerland.

Pelchovich, G., Omer-Bendori, S., and Gophna, U. (2013) Menaquinone and iron are essential for complex colony development in Bacillus subtilis. *PLoS One* **8**: e79488.

Peters, J.M., Colavin, A., Shi, H., Czarny, T.L., Larson, M.H., Wong, S., Hawkins, J.S., Lu, C.H.S., Koo, B.M., Marta, E., Shiver, A.L., Whitehead, E.H., Weissman, J.S., Brown, E.D., Qi, L.S., Huang, K.C., and Gross, C.A. (2016) A

Comprehensive, CRISPR-based Functional Analysis of Essential Genes in Bacteria. *Cell* **165**: 1493-1506.

Pick, T.R., Brautigam, A., Schulz, M.A., Obata, T., Fernie, A.R., and Weber, A.P. (2013) PLGG1, a plastidic glycolate glycerate transporter, is required for photorespiration and defines a unique class of metabolite transporters. *Proc Natl Acad Sci U S A* **110**: 3185-3190.

Pinu, F.R., and Villas-Boas, S.G. (2017) Extracellular Microbial Metabolomics: The State of the Art. *Metabolites* **7**.

Poncet, S., Soret, M., Mervelet, P., Deutscher, J., and Noirot, P. (2009) Transcriptional activator YesS is stimulated by histidine-phosphorylated HPr of the Bacillus subtilis phosphotransferase system. *J Biol Chem* **284**: 28188-28197.

Randazzo, P., Aubert-Frambourg, A., Guillot, A., and Auger, S. (2016) The MarR-like protein PchR (YvmB) regulates expression of genes involved in pulcherriminic acid biosynthesis and in the initiation of sporulation in Bacillus subtilis. *BMC Microbiol* **16**: 190.

Ratnayake-Lecamwasam, M., Serror, P., Wong, K.W., and Sonenshein, A.L. (2001) Bacillus subtilis CodY represses early-stationary-phase genes by sensing GTP levels. *Genes Dev* **15**: 1093-1103.

Reyes-Caballero, H., Guerra, A.J., Jacobsen, F.E., Kazmierczak, K.M., Cowart, D., Koppolu, U.M., Scott, R.A., Winkler, M.E., and Giedroc, D.P. (2010) The metalloregulatory zinc site in Streptococcus pneumoniae AdcR, a zinc-activated MarR family repressor. *J Mol Biol* **403**: 197-216.

Rice, K.C., and Bayles, K.W. (2008) Molecular control of bacterial death and lysis. *Microbiol Mol Biol Rev* **72**: 85-109, table of contents.

Rice, K.C., Turner, M.E., Carney, O.V., Gu, T., and Ahn, S.J. (2017) Modification of the Streptococcus mutans transcriptome by LrgAB and environmental stressors. *Microb Genom* **3**: e000104.

Rice, P., Longden, I., and Bleasby, A. (2000) EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet* **16**: 276-277.

Rigali, S., Derouaux, A., Giannotta, F., and Dusart, J. (2002) Subdivision of the helix-turn-helix GntR family of bacterial regulators in the FadR, HutC, MocR, and YtrA subfamilies. *J Biol Chem* **277**: 12507-12515.

Riley, M., Abe, T., Arnaud, M.B., Berlyn, M.K., Blattner, F.R., Chaudhuri, R.R., Glasner, J.D., Horiuchi, T., Keseler, I.M., Kosuge, T., Mori, H., Perna, N.T., Plunkett, G., 3rd, Rudd, K.E., Serres, M.H., Thomas, G.H., Thomson, N.R., Wishart, D., and Wanner, B.L. (2006) Escherichia coli K-12: a cooperatively developed annotation snapshot--2005. *Nucleic Acids Res* **34**: 1-9.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**: 139-140.

Rosenthal, A.Z., Qi, Y., Hormoz, S., Park, J., Li, S.H., and Elowitz, M.B. (2018) Metabolic interactions between dynamic bacterial subpopulations. *Elife* **7**.

Roy, A., Kucukural, A., and Zhang, Y. (2010) I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc* **5**: 725-738.

Rutherford, S.T., and Bassler, B.L. (2012) Bacterial quorum sensing: its role in virulence and possibilities for its control. *Cold Spring Harb Perspect Med* **2**.

Saier, M.H., Jr., Chauvaux, S., Deutscher, J., Reizer, J., and Ye, J.J. (1995) Protein phosphorylation and regulation of carbon metabolism in gram-negative versus gram-positive bacteria. *Trends Biochem Sci* **20**: 267-271.

Saito, K., Akama, H., Yoshihara, E., and Nakae, T. (2003) Mutations affecting DNA-binding activity of the MexR repressor of mexR-mexA-mexB-oprM operon expression. *J Bacteriol* **185**: 6195-6198.

Santiago, A.E., Yan, M.B., Tran, M., Wright, N., Luzader, D.H., Kendall, M.M., Ruiz-Perez, F., and Nataro, J.P. (2016) A large family of anti-activators accompanying XylS/AraC family regulatory proteins. *Mol Microbiol* **101**: 314-332.

Schibany, S., Kleine Borgmann, L.A.K., Rosch, T.C., Knust, T., Ulbrich, M.H., and Graumann, P.L. (2018) Single molecule tracking reveals that the bacterial SMC complex moves slowly relative to the diffusion of the chromosome. *Nucleic Acids Res* **46**: 7805-7819.

Schindler, B.D., Seo, S.M., Jacinto, P.L., Kumaraswami, M., Birukou, I., Brennan, R.G., and Kaatz, G.W. (2013) Functional consequences of substitution mutations in MepR, a repressor of the Staphylococcus aureus MepA multidrug efflux pump gene. *J Bacteriol* **195**: 3651-3662.

Scholefield, G., Errington, J., and Murray, H. (2012) Soj/ParA stalls DNA replication by inhibiting helix formation of the initiator protein DnaA. *EMBO J* **31**: 1542-1555.

Scholefield, G., and Murray, H. (2013) YabA and DnaD inhibit helix assembly of the DNA replication initiation protein DnaA. *Mol Microbiol* **90**: 147-159.

Seger, J.a.H.J.B. (1987) What is bet-hedging? *Oxford Surveys in Evolutionary Biology*: 182-211.

Sham, L.T., Butler, E.K., Lebar, M.D., Kahne, D., Bernhardt, T.G., and Ruiz, N. (2014) Bacterial cell wall. MurJ is the flippase of lipid-linked precursors for peptidoglycan biogenesis. *Science* **345**: 220-222.

Shearer, A.G., Altman, T., and Rhee, C.D. (2014) Finding sequences for over 270 orphan enzymes. *PLoS One* **9**: e97250.

Shemesh, M., Tam, A., Kott-Gutkowski, M., Feldman, M., and Steinberg, D. (2008) DNA-microarrays identification of Streptococcus mutans genes associated with biofilm thickness. *BMC Microbiol* **8**: 236.

Shepard, W., Soutourina, O., Courtois, E., England, P., Haouz, A., and Martin-Verstraete, I. (2011) Insights into the Rrf2 repressor family--the structure of CymR, the global cysteine regulator of Bacillus subtilis. *FEBS J* **278**: 2689-2701.

Shi, L., Pigeonneau, N., Ventroux, M., Derouiche, A., Bidnenko, V., Mijakovic, I., and Noirot-Gros, M.F. (2014) Protein-tyrosine phosphorylation interaction network in Bacillus subtilis reveals new substrates, kinase activators and kinase cross-talk. *Front Microbiol* **5**: 538.

Shim, J.H., Park, J.T., Hong, J.S., Kim, K.W., Kim, M.J., Auh, J.H., Kim, Y.W., Park, C.S., Boos, W., Kim, J.W., and Park, K.H. (2009) Role of maltogenic amylase and pullulanase in maltodextrin and glycogen metabolism of Bacillus subtilis 168. *J Bacteriol* **191**: 4835-4844.

Shivers, R.P., and Sonenshein, A.L. (2004) Activation of the Bacillus subtilis global regulator CodY by direct interaction with branched-chain amino acids. *Mol Microbiol* **53**: 599-611.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J., Thompson, J.D., and Higgins, D.G. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7**: 539.

Sonenshein, A.L. (2007) Control of key metabolic intersections in Bacillus subtilis. *Nat Rev Microbiol* **5**: 917-927.

Sorokina, M., Stam, M., Medigue, C., Lespinet, O., and Vallenet, D. (2014) Profiling the orphan enzymes. *Biol Direct* **9**: 10.

Srikumar, R., Tsang, E., and Poole, K. (1999) Contribution of the MexAB-OprM multidrug efflux system to the beta-lactam resistance of penicillin-binding protein and beta-lactamase-derepressed mutants of Pseudomonas aeruginosa. *J Antimicrob Chemother* **44**: 537-540.

Sterlini, J.M., and Mandelstam, J. (1969) Commitment to sporulation in Bacillus subtilis and its relationship to development of actinomycin resistance. *Biochem J* **113**: 29-37.

Sukdeo, N., and Honek, J.F. (2008) Microbial glyoxalase enzymes: metalloenzymes controlling cellular levels of methylglyoxal. *Drug Metabol Drug Interact* **23**: 29-50.

Suomalainen, H., and Ronkainen, P. (1968) Mechanism of diacetyl formation in yeast fermentation. *Nature* **220**: 792-793.

Tobes, R., and Ramos, J.L. (2002) AraC-XylS database: a family of positive transcriptional regulators in bacteria. *Nucleic Acids Res* **30**: 318-321.

van Bergen, B., Strasser, R., Cyr, N., Sheppard, J.D., and Jardim, A. (2006) Alpha,beta-dicarbonyl reduction by Saccharomyces D-arabinose dehydrogenase. *Biochim Biophys Acta* **1760**: 1636-1645.

van den Esker, M.H., Kovacs, A.T., and Kuipers, O.P. (2017) YsbA and LytST are essential for pyruvate utilization in Bacillus subtilis. *Environ Microbiol* **19**: 83-94.

van Opijnen, T., Bodi, K.L., and Camilli, A. (2009) Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. *Nat Methods* **6**: 767-772.

Vlamakis, H., Aguilar, C., Losick, R., and Kolter, R. (2008) Control of cell fate by the formation of an architecturally complex bacterial community. *Genes Dev* **22**: 945-953.

W., R., (1997-2015) *ImageJ*. Bethesda, Maryland: U.S. National Institutes of Health.

Wagner-Herman, J.K., Bernard, R., Dunne, R., Bisson-Filho, A.W., Kumar, K., Nguyen, T., Mulcahy, L., Koullias, J., Gueiros-Filho, F.J., and Rudner, D.Z. (2012) RefZ facilitates the switch from medial to polar division during spore formation in Bacillus subtilis. *J Bacteriol* **194**: 4608-4618.

Wang, X., Le, T.B., Lajoie, B.R., Dekker, J., Laub, M.T., and Rudner, D.Z. (2015) Condensin promotes the juxtaposition of DNA flanking its loading site in Bacillus subtilis. *Genes Dev* **29**: 1661-1675.

Warner, J.B., and Lolkema, J.S. (2003) CcpA-dependent carbon catabolite repression in bacteria. *Microbiol Mol Biol Rev* **67**: 475-490.

Wilke, M.S., Heller, M., Creagh, A.L., Haynes, C.A., McIntosh, L.P., Poole, K., and Strynadka, N.C. (2008) The crystal structure of MexR from Pseudomonas aeruginosa in complex with its antirepressor ArmR. *Proc Natl Acad Sci U S A* **105**: 14832-14837.

Wilkinson, S.P., and Grove, A. (2006) Ligand-responsive transcriptional regulation by members of the MarR family of winged helix proteins. *Curr Issues Mol Biol* **8**: 51-62.

Windham, I.H., Chaudhari, S.S., Bose, J.L., Thomas, V.C., and Bayles, K.W. (2016) SrrAB Modulates Staphylococcus aureus Cell Death through Regulation of cidABC Transcription. *J Bacteriol* **198**: 1114-1122.

Wolfe, A.J. (2005) The acetate switch. *Microbiol Mol Biol Rev* **69**: 12-50.

Wondrak, G.T., Cervantes-Laurean, D., Roberts, M.J., Qasem, J.G., Kim, M., Jacobson, E.L., and Jacobson, M.K. (2002) Identification of alpha-dicarbonyl scavengers for cellular protection against carbonyl stress. *Biochem Pharmacol* **63**: 361-373.

Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., and Zhang, Y. (2015) The I-TASSER Suite: protein structure and function prediction. *Nat Methods* **12**: 7-8.

Yang, S.J., Dunman, P.M., Projan, S.J., and Bayles, K.W. (2006) Characterization of the Staphylococcus aureus CidR regulon: elucidation of a novel role for acetoin metabolism in cell death and lysis. *Mol Microbiol* **60**: 458-468.

Yang, S.J., Rice, K.C., Brown, R.J., Patton, T.G., Liou, L.E., Park, Y.H., and Bayles, K.W. (2005) A LysR-type regulator, CidR, is required for induction of the Staphylococcus aureus cidABC operon. *J Bacteriol* **187**: 5893-5900.

Yang, Y., Jin, H., Chen, Y., Lin, W., Wang, C., Chen, Z., Han, N., Bian, H., Zhu, M., and Wang, J. (2012) A chloroplast envelope membrane protein containing a putative LrgB domain related to the control of bacterial death and lysis is required for chloroplast development in Arabidopsis thaliana. *New Phytol* **193**: 81-95.

Youngman, P.J., Perkins, J.B., and Losick, R. (1983) Genetic transposition and insertional mutagenesis in Bacillus subtilis with Streptococcus faecalis transposon Tn917. *Proc Natl Acad Sci U S A* **80**: 2305-2309.

Zhang, Y. (2008) I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* **9**: 40.

Zhu, J.Y., Fu, Y., Nettleton, M., Richman, A., and Han, Z. (2017a) High throughput in vivo functional validation of candidate congenital heart disease genes in Drosophila. *Elife* **6**.

Zhu, R., Hao, Z., Lou, H., Song, Y., Zhao, J., Chen, Y., Zhu, J., and Chen, P.R. (2017b) Structural characterization of the DNA-binding mechanism underlying the copper(II)-sensing MarR transcriptional regulator. *J Biol Inorg Chem* **22**: 685-693.