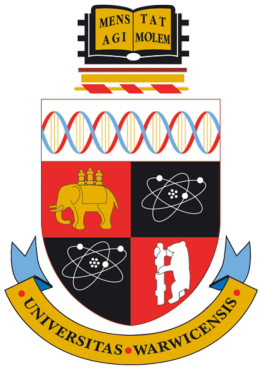
WARWICK

A Multi-Omics Approach in the Characterisation of the RNA-Binding Protein SpoVG in *Bacillus subtilis*



by

Swati Mahapatra

A thesis submitted in partial fulfilment of the
requirements for the degree of
Doctor of Philosophy (Medical Sciences)

Warwick Medical School, University of Warwick

November 2020

Table of Contents

List of Tables

List of Figures

Acknowledgements

First and foremost, I extend my sincere thanks and gratitude to my supervisor Dr. Emma L. Denham for giving me the opportunity to pursue my dream education (PhD). I still remember the joy of being selected for the PhD position and eagerly waiting to start my graduate school experience. I will remain forever indebted for her continued guidance, encouragement and mentorship throughout this journey. I am also extremely thankful for her kindness and continued support even during my cross-continental, long-distance and global pandemic hit writing phase. Thank you for everything, Emma! I will always cherish this experience.

I also express my immense gratitude and thanks to my supervisor Dr. Chrystala Constantinidou for being so kind, supportive and helpful throughout my PhD. She was always ready to take out time for me and support me whenever I needed any help. My work would not have been easy without the help of my supervisor, Dr. Mohammad Tauqeer Alam. He has been very kind in helping me to learn coding in R and assist me in data analysis. He is a great teacher and mentor who constantly directed me towards the right path. I am glad he agreed to become my secondary supervisor and take-out time in mentoring me. I also thank the University of Warwick for awarding me with Chancellor's International Scholarship for my PhD, which was crucial in aiding my stay in the UK.

My time at Warwick would have been incomplete without my ‘work-wife’ Dr. Jenna Lam. Even if our interaction at Warwick was short, I will cherish the bond that we built. She was always there for a rant, a cup of coffee or tea, pot-noodle, late-evening experiments and numerous weekend fun. I really wish we could spend more time together! Thank you, Leo, for helping me with the data analysis and taking time out to help with coding. I am also thankful to the previous members of Denham lab, Dr. Josie McKweon and Dr. Holly Hall for their help during my PhD, proofreading a few chapters of my thesis and suggesting useful changes. I wish I could spend more time

and learn more from both. I also thank all the past and current members of Microbiology and Infection Unit (MIU) who helped me during my PhD.

I am thankful to my Warwick friends Anjali, Farah and Ines for being good friends and sharing this journey with me. I express my dearest thanks to my housemates Aleida, Kate and Samir for being my pseudo family in Coventry. I also thank Sucheta being a good friend even from far away.

I was lucky to have worked with some amazing people at Bath. Firstly, I sincerely thank the Department of Biology and Biochemistry, University of Bath for quickly integrating me with the department. I am obliged to Dr. Susanne Gebhard, Dr. Tiffany Taylor, Dr. Hans-Wilhelm Nützmann, Dr Gernot Walko and Sunny Lau for providing me with the necessary equipment and reagents towards completing my research. I am grateful to my Bath friend and colleague Louise Flanagan for being a sincere friend and a lovely junior. I also thank Dr. Caro Kobras for proofreading one of my thesis chapters and sending me useful comments, and Dr. Stacy Ramakisoon and Dr. Bianca Reeksting for being great colleagues at work.

I will also remain forever thankful to my teachers Dr. Krishnaveni Mishra and Mrs. Renuka Madhu for recommending me for my PhD position. I thank Dr. Krushnamegh Kunte for being a great mentor and an inspiring scientist. Last but not the least, I express my extreme gratitude and thanks to my parents and siblings for always being there, tolerating the painstaking separation for so long and believing in me. They have been my greatest emotional support and turned into my closest friends over the last few years. Special thanks to my dearest sister for listening to my maximum rant and always being so calm and patient. I love you all!

Finally, I sincerely thank my husband Devarshi for being the constant in this ever-changing life, encouraging to achieve the best and the favourite friend to rely on. I really do not know how things would have turned out post our marriage if I continued living afar from you and survive the pandemic. I express my gratitude for your constant love and care especially in the tough times. I love you.

Declaration

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy (Medical Sciences). It has been composed by myself and has not been submitted in any previous application for any degree. The work presented (including data generated and data analysis) was carried out by the author except in the cases outlined below:

Chapter 3

1. The original Python script for filtering coverageBed read counts was written by Dr. Richard Brown and modified with the help of Mr. Leonidas Souliotis (Brown, 2016).
2. The Python script for de-multiplexing RNAtag-Seq fastq files was written with the help of Mr. Leonidas Souliotis and Dr. Jenna Lam.

Chapter 4

Mass-spectrometry run for the proteomic analysis, the initial peptide matching, and database search were carried out by Dr. Juan Ramon Hernandez-Fernaud at the Proteomics Research Technology Platform, University of Warwick.

Chapter 5

1. The RNA-protein UV cross-linking, immunoprecipitation and RNA sequencing experiments were carried out by Dr. Pedro Arede Rei, Granneman Lab, University of Edinburgh, UK.
2. The RNA sequencing mapping, filtration and data analysis was carried out by Dr. Sander Granneman, University of Edinburgh, UK.

Chapter 6

The mass-spectrometric run of the protein samples, peptide matching, and database search were carried out by Dr. Liangcui Chu, Granneman Lab, University of Edinburgh, UK.

Chapter 7

The biofilm assay on Δ*spoVG* in MSgg agar was carried out by Ms. Tetyana Sukhodub, Stanley-Wall Lab, University of Dundee, UK.

Abstract

Post-transcriptional gene regulation via small non-coding RNAs (sRNAs) have been well characterised in many bacteria. RNA-binding proteins (RBPs) are emerging as key players in mediating sRNA based gene regulation. Most of these studies have focused on investigating Gram-negative bacteria. Nonetheless, little is known about RBPs in Gram-positive bacteria. Furthermore, in many Gram-positive bacteria the known RBPs are either absent or do not function similar to RBPs in the Gram-negatives. Such evidence suggest alternate gene regulatory mechanisms in Gram-positives that either include alternative RBPs or regulatory pathways that do not involve RBPs. Emerging studies in various bacteria have identified SpoVG as a novel RBP. *spoVG* was initially identified in *B. subtilis* as a sporulation gene. The gene is highly conserved with constitutive expression in *B. subtilis*. However, nothing is known about its RNA-binding activities in *B. subtilis*.

In this thesis, it was hypothesised that SpoVG is an RBP *in vivo* and a post-transcriptional gene regulator in *B. subtilis*. To test this, a global multi-omics approach was taken. It was shown that deletion of *spoVG* perturbs both the total transcriptomic and proteomic landscapes of *B. subtilis* (Chapters 3 and 4). SpoVG was also identified as a global RBP (Chapter 5). A protein-protein interaction study indicated potential protein binding activity of SpoVG (Chapter 6). Finally, Δ*spoVG* was functionally characterised based on the -omics data from this thesis and novel phenotypes of Δ*spoVG* were reported (Chapter 7). Cumulatively, these results indicated a global regulatory role of SpoVG in *B. subtilis* beyond its role in sporulation. Nevertheless, its role as a global post-transcriptional gene regulator remains elusive. The findings reported in this study not only promote SpoVG as novel RBP with regulator functions in *B. subtilis*, but they will also benefit in uncovering the less explored realms of regulatory networks of RBPs in other bacteria.

Abbreviations

˚C	Degree Celsius
~	Approximately
bp	Base pairs
cDNA	Complementary DNA
CLASH	Cross-linking, Ligation and Sequencing of Hybrids
CLIP	Ultraviolet (UV) cross-linking and immunoprecipitation
co-IP	Co-immunoprecipitation
co-IPed	Co-immunoprecipitated
CRAC	UV crosslinking and analysis of cDNAs
DEG	Differentially expressed gene
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
DPP	Differentially produced protein
ECF	Extracytoplasmic factor
g	Relative centrifugal force
gDNA	Genomic DNA
GFP	Green fluorescent protein
GSR	General stress response
h	Hour
HRP	Horseradish peroxidase
HTF	His(6X)-TEV cleavage site-FLAG(3X)
IP	Immunoprecipitation
KEGG	Kyoto Encyclopedia of Genes and Genomes
LB	Lysogeny broth
LBA	Lysogeny broth agar
LBGM	LB with glycerol (1%) and manganese (1 mM $MnSO_4$)
LFQ	Label free quantification

M9	Minimal media with glucose
Mb	Million base pairs (bases)
min	Minute
mM	Millimolar
mRNA	Messenger RNA
MS/MS	Tandem mass spectrometry
MS^2	MS/MS spectra
ncRNA	Non-coding RNA
nm	Nanometre
NRP	Non-ribosomal peptide
nt	Nucleotide
O/N	Overnight
OD_{600}	Optical density at 600 nm
PBS	Phosphate-buffered saline
PC	Phenol:chloroform
PCA	Principal component analysis
PCI	Phenol:chloroform:Isoamyl-alcohol
PCR	Polymerase chain reaction
PK	Polyketide
PPI	Protein-protein interaction
PVDF	Polyvinylidene difluoride
QA	Quality analysis
R	Pearson's correlation coefficient
RBP	RNA-binding protein
RNA	Ribonucleic acid
RNase	Ribonuclease
rpm	Rotations per min
RPS	Ribosomal proteins
RT	Room temperature

s	Seconds
SDS-PAGE	Sodium dodecyl sulphate–polyacrylamide gel electrophoresis
SH2	Sporulation hour 2
SH5	Sporulation hour 5
sRNA	Small non-coding RNA
TCA	Trichloroacetic acid
TCA cycle	Tricarboxylic acid cycle
TF	Transcription factor
UMP	Uridine monophosphate
UTR	Untranslated region
UV	Ultraviolet
V	Volt
WB	Western blot
WT	Wild-type
µ	Micro

Chapter 1 Introduction

In order to sense and thrive in a constantly changing environment, bacteria need to quickly fine-tune their gene expression and adapt in response to the external cues. A traditional regulation of gene expression occurs at the promoter region of the gene of interest where the RNA polymerase binds in association with the suitable sigma factor, that determines the specificity of binding. In bacteria, several sigma factors are known with conserved functions across bacterial species. Nonetheless, many other sigma factors are also functional in a condition-dependent and bacteria specific manner which modulate the gene expression. The gene expression in bacteria also encompasses several transcriptional termination mechanisms wherein, the mode of termination of the transcription determines the fate of the expression of the genes. However, many of these processes are heavily energy consuming and needs time to execute. Nonetheless, to swiftly fine-tune the gene expression in rapidly changing environments, bacteria need mechanism that are quick and efficient.

Many of these quick methods of gene regulation include molecules such as small non-coding RNAs (sRNAs) and RNA-binding proteins (RBPs) that function together to modulate cellular functions. Although a lot has been studied regarding the functioning of sRNAs, only a little is known about the RBPs in Gram-positive bacteria. This thesis reports the discovery of the *in vivo* RNA-binding properties of SpoVG in *B. subtilis*. A whole-cell transcriptomics and proteomics study of the Δ*spoVG* knockout mutant under routine laboratory conditions identified several significantly differentially expressed genes and proteins. Nevertheless, it remains unknown whether these changes were the direct or the indirect effect of the *spoVG* deletion. To identify the RNA-RNA interactions mediated by SpoVG, an sRNA-mRNA immunoprecipitation experiment was also carried out. The study identified only a few sRNA-mRNA chimeras bound to SpoVG, which indicated that SpoVG did not have a global effect in mediating these

interactions in *B. subtilis*. Furthermore, an *in vivo* protein-protein interaction experiment with SpoVG was also carried out. The results from this experiment identified a few potential protein partners of SpoVG, suggesting perhaps a multi-functional role of SpoVG in *B. subtilis* which remain to be investigated further. Finally, novel phenotypes of Δ*spoVG* mutant are also reported in this thesis that sheds light on the previously understudied roles of SpoVG in *B. subtilis*. In conclusion, this thesis opens a number of avenues for investigating the role of SpoVG in the post-transcriptional gene regulation of *B. subtilis* and other Gram-positive bacteria.

1.1 Bacterial Post-Transcriptional Gene Regulation

In bacteria, the process of transcription and translation occurs simultaneously. When the protein of interest is no longer needed by a cell, the transcription of the gene terminates, hence stopping the translation of the gene. Therefore, a primary method to control the production of proteins and their levels in bacteria are regulated at the DNA level. The initiation of transcription of a gene can be regulated by the binding of specific sigma factors (in association with RNA polymerase) to the promoters of the genes. Sigma factors are the subunits of RNA polymerase that determine the specificity of binding to promoter sequence and the efficient initiation of transcription. Furthermore, regulation of gene expression is also mediated through the operator region of an operon (cluster of genes involved in the same pathway that are transcribed from a single promoter). The operons can be turned on/off based on the binding of regulatory proteins to these DNA sequences, or by binding of small metabolites to these regulatory proteins. In addition to these mechanisms, bacteria can also regulate their gene expression by chemically modifying proteins post-translationally. Some of the common protein modifications in bacteria include protein phosphorylation, acetylation, succinylation and glycosylation. These modified proteins participate in a diverse range of cellular processes such as cellular signalling in cell cycle, morphological differentiation, sporulation,

cellular metabolism, stationary phase functions, virulence and so on, hence regulating genes at protein levels.

Nonetheless, after the process of transcription but before translation is initiated, the gene expression in bacteria can also undergo several layers of regulation in collaboration with regulatory RNAs and RNA-binding proteins (RBPs) (Holmqvist and Vogel, 2018; Babitzke *et al.*, 2019; Quendera *et al.*, 2020). Such regulatory events are commonly referred to as post-transcriptional gene regulation. Post-transcriptional gene regulation controls gene expression at the RNA level. Major post-transcriptional gene regulation include processes such as RNA decay, mRNA stability or translation initiation efficiency (Van Assche *et al.*, 2015). Post-transcriptional gene regulation enable bacteria to rapidly adjust to their environments by fine tuning their gene expression to the external stimuli. For example, post-transcriptional regulation is known to regulate processes involved in response to environmental stresses, biofilm formation, toxin-antitoxin systems, initiation of translation, pathogenesis and response to cold shock to name a few (Bertram and Schuster, 2014; Martínez and Vadyvaloo, 2014; Potts *et al.*, 2017; Sauder and Kendall, 2018; Babitzke *et al.*, 2019). The following sections will further describe the type of regulatory RNAs and the roles of RBPs in post-transcriptional gene regulation.

1.1.1 Regulatory RNAs in Bacteria

The development of advanced sequencing technologies has enabled the identification of many RNA molecules that are not of protein coding function, i.e., they are non-coding RNAs (ncRNAs). Many of these ncRNAs are thought to act as post-transcriptional gene regulators. These regulatory RNAs were first identified in bacteria and comprised of a heterogenous group of molecules, which can modulate a range of physiological responses through different regulatory mechanisms (reviewed in Waters and Storz, 2009). Moreover, the regulatory RNAs can be classified based on a few parameters which may include their site of action, the length of the

transcripts and their mode of action. For example, based on their site of action, the regulatory RNAs can be classified as either *cis*-acting or *trans*-acting. The *cis*-acting RNAs are the regulatory RNAs synthesised at or near the site of their target transcripts. On the contrary, the *trans*-acting regulatory RNAs target transcripts that are encoded away from the loci of their origin. Some examples of well-studied *trans*-acting regulatory RNAs in *B. subtilis* include FsrA/S512, SR1/ykzW, RsaE/S415/RoxS, and RnaC/S1022 (reviewed in Mars *et al.*, (2016)). Furthermore, based on the mode of action, the *cis*-acting ncRNAs can be classified as i) *cis*-acting 5' elements ii) *cis*-acting 3' elements (Mars *et al.*, 2016) and iii) *cis*-encoded anti-sense RNAs (asRNAs). Some examples of *cis*-acting anti-sense RNAs in *B. subtilis* include SR5, *ratA* (S976) and *as-bsrH*. SR5 is anti-sense to the *bsrE* gene and forms a part of the type I toxin/antitoxin system in *B. subtilis* (Meißner, Jahn and Brantl, 2016). *ratA* also forms a part of the type I toxin/antitoxin system in *B. subtilis* where it is anti-sense to the expression of *txpA* toxin gene (Nicolas *et al.*, 2012). Whereas *as-bsrH* is anti-sense to the expression of the *bsrH* gene and the binding of *as-bsrH* to the *bsrH* mRNA triggers its degradation.

Furthermore, based on the functions and the length of transcripts, the regulatory RNAs can also be classified as 1) riboswitches, 2) small transcripts that bind proteins, and 3) small non-coding RNAs (sRNAs). The example of *cis*-acting RNAs comprise of riboswitches and anti-sense RNAs (asRNAs). Riboswitches are highly structured regulatory RNAs that are usually found in the 5' untranslated region (UTR) of the mRNA that they regulate (Waters and Storz, 2009; Serganov and Nudler, 2013). The structured folds of riboswitches respond to diverse physiochemical signals/ligands, such as temperatures, pH and or ligands e.g., vitamins, amino acids, nucleotides, ions or transfer RNAs (tRNAs), to exert their regulatory functions (Bastet *et al.*, 2018; Breaker, 2018). In contrast to riboswitches, the asRNAs are transcribed from the opposite strand of an established protein-coding sequence (Waters and Storz, 2009; Storz, Vogel and Wassarman, 2011). Even though asRNAs have regulatory functions,

not all exhibit true regulation. That is, some asRNAs exert their regulation by virtue of their proximity to their target mRNAs rather than their intrinsic regulatory functions (Storz, Vogel and Wassarman, 2011). Figure 1.1 depicts a schematic of the types of regulatory RNAs in bacteria.

The *trans*-acting regulatory RNAs comprise of small non-coding RNAs (sRNAs) or protein-binding RNAs. By binding to proteins or RNAs the *trans*-acting RNAs modulate the functions of their targets. For example, sRNAs can either stabilize or de-stabilize their target mRNAs or block translation, thereby regulating the gene expressions (reviewed in Barquist and Vogel, 2015). A few well-studied small transcripts that bind proteins (protein-binding sRNAs) are the highly conserved *E. coli* CsrB and 6S RNAs (Storz, Vogel and Wassarman, 2011). These small RNAs directly bind to their target proteins rather than RNA targets (Willkomm and Hartmann, 2005; Babitzke and Romeo, 2007; Wassarman, 2007).

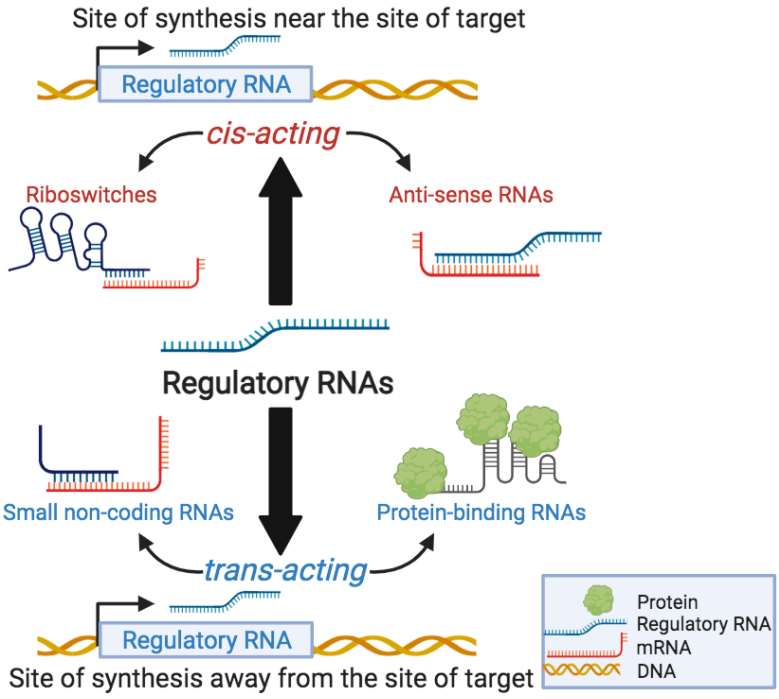


Figure 1.1: Schematic of the different types of regulatory RNAs in bacteria

sRNAs are small non-coding RNAs transcripts that are typically 50-200 nucleotides (nt) long (Gottesman, 2004; Storz, Vogel and Wassarman, 2011; Ahmed, Zheng and Liu, 2016). sRNAs form imperfect base-pairs and interact with their target RNAs (Waters and Storz, 2009). Many sRNAs bind at or near the ribosomal binding sites on their target mRNAs thereby blocking the access to the ribosome, therefore blocking translation (Storz, Vogel and Wassarman, 2011). Several sRNA-mRNA interactions are stabilized by RBPs that function as 'RNA chaperones'. However, the need of RBPs in mediating the sRNA-mRNA interactions is not absolute, i.e., in conditions where the interaction between sRNA-mRNA is stable, requirement of a RBP can be omitted. Furthermore, various studies have also identified roles of RBPs beyond their chaperone activity (reviewed in Quendera *et al.*, 2020). The following section will discuss the roles and properties of RBPs in detail.

1.1.2 RNA-Binding Proteins in Bacteria

RNA-binding proteins are ubiquitous in all living organisms. Nonetheless, many of their precise roles in bacteria are still emerging. In bacteria, RBPs are used as both structural components of larger complexes such as ribosomes, and as regulators of many cellular processes including synthesis, modification, translation, processing and degradation of RNA) (Van Assche *et al.*, 2015; Holmqvist and Vogel, 2018; Babitzke *et al.*, 2019; Holmqvist, Berggren and Rizvanovic, 2020).

1.1.2.1 RNA-binding domains

RBPs interact and bind with their ligands via one or multiple specific domains which are known as the RNA-binding domains (reviewed in Hentze *et al.*, 2018). Some of the classical RNA-binding domains include the S1 domain, cold shock domain (CSD), K homology (KH) domain, RNA recognition motif (RRM), FinO domain and Sm-like domain amongst many

more (reviewed in Hentze *et al.*, 2018). These domains bind to short RNA sequences and form ribonucleoprotein complexes (RNPs), which may have either specific regulatory functions or could be formed due to non-specific binding events (Ferré-D'Amaré, 2016; Helder *et al.*, 2016). The regulatory activity of RBPs can be further influenced by the binding of metabolites, small RNAs, binding to other binding proteins or the chemical modification such as phosphorylation (reviewed in Babitzke *et al.*, 2019). However, a growing number of bacterial RBPs are being recognised that lack a structured RNA-binding domain (Beckmann, Castello and Medenbach, 2016; Pagliuso *et al.*, 2019).

1.1.2.2 RBPs in post-transcriptional gene regulation

In several bacterial post-transcriptional gene regulation mechanisms, the presence of RBPs greatly accelerates the base-pairing process between the interacting sRNA-mRNA molecules (Rajkowitsch and Schroeder, 2007; Panja, Schu and Woodson, 2013). The mode of action of RBPs and their final effects may vary depending on different interactions with sRNAs in four basic ways. The RBPs can (1) inhibit protein translation, by helping the sRNA to bind to the 5' end of its target mRNA; (2) increase protein translation, by assisting the sRNAs to bind to the 5' end of the mRNA hence disrupting the secondary structures which would have otherwise blocked translation; (3) protect some of the sRNAs from the ribonuclease activity or present sRNAs to promote cleavage of target mRNAs; and (4) enhance RNA turnover by making the 3' end of the RNAs accessible for polyadenylation, which will be then processed in a 3' → 5' direction by the exonucleases (Vogel and Luisi, 2011). However, the function of the RBPs appears to be greatly dependent on the structure and conformation of the RNA that it binds to.

Based on the mode of participation of RBPs, the RBPs are generally categorised as (1) Csr (Rsm) Family Proteins, (2) RapZ and (3) ProQ/FinO Family Proteins (reviewed in Babitzke *et al.*, 2019). To understand the roles

of RBPs in post-transcriptional gene regulation, extensive studies have been carried out in the Gram-negative bacteria *Escherichia coli, Salmonella typhimurium* and *Vibrio cholerae.* The RBPs Hfq, CsrA and ProQ are some of the most thoroughly studied RBPs in these species (Quendera *et al.*, 2020; Ng Kwan Lim *et al.*, 2021).

1.1.2.3 Hfq

Hfq is one of the most well-studied RBPs in bacteria. Hfq was initially identified as an essential host factor in *E. coli* for the RNA bacteriophage Qβ (Franze De Fernandez, Eoyang and August, 1968; Vogel and Luisi, 2011). Hfq is highly conserved across many bacterial species including both Gram-positive and negative species. The main role of Hfq is as a mediator of the sRNA-mRNA interactions in the *Enterobacteriaceae E. coli* and *Salmonella* (Soper *et al.*, 2010; Vogel and Luisi, 2011; Santiago-Frangos and Woodson, 2018). Hfq helps sRNAs in finding and regulating the trans-encoded mRNAs via short and imperfect base-pairing (Gorski, Vogel and Doudna, 2017). However, the requirement of Hfq in facilitating RNA-RNA interactions can be bypassed if the stability of the sRNA-mRNA interaction is sufficiently high (Soper *et al.*, 2010). In several of the well-studied Gram-negative bacteria, the absence of Hfq in cells exhibit highly pleiotropic phenotypes and perturbs a range of sRNA-mRNA interactions indicating its global regulatory roles in these bacteria (Sobrero and Valverde, 2012).

Based on these interactions, Hfq has also been extensively investigated in Gram-positive bacteria such as *Listeria monocytogenes* (Nielsen *et al.*, 2009; Kovach *et al.*, 2014), *Clostridium difficile* (Salim *et al.*, 2012; Boudry *et al.*, 2014), *Bacillus anthracis* (Keefer *et al.*, 2017) and *Bacillus subtilis* (Dambach, Irnov and Winkler, 2013). In the *Firmicutes* phylum, Hfq is either absent (e.g., *Lactobacillales*) or, when present, shows minimal participation in gene regulation and bacterial adaptation (Bouloc and Repoila, 2016). The deletion of the genes encoding Hfq in *B. subtilis* (Rochat *et al.*, 2015) and *Staphylococcus aureus* (Bohn, Rigoulay and Bouloc, 2007) do not alter

growth of these bacteria in the 2,000 growth conditions tested, nor in *L. monocytogenes* in the standard laboratory growth conditions (Christiansen *et al.*, 2004). Although Hfq has been shown to bind to several RNAs, no sRNA-mRNA mediated regulation facilitated by Hfq have been identified in *B. subtilis* (Hämmerle *et al.*, 2014; Mars *et al.*, 2015; Rochat *et al.*, 2015). These evidence suggest either the existence of alternative mechanisms without the need of RBPs in the bacterial post-transcriptional gene regulation, the presence of undiscovered RBPs with global regulatory functions, or specific condition dependent RBPs which would eliminate the need of global RBPs.

1.1.2.4 RBPs with dual functions

In addition to mediating sRNA-mRNA interactions, several RBPs are also known to participate in functions other than RNA-binding (Beckmann, Castello and Medenbach, 2016; Holmqvist and Vogel, 2018). The RBPs with more than one function generally lack a conserved RNA-binding domain and are known to possess protein domains with enzymatic activity (Beckmann, Castello and Medenbach, 2016). Such additional activities of RBPs other than their RNA-binding protein activities are often termed as 'moonlighting'.

A classic example of RBP with dual function is the protein aconitase. Aconitase is an iron-containing tricarboxylic acid (TCA) cycle protein. In *B. subtilis* aconitase participates in the sporulation process (Alén and Sonenshein, 1999). Whereas it also autoregulates its mRNA as an RNA-binding protein (Benjamin and Massé, 2014). Other examples of RBPs with enzymatic activity include RodZ, YopD, TruB and RelQ (reviewed in Holmqvist and Vogel, 2018). In addition to the enzymatic activities, many RBPs are also known to have DNA-binding activity (reviewed in Holmqvist and Vogel, 2018). The RBPs such as Hfq also tend to be a part of bigger protein complexes whose nature have been investigated in both *in vitro* and *in vivo* protein-protein interaction studies (Mann, 2006; Gavin, Maeda and

Kühner, 2011; Petschnigg, Snider and Stagljar, 2011; Zhang *et al.*, 2013; Sternburg and Karginov, 2020). These emerging multi-dimensional functions of the RBPs broaden the understanding of the RBPs beyond the sRNA-mRNA interactions, which needs to be understood further.

1.1.2.5 Methods to study RBPs

Several studies have investigated bacterial RBPs, nonetheless the exact number of RBPs and their precise functional roles remain at remain extremely limited even in the model bacterial systems. Global approaches, such as Gradient profiling by Sequencing (Grad-seq) and gradient sedimentation with RNase treatment and mass spectrometry (GradR), have been employed to discover and catalogue a comprehensive repository of microbial RBPs which still remains to be completed (Smirnov *et al.*, 2016; Gerovac *et al.*, 2020). Nevertheless, identifying new RBPs and studying their RNA binding properties *in vivo* is a considerable research challenge and has employed genome-wide high-throughput technologies (Ng Kwan Lim *et al.*, 2021).

A traditional approach to identify the RNA-binding activity of a protein is to identify the RNAs directly binding to the protein of interest *in vivo* and separating the RNA-protein complex via immunoprecipitation. For *in vivo* studies, the vast majority of the methods use cross-linking with ultraviolet (UV) radiation or formaldehyde. UV-cross-linking takes advantage of the covalent linking of amino acids when irradiated with λ = 254 nm which permanently cross-links nucleic acid molecules to the protein (Greenberg, 1979; Lee and Ule, 2018). Whereas, formaldehyde can readily permeate cells and create reversible cross-links between protein–protein, protein–DNA, and protein–RNA complexes (Sutherland, Toews and Kast, 2008; Ramanathan *et al.*, 2019). As RBPs also form multimeric protein complexes, high-throughput proteomic approaches to investigate the roles of RBPs are also employed (Mann, 2006; Gavin, Maeda and Kühner, 2011; Petschnigg,

Snider and Stagljar, 2011; Zhang *et al.*, 2013; Sternburg and Karginov, 2020).

Although the high-throughput methods to study RBPs have huge advantages, such as providing snap-shots of *in vivo* interactions, yet, they are only partly efficient. Many of the limitations of these studies include 1) inclusion of technical variations e.g., partial capture of RNA-binding activity during transient interactions, 2) incorporation of stresses during sample handling 3) the expensive costs of chemicals and equipment involved in these studies that do not make the experiments cost effective, and 4) probability of generation of low signal-to-noise ratios making the data interpretation difficult. Additionally, these sequencing studies do not have a universal method to process the data generated and are influenced by a number of factors.

Additionally, one of the central challenges of RNA-binding studies of RBPs is to determine the binding specificity and the set of targets of the RBP of interest. *In vitro* binding assays coupled with high-throughput sequencing have helped in identifying the binding preference of many RBPs. To substantiate the tentative RNA-protein interactions *in vivo*, methods such as RNA immunoprecipitation and sequencing (RIP-seq) have been developed (Keene, Komisarow and Friedersdorf, 2006; Lu *et al.*, 2014; Sternburg and Karginov, 2020). The related analyses use UV cross-linking of the RNA with proteins followed by the sequencing (CLIP-Seq) of the partially digested short fragment RNAs (Ule *et al.*, 2003, 2005). Several studies also investigated the precise binding site of the co-immunoprecipitated (co-IPed) RNAs and identifying the enriched binding motifs of the RBPs (reviewed in Sternburg and Karginov, 2020).

Similar challenges are also observed in the protein based studied of the RBPs. Based on the labelling techniques used to identify the peptides, the proteomic approaches could be categorised as labelled quantifications or label-free quantifications (LFQ). The use of different instruments for targeted protein quantification with high mass accuracy and high sensitivity enable a robust method to identify hundreds of thousands of peptides from

a pool of peptides (reviewed in Sternburg and Karginov, 2020). The proteomic data analysis is also challenging and varies according to the need of the study. No universal pipeline exists to analyse label-free quantifications of the peptides. However, several bioinformatics tool and packages are available to statistically quantify the changes and identify the biologically significant changes in the proteomics data sets (Röst *et al.*, 2016; Bose *et al.*, 2019).

1.2 Bacillus subtilis

1.2.1 *B. subtilis* as a Model Species

B. subtilis is a Gram-positive bacterium which is commonly found in the soil, roots of plants and in the gastrointestinal tracts of ruminants and humans. It is a rod-shaped bacterium that can form a tough protective endospore in response to extreme environmental stress such as heat and desiccation. *B. subtilis* is the model organism of Gram-positive bacteria and is generally regarded as safe (GRAS) by the Food and Drug Administration (FDA). *B. subtilis* is a mesophilic bacterium that grows optimally in the temperature range of 25-37 ˚C.

The genome of *B. subtilis* was the first completely sequenced Gram-positive bacteria, and harbours a single circular genome of 4,215,606 base pairs (bp) comprising 4,100 protein-coding genes and 174 RNA genes (Kunst *et al.*, 1997), although only 257 protein-coding genes are essential for the cultivation of *B. subtilis* in the LB (Koo *et al.*, 2017). *B. subtilis* is one of the best-characterized Gram-positive bacteria and has a low GC genomic content (43%) (Kunst *et al.*, 1997). It also secretes a variety of secondary metabolites and antibiotics that are typically associated with *Streptomyces* species (Kunst *et al.*, 1997).

Another property of *B. subtilis* is its ability to directly secrete large amounts of proteins into its culture media (reviewed in Tjalsma *et al.*, 2000, 2004).

Many of the secreted proteins belong to the class of enzymes such as RNases, DNases, proteases, lipases and carbohydrases that participate in the hydrolysis of polymers found in the natural habitat of *B. subtilis* (Tjalsma *et al.*, 2000, 2004). By secreting these degradative enzymes, the bacterium attempts to adapt to its environment and derive the available resources as part of an adaptive response (Ferrari, Jarnagin and Schmidt, 1993; Tjalsma *et al.*, 2000). This property of *B. subtilis* has been utilized in the biotechnology industries for the production of enzymes such as amylases, proteases, pullulanases, chitinases, xylanases and lipases (reviewed in (van Dijl and Hecker, 2013)). *B. subtilis* is exploited in the biotechnology industry robustly for the production of such enzymes, which comprises of approximately 60% of the total industrially produced enzymes (Harwood, 1992; Morikawa, 2006; van Dijl and Hecker, 2013). Thus, this makes *B. subtilis* an important biotechnological tool and is therefore also referred to as a molecular biological and industrial workhorse (Harwood, 1992; Morikawa, 2006; van Dijl and Hecker, 2013). Figure 1.2 shows an overview of *B. subtilis* and several of its properties.

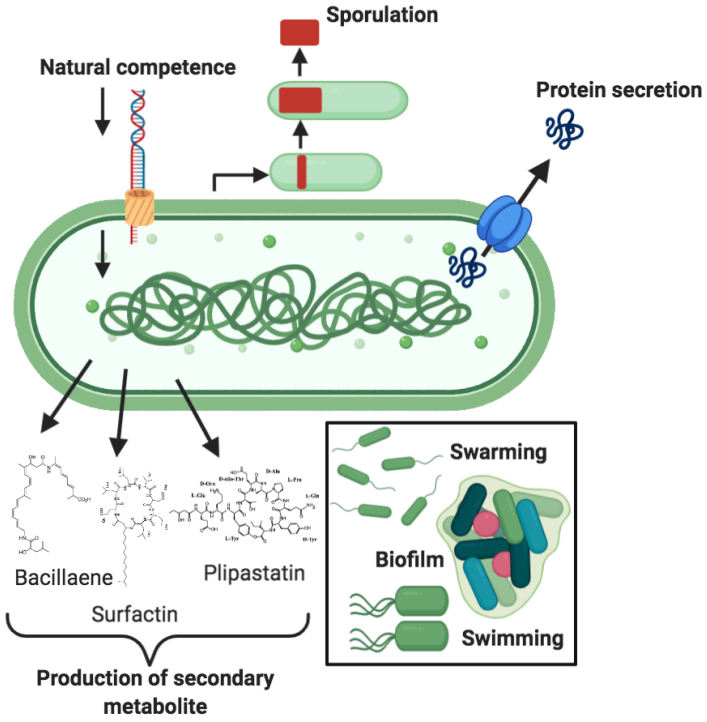


Figure 1.2: Overview of several properties of *B. subtilis*

B. subtilis can differentiate into a variety of cell types in different growth conditions (Lopez, Vlamakis and Kolter, 2009). For example, (1) at the onset of stationary phase, some cells differentiate into competent cells that are capable to take up exogenous DNA from their surroundings; (2) under nutritional starvation conditions *B. subtilis* may sporulate and differentiate into dormant cells that are highly resistant to external stress; (3) while some cells produce killing factor and toxin to kill cells that have not yet started sporulation and are thus referred to as cannibalistic cells and (4) some cells differentiate into extracellular matrix producing cells that assist biofilm formation (reviewed in Lopez, Vlamakis and Kolter, 2009). In addition to these properties, *B. subtilis* also displays different kinds of motility in liquid and on solid surfaces (e.g., swimming, swarming and sliding), chemotaxis and pigment formation (Chu *et al.*, 2006). In summary, *B. subtilis* has become a model organism to study attachment to plant root or fungal hyphae, biofilm development, cell division, protein secretion, secondary metabolite production and surface motility (swimming, swarming, and sliding).

1.2.2 Gene Regulation in *B. subtilis*

Free-living bacteria must sense and respond to their environment rapidly to survive. In doing so, they regulate their gene expression and undergo complex signalling processes (Arrieta-Ortiz *et al.*, 2015; Escorcia-Rodríguez, Tauch and Freyre-González, 2020). The changes in gene expression could be either be a direct response to the external stimuli or can occur stochastically (by chance). Sometimes, these changes in gene regulation could be a result of both the direct and the stochastic changes. *B. subtilis* is one of the most versatile bacteria and exhibits a large range of responses such as motility to the formation of specialised cell types in different growth conditions (Losick, 2020). These responses can further be attributed to either their single cellular responses, or to the multi-cellular responses. Some examples of cellular responses of *B. subtilis* include

motility (swimming and swarming) and biofilm formation. While motility in *B. subtilis* could be a single cell property, the formation of biofilms in different natural habitats such as roots of the plants or in chemically defined media in the laboratories is largely a multi-cellular process which encompasses complex signalling and gene regulation mechanisms.

Another interesting property of *B. subtilis* is its state of competence (ability to take up exogeneous DNA from the environment) and the process of sporulation. Competence and sporulation in *B. subtilis* are generally used as model system to study cell differentiation in many other bacteria. In both these processes, the gene expression is tightly regulated by an array of sigma factors. Including the sigma factors involved in competence and sporulation, there are ~20 different sigma factors known in *B. subtilis* (Zhu and Stülke, 2018; Burton *et al.*, 2019). Table 1.1 summarises few of the commonly known sigma factors in *B. subtilis*.

Table 1.1: Overview of sigma factors in *B. subtilis*

Sigma factor	Category	Regulates genes related to
SigA	Household	household
SigB	Stress	general stress
SigD	Motility	chemotaxis and motility genes
SigE	Sporulation	early mother cell-specific sporulation
SigF	Sporulation	early forespore-specific sporulation
SigG	Sporulation	late forespore-specific sporulation
SigH	Transition	transition phase
SigK	Sporulation	late mother cell-specific sporulation
SigL	Enhancer	enhancer-dependent (Sigma-54 family)
SigM	ECF	Extracytoplasmic function (ECF), high salt concentrations
SigV	ECF	ECF
SigW	ECF	ECF, cell wall stress
SigX	ECF	ECF
SigY	ECF	ECF
SigZ	ECF	ECF

Amongst all the sigma factors in *B. subtilis*, SigA is the primary sigma factor. SigA is active in the exponential/vegetative growth of *B. subtilis* and regulates over 800 genes (Ramaniuk *et al.*, 2017). For cells experiencing general stress, such as heat or osmotic stress, SigB serves as the primary sigma factor that regulates more than 200 genes (Schumann, 2016). Whereas properties such as motility and chemotaxis in *B. subtilis* are regulated by the sigma factor SigD.

The genes under the control of sigma factors SigM, V, W, X, Y and Z are regulated for the expression of extra cytoplasmic functions (ECF) such as the production of extracellular matrix and biofilm formation. For the formation of an endospore, >500 genes are cumulatively expressed and regulated by the sigma factors SigE, F, G, and K. The gene expression in the cells undergoing the transition state of cell cycles, SigH serves as the main sigma factor. Whereas, the genes regulated by the SigL have emerged as an important link between central carbon and nitrogen metabolism in *B. subtilis* (Choi and Saier, 2005).

1.2.3 Sporulation in *B. subtilis*

The study of sporulation sheds insights in both the processes of cellular differentiation and developmental processes in bacteria. Sporulation in *B. subtilis* is induced under nutritional starvation conditions which results in the asymmetric cell division and in the formation of an endospore – a metabolically inert body (Dawes, Kay and Mandelstam, 1969; Nicholson *et al.*, 2000; Errington, 2003). However, a number of other methods are also adopted by cells to escape sporulation before finally committing to the irreversible sporulation pathway. For example, some cells activate their flagellar motility to look for more food via chemotaxis, production of antibiotic compounds to kill competing bacteria in the shared ecological niche, increase secretion of degradative enzymes to look for extracellular proteins and polysaccharides from the degraded bacteria and materials around their surroundings (Schultz *et al.*, 2009). The stressed cells also examine the

state of the chromosomal integrity and chromosomal replication to ensure successful completion of sporulation once committed to the process. In situations where the alternate processes fail to support the cells, the sporulation process is initiated as the final fate of the stressed cells. An overview of the stages of sporulation is illustrated in Figure 1.3.

Endospores (henceforth referred as spores) can survive long periods of dormancy and harsh conditions until favourable conditions are achieved by the cells to germinate and resume their vegetative growth. The sporulation process in *B. subtilis* goes through eight major stages of morphogenesis starting from the vegetative cell at Stage 0 and reaches till a free spore is released at Stage VII (Narula *et al.*, 2012; Tan and Ramamurthi, 2014; Khanna, Lopez-Garrido and Pogliano, 2020). Sporulation is triggered as *B. subtilis* undergoes asymmetric cell division in response to the levels of the sporulation master regulator Spo0A (reviewed in Schultz *et al.*, 2009).

Once the commitment stage is reached, the asymmetric division leads to the formation of two compartments with the cell. These compartments are genetically identical yet morphologically different and are termed as the ‘mother cell’ (larger) and the ‘forespore’ (smaller). Both the compartments briefly remain attached side-by-side. The mother cell engulfs the forespore and eventually the forespore becomes the endospore. Whereas the mother cell perishes, as the sporulation cycle is complete.

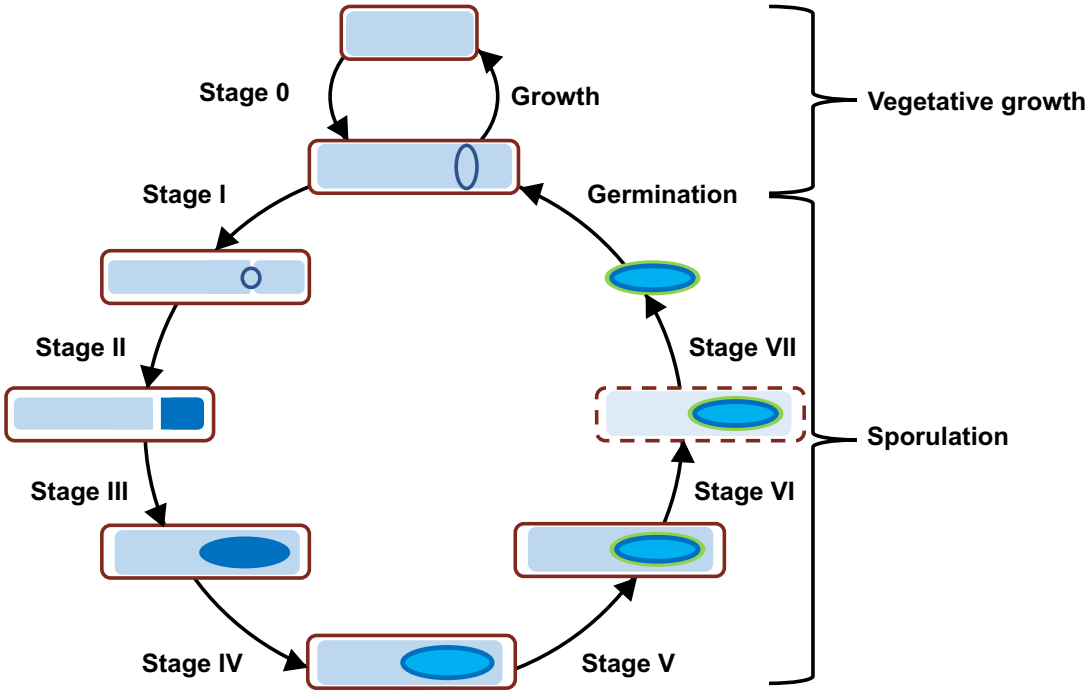


Figure 1.3: Stages of sporulation in *B. subtilis*

The scheme represents the key architectural transformations that occur in sporulation. The detailed stages are reviewed in Khanna, Lopez-Garrido and Pogliano, 2020. In summary, sporulation is initiated (Stage 0) from the vegetative *B. subtilis* cells cycle and enters Stage I under unfavourable growth conditions (anti clockwise). Stage I – marks the initiation of formation of an axial DNA filament. Stage II – represents the formation of polar septum, creating the two compartments - mother cell (larger) and forespore (smaller) and the initiation of engulfment of the forespore by the mother cell. The septa dividing the mother cell and the forespore also goes through stages of flat, curved and engulfing septa formation. Stage III – symbolises the completion of engulfment of the forespore by the mother cell. Stage IV – initiates the assembly of cortex between engulfed forespore. Stage V – completes the formation of the spore coat around the engulfed forespore. Stage VI – the spore gains resistance to heat and prepares for the release. Stage VII – marks the initiation of the lysis of the mother cell and release of the free endospore to the environment. Stage VIII – involves the germination of the free endospore back to the vegetative *B. subtilis* cell in the presence of enough nutrition.

1.2.4 RNA-Binding Proteins in *B. subtilis*

Some of the earliest known RBPs in *B. subtilis* include the *trp* RNA-binding attenuation protein (TRAP) (Babitzke, Bear and Yanofsky, 1995), CsrA (Romeo, 1998; Yakhnin *et al.*, 2007) and the protein aconitase (Alén and Sonenshein, 1999). Investigation of Hfq in *B. subtilis* also found it to be an RBP (Dambach, Irnov and Winkler, 2013), however it plays no role in the sRNA-mRNA mediated post-transcriptional gene regulation of the bacterium (Hämmerle *et al.*, 2014; Rochat *et al.*, 2015). Also, among the RBPs in *B. subtilis*, only a handful are known to have regulatory functions. Some of these regulatory RBPs include CitB, CsrA, HutP, LicT, PyrR, and SacY and RNases such as RNase Y (global endoribonuclease), RNase J1 (global exoribonuclease), RNase III and polynucleotide phosphorylase (PNPase) (Stülke, 2002; Durand *et al.*, 2012; Lehnik-Habrink *et al.*, 2012; Babitzke *et al.*, 2019; Quendera *et al.*, 2020). Although these RBPs support gene regulation at multiple layers, none of the RNA chaperones function as global regulators of gene expression in *B. subtilis*.

A study by Burke and Portnoy (2016) reported SpoVG as a novel RBP in the Gram-positive pathogen *L. monocytogenes in vitro* and proposed that it might act as a global post-transcriptional gene regulator *in vivo*. Nevertheless, it is not clear if this protein is functional *in vivo* and if it has similar properties in other Gram-positive bacteria. However, a study in *S. aureus* highlighted participation of SpoVG in the bacterial infection as a transcription factor (Bischoff *et al.*, 2016). SpoVG and its homologs are commonly found proteins in eubacteria and have been studied in several bacteria as also having DNA binding properties (Jutras *et al.*, 2013; Bischoff *et al.*, 2016). SpoVG belongs to one of the most conserved family of proteins (Kim, Lee and Kwon, 2010). Nevertheless, nothing is known about the roles of SpoVG in *B. subtilis* as an RNA-binding protein or as a global regulator. Hence, these evidence provide scope to test and prove the RNA-binding properties of SpoVG *in vivo* and investigate their interactions with sRNAs in *B. subtilis*.

1.3 SpoVG in *B. subtilis* and Beyond

spoVG was first identified in *B. subtilis* and the initial investigation into its functional role identified it as a highly-conserved gene across bacterial species (Margolis, Driks and Losick, 1993; Matsuno and Sonenshein, 1999; Kim, Lee and Kwon, 2010; Jutras *et al.*, 2013). According to the existing literature, *spoVG* is described as a sporulation gene in *B. subtilis* which is 291 bp in length and is translated into a 97 amino acid protein (Sandman, Losick and Youngman, 1987; Matsuno and Sonenshein, 1999). The protein is known to be phosphorylated on Ser-66/Ser-67/Thr-68 and belongs to the SpoVG family (according to Swiss-Prot) of proteins (Macek *et al.*, 2007; Michna *et al.*, 2016). Researchers gained interest in *spoVG* in *B. subtilis* due to several of its intriguing properties such as 1) transcription of the gene during sporulation starts very early (stage II), yet the mutant phenotype (i.e., blockage in the sporulation process) is not detected until 5 hours into the process of sporulation (stage V), 2) *spoVG* has two condition specific overlapping promoters, which are active differentially based on different developmental states, 3) the protein SpoVG has partially overlapping functions with another sporulation specific protein SpoIIB (Moran Jr. *et al.*, 1981; Rosenbluh *et al.*, 1981; Wang and Doi, 1984; Sandman, Losick and Youngman, 1987; Zuber, Healy and Losick, 1987; Margolis, Driks and Losick, 1993; Matsuno and Sonenshein, 1999). SpoVG has been classified as a negative regulator of asymmetric septation during sporulation (Matsuno and Sonenshein, 1999).

The transcription of *spoVG* is under the control of the sigma factor SigH (Johnson, Moran and Losick, 1983; Carter and Moran, 1986). Furthermore, research by Hudspeth and group suggest a putative Rho-independent transcriptional termination of *spoVG* transcription which was confirmed by the transcription profiling of *B. subtilis* (Hudspeth and Vary, 1992; Nicolas *et al.*, 2012). Researchers interested in *B. subtilis* regulation identified mutations in early sporulation genes such as *spo0A*, *spo0B*, *spo0E*, *spo0F* and *spo0H* that have a limiting effect on the expression of *spoVG* (Moran Jr.

et al., 1981; Johnson, Moran and Losick, 1983; Zuber and Losick, 1987). This research suggested regulation of *spoVG* by *spo0* (*A, B, E, F, H*) during sporulation. Following this, a study of the transcription regulator, AbrB (transcriptional regulator of gene expression during transition from the exponential phase to stationary phase) highlighted that AbrB has a binding site in two genes, *tycA* and *spoVG*, where it binds to *spoVG* in the A+T rich region upstream of the promoter (Marahiep, Republic and Highway, 1991). Additionally, it was shown that AbrB negatively regulates *spoVG* by blocking the expression of *spo0A* and *spo0B*, hence, inhibiting the transcription of *spoVG* during sporulation (Zuber and Losick, 1987). Research from other groups also identified repression of *spoVG* by SinR, a master regulator of multicellularity and formation of biofilms in *B. subtilis* (Chu et al., 2006).

Despite the knowledge that we have about *spoVG*, its exact roles and mechanisms of action in *B. subtilis* still remain unclear. Research from Margolis showed that function of SpoIIB (another sporulation protein in *B. subtilis*) is redundant to SpoVG in sporulation and that the allelic variation in the *spoVG* mutant varies based on different laboratory strains of *B. subtilis* (Margolis, Driks and Losick, 1993). Based on the results from the *spoIIB spoVG* double mutant (strongly blocked at septation), they also postulated that these genes have implications in the maturation of sporulation septum and activation of sporulation sigma factors SigE and SigF (Margolis, Driks and Losick, 1993). In addition to this, work from Resnekov proposes a similar property of *spoVG* with *spoIIB*, where they observed that a single deletion mutation of *spoVG* does not restrict the cell to undergo sporulation to a great extent, whereas, the double mutation of *spoVG* and *spoIIB* halts the sporulation completely (Resnekov *et al.*, 1995). However, this blockage was overcome by a third mutation in another gene, *spoVS*. Hence, it indicated an interplay and participation of three different genes under sporulation, underlining the complexity of bacterial regulation.

The discovery by Burke and Portnoya, (2016) showed that SpoVG is an RBP in *L. monocytogenes in vitro*. In the study it was shown that the 5' UTR of *spoVG* interacts with the non-coding RNA Rli31 *in vitro* (Burke and

Portnoy, 2016). SpoVG is also a known RNA-binding and DNA-binding protein in *Borrelia burgdorferi* (Savage *et al.*, 2018). The correlation of the protein and the mRNA levels of *spoVG* mutant did not corelate in *Borrelia*, suggesting protein levels are controlled in part via post-transcriptional mechanism(s) (Savage *et al.*, 2018). Studies in *S. aureus* have also identified *spoVG* to be negatively regulated by the small RNA SprX (RsaOR) (Eyraud *et al.*, 2014). Recent investigations in *B. anthracis* have established the indispensable role of SpoVG in the regulation of sporulation (Chen *et al.*, 2020). Figure 1.4 summarises the properties and functions of SpoVG in *B. subtilis* and other bacteria.

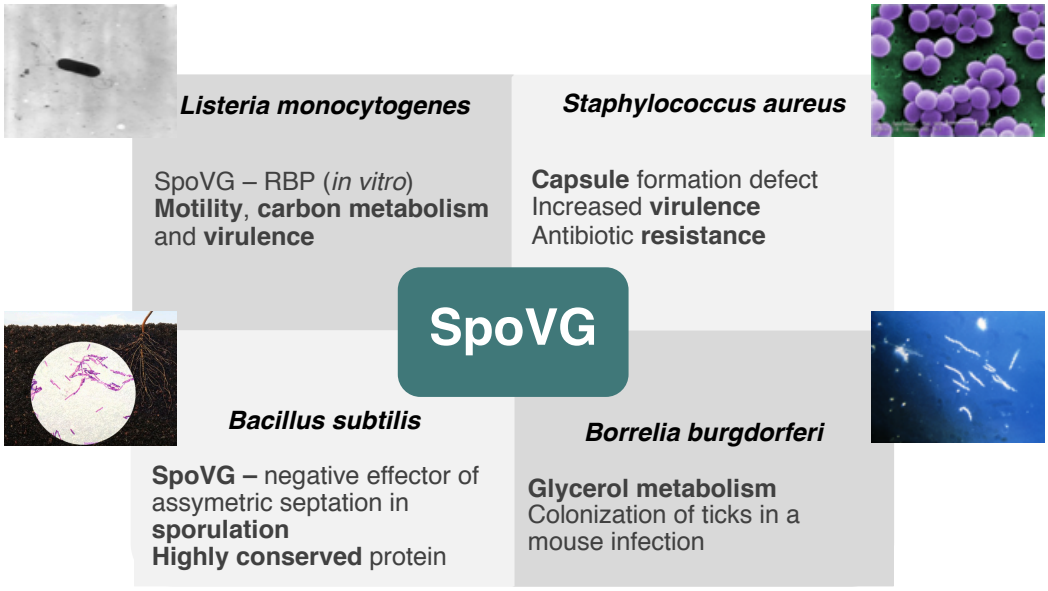


Figure 1.4: Overview of roles and function of SpoVG in *B. subtilis* and other bacteria

From the work of Nicolas *et al.*, (2012) it is known that *spoVG* has constitutive expression in all the growth conditions tested in *B. subtilis*. Additionally, the expression of *spoVG* is also relatively higher than many known RBPs in *B. subtilis* (Figure 1.5). Therefore, it becomes intriguing to understand the role and properties of *spoVG* in *B. subtilis* in further detail.

1.4 Summary and Thesis Structure

Despite the advances in the field of RBPs and the understanding of post-transcriptional gene regulation, there is a paucity of knowledge regarding global RNA-binding regulators in *B. subtilis* (Ul Haq, Müller and Brantl, 2020). The lack of a global regulator of pots-transcriptional gene regulation in *B. subtilis* is intriguing and prompts us to question the mode of post-transcriptional gene regulation in this bacterium.

SpoVG has been presented as a lead candidate as an RNA-binding protein and a potential global gene regulator in *L. monocytogenes*, *S. aureus* and *B. burgdorferi*. Therefore, in this thesis I present a multi-omics approach in characterising SpoVG for its RNA-binding activity and regulatory effects in *B. subtilis*.

Chapter 2 describes the cumulative materials and methods including all the computational and the wet-lab approaches employed in this investigation. Chapter 3 presents the global transcriptomic analysis of the Δ*spoVG* in *B. subtilis* in three different growth conditions (Figure 1.6). The data analysis of the study comprises the investigations of the differentially expressed genes in all the conditions and the analysis of the biological pathways perturbed in the absence of *spoVG*.

Whereas Chapter 4 discusses the results from a comparative total proteomic analysis of the Δ*spoVG* mutant as in Chapter 3 (Figure 1.6). The investigation also includes a nutrient rich medium growth condition and assesses the differential protein production in the mutant. It also discusses the changes in the biological pathways affected in the absence of SpoVG.

The Chapter 5 explores the RNA-binding activity of SpoVG in *B. subtilis in vivo* and identifies the bound RNAs to SpoVG using UV-crosslinking and sequencing techniques. Further, Chapter 6 examines whether SpoVG interacts with any proteins and forms protein-protein complexes, via protein-protein cross-linking and proteomic analysis.

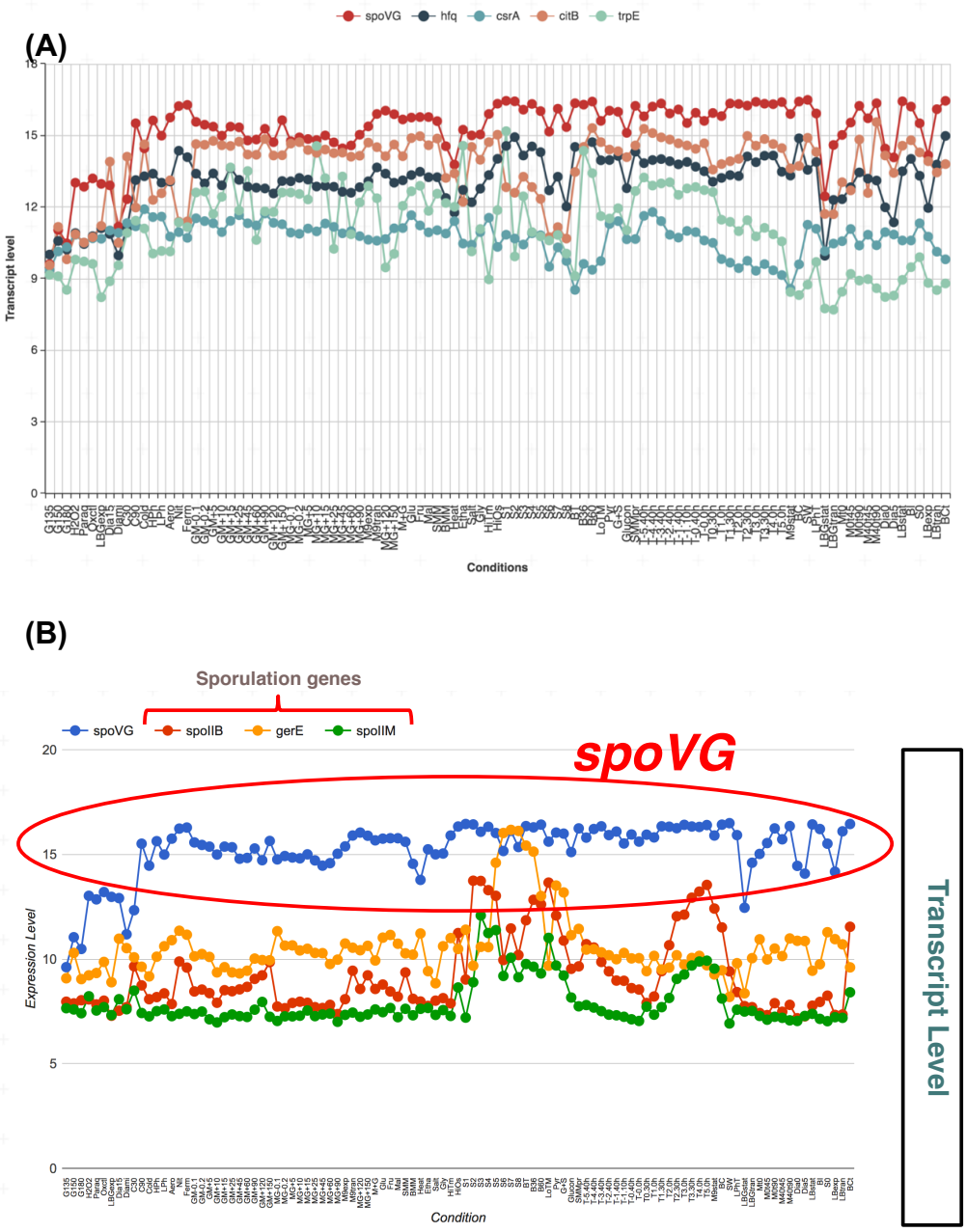


Figure 1.5: Level of transcription of *spoVG* across some conditions compared to

(A) Some of the RBPs (genes indicated on the top of the graph) in *B. subtilis* tested in the global gene expression profile in Nicolas *et al.*, (2012). The conditions indicate the growth conditions in which *B. subtilis* was grown. **(B)** Sporulation genes in *B. subtilis.*

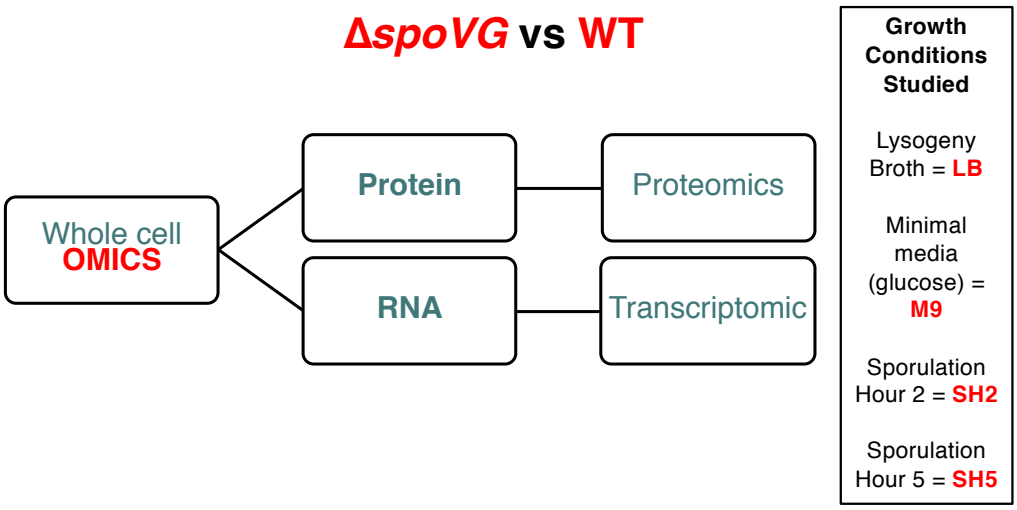


Figure 1.6: Approach taken in this thesis to study global -oimcs of Δ*spoVG* in *B. subtilis*

Next, Chapter 7 functionally characterises the Δ*spoVG* mutant in connection to the changes identified in the -omics study (Chapter 3 and 4) using wet-lab techniques. Finally, the Chapter 8 discusses the overall outcomes of these investigations with comments on the future directions of this work.

Chapter 2 Materials and Methods

2.1 Bacterial Cell Biology

2.1.1 Growth Media

2.1.1.1 LB

Bacillus subtilis strains were routinely cultured in lysogeny broth (LB). Typically, a fresh colony was inoculated in LB with appropriate antibiotics wherever necessary and grown at 37˚C with continuous shaking at 200 rpm overnight. Precultures were set up for some experiments, where cultures were diluted in 1 in 100 and grown for ~2.5 hours prior to the final culture. The final growth cultures were set up with a starting optical density (OD_{600}) of 0.05 and grown for 8 hours or up to the required OD_{600}.

Concentrations of antibiotics: Ampicillin: 100 µg/mL; Chloramphenicol: 5 or 35 µg/mL; Erythromycin: 2 µg/mL; Phleomycin: 4 µg/mL; Spectinomycin: 100 µg/mL; Tetracycline: 10 µg/mL.

2.1.1.2 M9

Cultures were set up in minimal media M9 as per standard operating procedure (SOP) of the BaSysBio project (Botella *et al.*, 2010). In summary, a fresh colony of *B. subtilis* was picked from an LBA plate with appropriate antibiotics and inoculated into LB with antibiotic and grown for 2-4 hours at 37˚C shaking at 200 rpm to an OD_{600} of 0.3-1.0. For regular experiments, M9 was supplemented with glucose (0.3%) (Nicolas *et al.*, 2012). Alternate carbon sources were used if required (e.g. sucrose 0.3 %, fructose (0.3%), succinate (0.3%), galactose (0.3%), glycerol (0.3%), arabinose (0.3%), xylose (0.3%), myo-inositol (10 mM), malate (0.3%) (Yoshida *et al.*, 2008;

Nicolas *et al.*, 2012). 18 mL of M9 medium was added to 250 mL culture flasks and precultures were set up in three dilutions – 2,000x, 4,000x and 6,000x for 12-16 hours (overnight). Preculture with OD_{600} 0.2-1.0 was selected and final culture was set up with a starting OD_{600} of 0.05. Cultures were grown at 37˚C in shaking water bath or incubator until the required OD_{600}.

2.1.1.3 Sporulation

The sporulation experiments were adapted from the Sterlini-Mandelstam protocol (Sterlini and Mandelstam, 1969). In summary, a freshly streaked colony of desired *B. subtilis* strain was re-suspended in 200 µL of Casein hydrolysate (CH) media (composition described in Appendices). A 1 in 1,000 dilution was made in fresh CH media and incubated overnight with shaking (200 rpm) at 37˚C. The overnight culture was diluted 1 in 20 and grown until OD_{600} 0.2 at which point it was then diluted to OD_{600} 0.1 and allowed to grow until OD_{600} 0.6-0.8. The cells were pelleted by centrifugation (12,000 x *g*, 10 min, 25˚C) and re-suspended in re-suspension media (RM) (components described in Appendices) to enable the process of sporulation. The cells were then cultured at 37˚C from 0 – 24 h depending on the requirements of the experiment.

2.1.2 Bacterial Strains

For all the routine laboratory experiments and the -omics' studies a Trp+ version of the 168 strain (trpC2), BSB1, was used. For some of the phenotypic experiments described in Chapter 7, the undomesticated NCIB 3610 strain was also used. All bacterial strains used throughout this project are listed in Table 2.1.

Table 2.1: *B. subtilis* strains used in this study

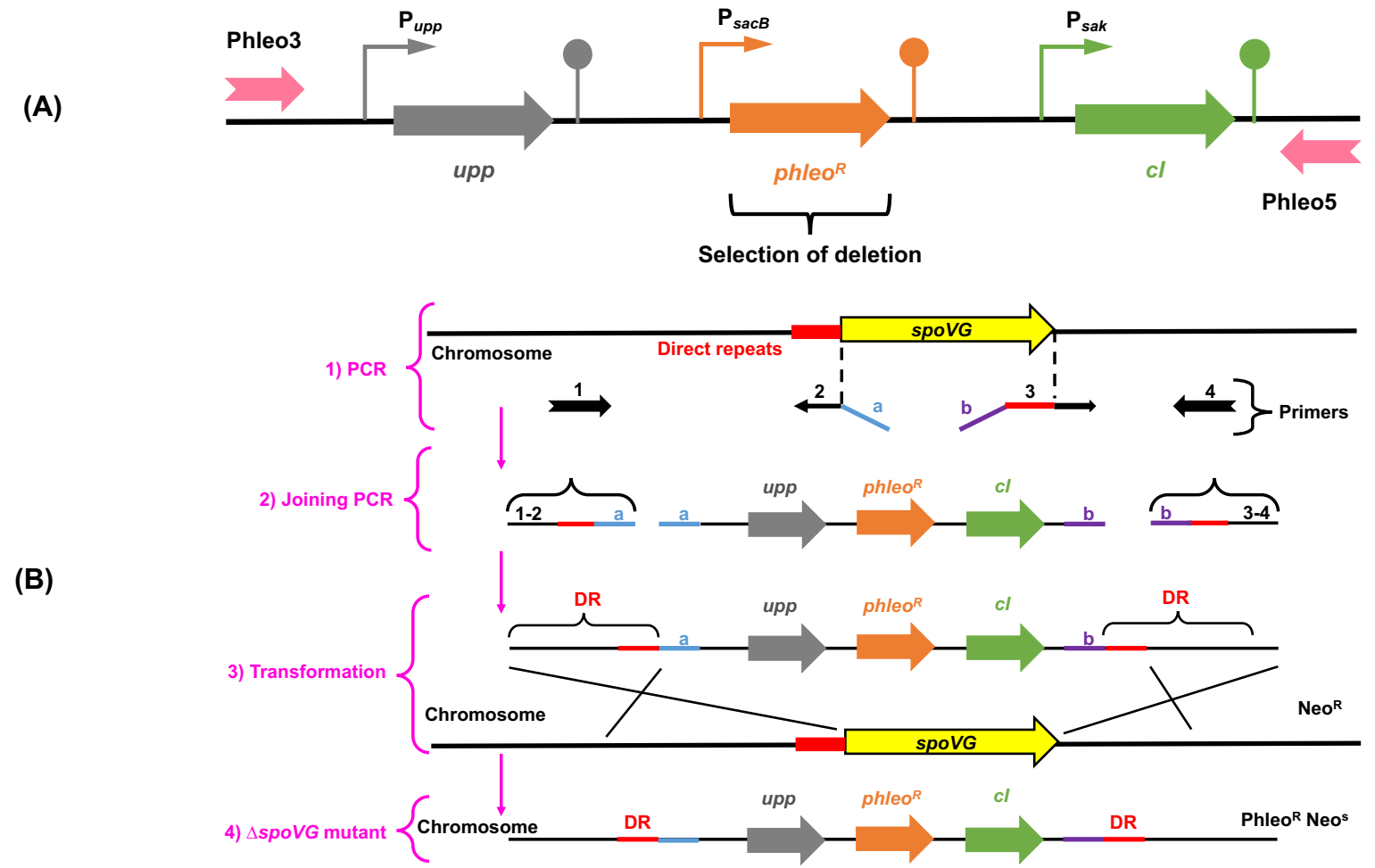
Strains	Description	Resistance	Source
BSB1 (168, *trp*+)	Laboratory strain wild-type *B. subtilis*	None	(Nicolas *et al.*, 2012)
BSB1 *spoVG::phl* (Δ*spoVG*)	Chromosomal deletion of *spoVG* by a reciprocal recombination and replacement with the *phl* gene (encoding for phleomycin)	Phleomycin	Emma L Denham
BSB1 *spoVG*::*phl, amyE*::*spoVG* (Δ*spoVG*::s*poVG)*	Chromosomal complement of *spoVG* at *amyE* locus	Tetracycline	This study
168, *trp*+ SpoVG-Flag	Chromosomal integration of SpoVG-FLAG at *amyE* locus in the Δ*spoVG* strain	Tetracycline	This study
168, *trp*+ SpoVG-HTF	Chromosomal integration of SpoVG-HTF at *amyE* locus in the Δ*spoVG* strain	Tetracycline	This study
NCIB 3610	Undomesticated strain wild-type *B. subtilis*	None	BGSC
NCIB 3610 Δ*spoVG*	Chromosomal deletion of Δ*spoVG*	Phleomycin	This study
NCIB 3610 Δ*spoVG*::*spoVG*	Chromosomal complement of *spoVG* at *amyE* locus	Tetracycline	This study

2.1.2.1 Construction of ∆spoVG

The ∆*spoVG* knockout deletion strain was constructed according to the Standard Operating Procedure (SOP) BaSynthec (Fabret, Ehrlich and Noirot, 2002). Briefly, ~1.5 kb upstream and downstream of *spoVG* was amplified using the primer set P1_spoVG-P2_spoVG and P3_spoVG-P4_spoVG (Table 2.3). The phleomycin deletion cassette was amplified from a plasmid (pUC19-K7-010) (from Noirot laboratory) using the primers phleo3 and phleo5 (Table 2.3). Using joining PCR, the upstream and the downstream fragments of *spoVG* were ligated to the deletion cassette (Figure 2.1). The successful construct was transformed into the *B. subtilis* master strain (Westers *et al.*, 2002), with genotype ΔSPβ, ΔSkin, ΔPBSX *upp*::P_{lambda}*neo* (*upp*::P_{lambda}*neo* is essential for the pop-out system (Itaya *et al.*, 2005)), which would replace *spoVG* with homologous recombination (Figure 2.1). The transformed cells were selected on LBA phleomycin plates. The successful colonies with ∆*spoVG* deletion were confirmed by colony PCR using primer set P0_spoVG-P4_spoVG. Glycerol-LB stocks were made for the successful deletion colonies.

2.1.2.2 Construction of BSB1 ∆spoVG::spoVG

For constructing the complement strain of *spoVG* (∆*spoVG*::*spoVG*)*, spoVG* with its native promoter and terminator was cloned into the shuttle vector pRMC (construction described in section 2.2.3.4) (Figure 2.2). The successful clone was transformed into the ∆*spoVG* mutant and selected on LBA tetracycline plates. To test the correct chromosomal integration of plasmid spoVG_pRMC into the *B. subtilis* genome (Figure 2.2), single colonies from the transformed plates were tested on LB agar containing 1% starch. A loss of a zone of starch clearing when exposed to iodine was used to confirm the correct integration due to the loss of *amyE*.


(A)
Phleo3
P upp
P sacB
P sak
upp
phleo R
cI
Phleo5
Selection of deletion
(B)
1) PCR
Chromosome
Direct repeats
spoVG
1
2
a
b
3
4
Primers
2) Joining PCR
1-2
a
upp
phleo R
cI
b
3-4
3) Transformation
DR
Chromosome
Neo R
4) ΔspoVG mutant
Phleo R Neo S

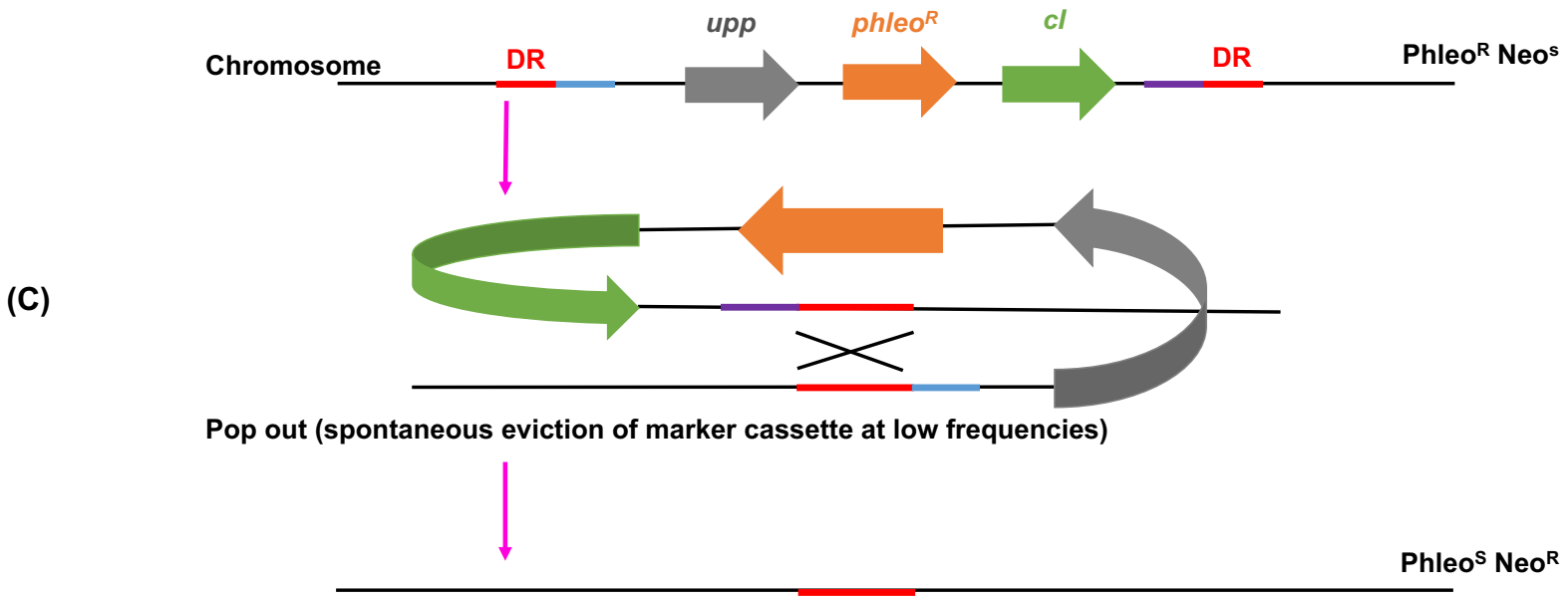


Figure 2.1: Schematic of the deletion system used to generate the Δ*spoVG* mutant

The system shows the **(A)** schematic of the drug resistant phleomycin deletion marker cassette used to select the single mutant strain. **(B)** A schematic representation of the construction of the Δ*spoVG* mutant construction using the deletion marker cassette in *B. subtilis*. **(C)** Spontaneous eviction of deletion marker cassette happens at lower frequencies, hence generating marker-less Δ*spoVG* deletion mutant.

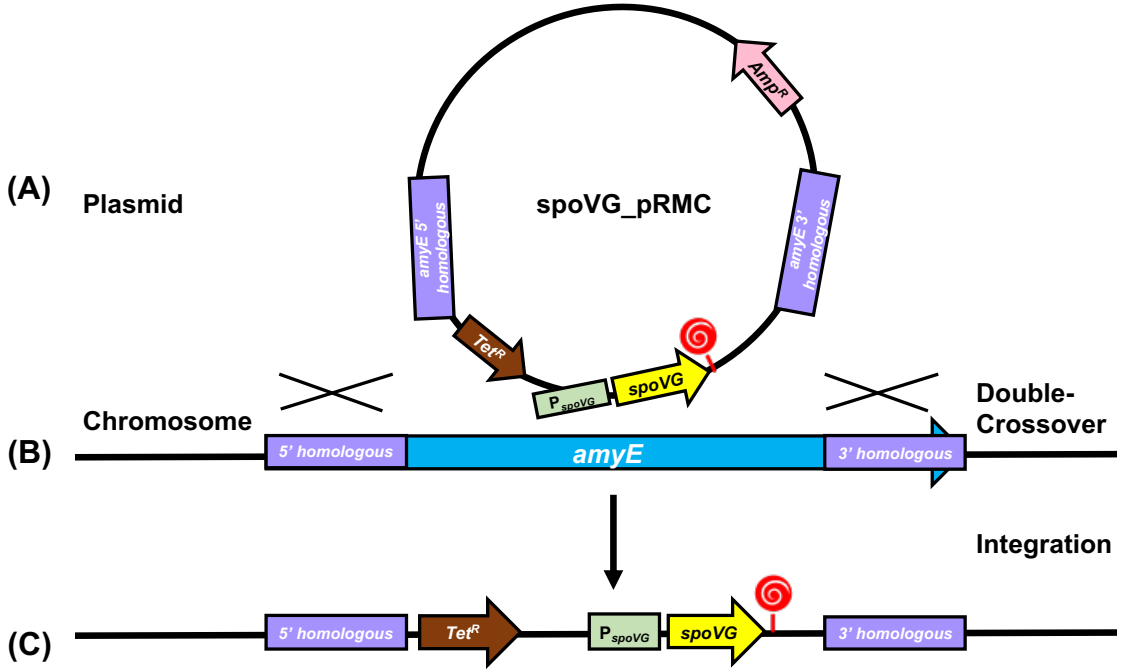


Figure 2.2: Schematic of construction of the *spoVG* complement strain (BSB1 Δ*spoVG*::*spoVG*)

The schematic shows the mechanism of complementation of *spoVG* into the Δ*spoVG* mutant. **(A)** The plasmid spoVG_pRMC (shuttle vector) carries *spoVG* with its native promoter and transcription termination sequence (illustrated with a red lollipop), a tetracycline resistance gene (Tet^R); and 5' and 3' homologous sequence of *amyE* gene. **(B)** Represents the homologous recombination between the plasmid spoVG_pRMC and the *amyE* locus in Δ*spoVG* mutant. **(C)** Illustrates the construct of the complemented BSB1 Δ*spoVG*::*spoVG* in *B. subtilis*.

2.1.3 Bacterial Transformations

2.1.3.1 Transformation of B. subtilis in Paris medium

B. subtilis strains were routinely genetically transformed in Paris medium (PM) as described in the 'Molecular biological methods for *Bacillus'* (Harwood and Cutting, 1990). Briefly, a single colony of *B. subtilis* was grown in PM overnight at 37˚C with shaking. The following day, the culture was diluted 1 in 50 into fresh PM and grown for 3 hours. Genomic DNA, plasmids (from TG1 *E. coli*) or PCR products were added and cultured for further 5 hours. To select for transformants the culture was plated out on LBA plates supplemented with suitable antibiotics.

2.1.3.2 Transformation of E. coli

Preparation and transformation of calcium-competent *E. coli* DH5-α

Making calcium-competent E. coli DH5-α

A freshly streaked colony of DH5-α *E. coli* was cultured in 5 mL LB broth at 37˚C with 200 rpm shaking overnight. The overnight culture was diluted 1 in 100 in fresh LB and grown at 37˚C with 200 rpm shaking until OD_{600} 0.3-0.4 (~1-2 h). The culture was then incubated on ice for 30 min and the bacteria were pelleted by centrifugation at 2,000 x *g* for 15 min at 4˚C. The supernatant was removed and discarded. The pellet was gently re-suspended in 1 in 5 volume of cold 0.1 M $CaCl_2$ by swirling the flask/tube and spun at 1,100 *g* for 15 min at 4˚C. This step was repeated twice. Finally, the cells were gently re-suspended in 1 in 25 volume of 0.1 M $CaCl_2$ and 15% glycerol solution. The re-suspended cells were split into aliquots of 50-100 µL and stored at -80˚C until transformations.

Transformation of E. coli DH5-α

2-5 µL of ligation mixture or plasmid was added to 50-100 µL of calcium competent cells on ice. The cells were incubated on ice for 30 min, heat shocked at 42˚C for 30-45 sec and incubated on ice for 2 min. 900 µL of SOC media (2% tryptone, 0.5% yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, and 20 mM glucose) was added and incubated at 37˚C with shaking for 1 hour. The cells were pelleted by centrifugation and re-suspended in 200 µL SOC and plated onto LB agar plates containing the required antibiotic.

Preparation and Transformation into *Mix and Go! E. coli* TG1

Preparation of Mix and Go! E. coli TG1

A freshly streaked colony of TG1 *E. coli* was cultured in 5 mL Super Optimal Broth (SOB) (2% tryptone, 0.5% yeast extract, 8.56 mM or 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$ and 10 mM $MgSO_4.7H_2O$) broth at 18˚C with 200 rpm shaking for 12 h. The overnight culture was diluted 1 in 100 into fresh SOB and cultured at 18˚C at 200 rpm shaking until OD_{600} 0.4-0.6 was reached (~3-4 h). In the meantime, the 'wash and competent buffers' were diluted to 1X by adding required amount of dilution buffer and stored at 4˚C. The culture of required OD_{600} was transferred and incubated on ice for 10 min. The cells were then pelleted by centrifugation at 1,600-2,500 x *g* for 10 min at 0-4°C. Supernatant was removed and pellets were gently re-suspended in 5 mL ice-cold 1X wash buffer. This step was repeated 2 times. Finally, the supernatant was completely removed, and pellets were gently re-suspended in 5 mL ice-cold 1X competent buffer. Cells were aliquoted (on ice) into 50-100 µL of the cell suspension into sterile microcentrifuge tubes and stored at -80°C until transformations.

Transformation of E. coli TG1

2-5 µL of ligation mixture or plasmid was added to 10-20 µL of competent TG1 cells (*Magic mix and go cells*) on ice. The cells were incubated on ice for 5-10 min and then directly plated onto the LBA plates supplemented with the required antibiotics.

2.2 Molecular Biology

2.2.1 Genomic DNA Extractions

The genomic DNA (gDNA) of *B. subtilis* was extracted using phenol-chloroform. Pellets from 2 mL of overnight cultures were collected and re-suspended in 500 µL of solution A (Birnboim A: EDTA 10 mM, Tris 25 mM, glucose 50 mM – stored at 4˚C) with 0.5 mg/mL lysozyme and incubated for 10 min at 37˚C. To the suspension, 20 µL of RNase A (20 mg/mL), 20 µL proteinase K (20 mg/mL) and 25 µL of 10% SDS was added and incubated at 37°C for 10 minutes and then at 60°C for further 45 min. 1 volume of Phenol/Chloroform/Isoamyl-alcohol (PCI) was added and vortexed. The aqueous phase was separated by centrifugation for 10 minutes at 16,000 x *g*. The clear upper aqueous phase was transferred into a fresh tube. Extraction of gDNA from the aqueous supernatant with PCI was repeated twice. 0.1 volume of 3M sodium acetate (pH 5.2) and 1 volume of 96% ethanol was added and inverted until the gDNA spooled. The precipitated gDNA was collected by centrifugation at 12,000 x *g* (Sigma 1-14K) for 5 min. The supernatant was discarded and the gDNA pellet was washed with 70% ethanol followed by centrifugation for 2 min. All the ethanol was removed, and the pellet was air-dried. The gDNA was re-suspended in 100-500 µL of nuclease free water with gentle pipetting and stored at -80˚C until further use.

2.2.2 Plasmid Extractions

Plasmids were extracted using the manufacturer's protocol of GenElute™ Plasmid Miniprep Kit (Sigma-Aldrich). Briefly, 2 mL of overnight cultures of *E. coli* were pelleted by microcentrifugation at top speed for 2 min at room temperature (RT). The cells were re-suspended in 200 µl of re-suspension solution and vortexed. To this 200 µL of lysis solution was added and immediately inversed and mixed until the mixture became clear and viscous. The mix was incubated at RT for no more than 5 min. The cell debris was precipitated by adding 350 µL of neutralizing/ binding solution and pelleted by centrifugation at 12,000 x *g* for 10 min. The GenElute Miniprep Binding Columns were prepared by adding 500 µL of the column preparation solution and cleared by centrifugation at 12,000 x *g* for up to 1 min. The clear lysate from the previous step was loaded on the prepared Miniprep Columns and centrifuged for 1 min at 12,000 x *g*. The column was washed with 500 µL of optional wash solution followed by 750 µL of wash solution. The column was spun for 1 min at top speed. The centrifugation was repeated to remove the excess solution. The purified plasmid was eluted with 100 µL of nuclease free water and was either used immediately or stored at -20˚C.

2.2.3 Cloning

2.2.3.1 Construction of FLAG-tagged SpoVG

The *spoVG* gene including its native promoter and terminator was amplified by PCR from the wild-type, 168, *trp*+ gDNA using primers spoVG_FLAG_FP and spoVG_FLAG_RP by Phusion polymerase (NEB) (Table 2.3). The vector ECE263 containing 3X-FLAG was digested using the restriction enzyme *Ngo*MIV (New England BioLabs). The resulting PCR product and linearized vector were purified (Monarch® PCR & DNA Cleanup Kit NEB). The ligation mixture was introduced into *E. coli* DH5-α strain via

transformation and selected on LBA with chloramphenicol. Correct clones were selected and used to PCR amplify *spoVG-FLAG* using the spoVG_pRMC_FP and spoVG_pRMC_RP primers (Table 2.3). The resulting PCR product was cloned into the *Asc*I restriction site of pRMC using Gibson assembly. The ligation mixture was introduced into *E. coli* by transformation and selected on LB ampicillin plates. A correct clone of pRMC-*spoVG-FLAG* was transformed into TG1 strain of *E. coli* and selected on ampicillin. The plasmid was also sequenced using pRMC specific primers to confirm the final construct, as designed. The plasmid was extracted, and the purified pRMC-*spoVG-FLAG* was introduced into *B. subtilis* and selected on tetracycline plates as described in section 2.1.3.1. pRMC integrates by homologous recombination at the *amyE* locus. To test the correct chromosomal integration of pRMC-*spoVG-FLAG*, colonies were tested on LB agar containing 1% starch plates. A loss of a zone of starch clearing when exposed to iodine was used to confirm the correct integration due to the loss of amylase production.

2.2.3.2 Gibson assembly

For Gibson assembly, the 2-3 fragment assembly protocol was followed for as per the manufacturer's guidelines. For each experiment 5 µL of the Gibson Assembly Master Mix (2X) was used instead of the recommended 10 µL volume. The total mix was made up to 10 µL with a 3:1 ratio of the insert: plasmid.

2.2.3.3 Construction of pRMC-spoVG-HTF

The *spoVG* gene with its native promoter and terminator was synthetically synthesized with the His(6X)-TEV cleavage site-FLAG(3X) (HTF) tag on gBlock by IDT. Gibson ends matching to the pRMC linearized with *Asc*I restriction site were added to the *spoVG*-HTF gBlock for synthetic DNA

synthesis. The synthesized gBlock was ligated to the linearized and purified pRMC vector via standard Gibson assembly protocol.

List of plasmids used in this study are listed in Table 2.2.

Table 2.2: Plasmids used in this study

Plasmids	Description	Resistance	Source
pFLAG (ECE263)	Cloning vector with tagging genes with 3X-FLAG. Does not integrate into *B. subtilis*	Chloramphenicol in *E. coli* and Tetracycline in *B. subtilis*	BGSC
pRMC	Cloning vector for the complementation of deletion mutants. Integrates into the *amyE* locus of *B. subtilis*	Ampicillin in *E. coli* and Tetracycline in *B. subtilis*	Ruben AT Mars (Mars *et al.*, 2015)
pRMC-*spoVG*-FLAG	pRMC containing *spoVG-FLAG*	Tetracycline	This study
pRMC-*spoVG*-HTF	pRMC containing *spoVG*-HTF	Tetracycline	This study

Table 2.3: Primers used in this study

FP = forward primer and RP = reverse primer (ELD* = Emma L Denham)

Primer name	Primer description	Primer sequence (5' → 3')	Type	Source
spoVG_FLAG_FP	3X FLAG plasmid to *spoVG*	ataaggaggaactactatggATTTACCTTATGCCCG AAATGAAAG	FP	This study
spoVG_FLAG_RP	*spoVG* to 3X FLAG plasmid	caccatcatgatccttataatcgGGAGAAGCTCCAGC TTCTTC	RP	This study
spoVG_pRMC_FP	pRMC to *spoVG*	tgcgttcacagggttcctggATTTACCTTATCCCCGA AATG	FP	This study
spoVG_pRMC_RP	3X FLAG to pRMC	atatcactagattgggctggCGACCGAGCGCAGCGA GT	RP	This study
Bac27F	Amplifies 16S rRNA	AGAGTTTGGATCMTGGCTCAG	FP	Illumina
Univ1492R	Amplifies 16S rRNA	CGGTTACCTTGTTACGACTT	RP	Illumina
AR2	Barcoded adapter bind to the 5' of RNA	TACACGACGCTCTTCCGAT	FP	Illumina
3Tr3	Barcoded adapter bind to the 3' of RNA	AGATCGGAAGAGCACACGTCTG	RP	Illumina
P0_spoVG	Deletion of *spoVG*	gtgcttgaatctgttacactc	FP	Emma L Denham
P1_spoVG	Deletion of *spoVG*	acgctcttgattaaatccgta	FP	Emma L Denham
P2_spoVG	Deletion of *spoVG*	CGACCTGCAGGCATGCAAGCTacagtagttcacc acctttc	RP	Emma L Denham

P3_spoVG	Deletion of *spoVG*	CGAGCTCGAATTCACTGGCCGTCGcaaggactgctgaaaggg	FP	Emma L Denham
P4_spoVG	Deletion of *spoVG*	catcttccgtaacagttgaac	RP	Emma L Denham
phleo3	Phleomycin deletion cassette	AGCTTGCATGCCTGCAGGTCG	FP	Noirot lab
phleo5	Phleomycin deletion cassette	CGACGGCCAGTGAATTCGAGCTCG	RP	Noirot lab
spoVG_pRMC_comp_FP	Inserts *spoVG* into pRMC	atatcactagattgggctggAAAGTGATTCTGGGAGAG	FP	This study
spoVG_pRMC_comp_RP	Inserts *spoVG* into pRMC	tgcgttcacagggttcctggAAGTTTGAAAAATCAGAATTAAAAAAG	RP	This study
pRMC_seq_F	pRMC sequencing primer	GACCGGCGCTCAGGATCT	FP	Denham lab
pRMC_seq_R	pRMC sequencing primer	GTCGACGTCATATGGATCCC	RP	Denham lab

2.2.3.4 Construction of spoVG_pRMC

The *spoVG* gene including its native promoter and terminator was amplified by PCR from the wild-type, 168, *trp*+ gDNA using primers spoVG_pRMC_comp_FP and spoVG_pRMC_comp_RP by Phusion polymerase (NEB) (Table 2.3). The vector pRMC was linearised by the restriction enzyme *AscI*. The purified PCR product of spoVG_pRMC amplicon was ligated into the linearised and purified pRMC using Gibson assembly protocol. The ligation mixture was introduced into *E. coli* DH5-α by transformation and selected on LB ampicillin plates. A correct clone of spoVG_pRMC was transformed into TG1 strain of *E. coli* and selected on ampicillin. The plasmid was also sequenced using pRMC specific primer set (pRMC_seq_F, pRMC_seq_R) to confirm the final construct (Table 2.3). The confirmed spoVG_pRMC plasmid was extracted and purified before introducing into *B. subtilis.* The positively transformed *B. subtilis* cells with spoVG_pRMC were selected on tetracycline plates as described in section 2.1.3.1. To test the correct chromosomal integration of spoVG_pRMC at *amyE* locus, colonies were tested on LBA-starch (1%) plates. A loss of a zone of starch clearing when exposed to iodine was used to confirm the correct integration due to the loss of amylase production.

2.2.4 Preparation of Protein Lysate

Desired OD_{600} units of cells were pelleted by centrifugation (12,000 x *g*, 10 min, 4˚C) and re-suspended in 1-2 mL of 1X PBS buffer. This was loaded onto FastPrep® Matrix tubes containing 1 mg of lysing matrix B beads (0.1 mm silica spheres) and run on FastPrep®-24 5G homogenizer at 4˚C with the following settings:

Spore CoolPrep

Adapter : CoolPrep

Speed : 6.0 m/s

Bead beating time : 40 s

Pause time : 5 min (tubes left on ice during the pause time)

Cycles : 3 cycles

Pause time : 300 s.

Cell debris was pelleted by centrifugation at 4˚C, 12,000 x *g* (Sigma 1-14K) for 10 min. The cleared lysate was transferred to a centrifuge tube and the spin was repeated to get a clearer lysate. The lysate was stored at -20˚C.

2.2.5 SDS-PAGE and Western Blotting

Protein lysates were combined with 4X Laemmli sample buffer and boiled at 95˚C for 5-10 min. The samples were loaded on a 12% pre-cast SDS-polyacrylamide gel (Bio-Rad) and proteins were separated with 1X SDS-running buffer at 140 V until the loading dye reached the end of the gel. The proteins were transferred to PVDF membrane (Bio-Rad) using a trans-blotter (Bio-Rad) according to manufacturer's instructions. The blotted membrane was stained using Ponceau S (Sigma-Aldrich) to determine transfer efficiency. The membrane was blocked using 3% bovine serum albumin (BSA) (Sigma-Aldrich) in 1X TBS-Tween (0.01%) or 5% skimmed milk (Sigma-Aldrich) TBST for 1 hour at room temperature or at 4˚C overnight with gentle shaking. Following this, the membrane was then washed three times with 1X TBST for 5 min, shaking at RT. Next, the membrane was incubated with 1˚ antibody (monoclonal anti-FLAG M2 antibody (mouse) (Sigma-Aldrich) 1:10,000 and polyclonal anti-TrxA (rabbit) 1:500) for 1 h shaking at RT or overnight at 4˚C. Then, the membrane was washed three times with 1X TBST for 5 min each and incubated with HRP-labelled/chemiluminescent 2˚ antibody (anti-mouse – LI-COR IRDye® 800CW - 1:15,000; anti-rabbit – Cell Signalling – 1:1,000) for 1 h shaking at RT. The membrane was washed three times with 1X TBST for 5 min each, before developing with horseradish peroxidase (Thermo Fisher Scientific) or irradiated by IR740 (filter LY800) and imaged using the Syngene G:BOX.

2.3 Omics Experiments

2.3.1 Transcriptomics with RNAtag-Sequencing

2.3.1.1 Cell culture and sample collection

Three biological replicates each of the WT 168, *trp*+ (BSB1) and Δ*spoVG* strains were grown in M9 and sporulation media as per the per standard laboratory protocols. 20 OD units of cells were harvested for each of the conditions into 0.5 volume of semi-frozen killing buffer (20 mM Tris-HCl pH 7.5, 5 mM $MgCl_2$, 20 mM NaN_3). For M9, cells were harvested at OD_{600} 0.5 (40 mL) and for sporulation conditions, cells were collected at hour 2 (SH2) and hour 5 (SH5) from the same shake flask.

Cells with killing buffer were mixed well (samples on ice or at 4˚C this step) and pelleted by centrifugation at 3,900 x *g* on Eppendorf 5810 R swing bucket rotor for 10 min at 4˚C. The supernatant was decanted off the tubes. The pellets were snap frozen in liquid nitrogen and stored at -80˚C until further use.

2.3.1.2 RNA collection and extraction

The RNA workspaces and hardware were cleaned with RNaseZap™ (Sigma-Aldrich) before RNA extractions were carried out. For all the further steps, nuclease-free molecular grade water (Sigma-Aldrich), RNase-free filter tips (Sarstedt), microfuge tubes (Sarstedt) and conical centrifuge tubes (Falcon) were used. To the thawed bacterial pellets, 600 µL of LETS buffer (0.1 M LiCl, 0.01 M Na_2EDTA, 0.01M Tris-Cl pH 7.4, 0.2% SDS) was added. The pellets were re-suspended gently by pipetting up and down. The suspensions were transferred to Lysing Matrix B tubes (2 mL) with 0.1 mm silica spheres (MP biomedicals™). The cells were bead bet and cleared once as described in section 2.2.4.

1 volume (600 µL) of phenol:choloform (PC) (5:1) pH 5.2 ± 0.2 was added to the cleared cell lysates at RT. The tubes were shaken on a shaking heating block (Eppendorf® Thermomixer® C) at 2,000 rpm for 5 min at RT. The cell debris was pelleted by centrifugation at 12,000 x *g* (Sigma 1-14K) at RT for 5 min. The top layer was transferred to 2 mL tube and 1 volume of PC was added and steps were repeated once more until the clear aqueous upper layer of RNA was retrieved. To the upper layer solution, 1 volume of choloroform:isoamyl-alcohol (24:1) was added and shaken at 1,200 rpm for 5 min at RT. The top layer was transferred to a 1.5 mL tube while the tubes were placed on ice. To this, 0.1 volume 3M sodium acetate pH 5.2 and 1 volume cold isopropanol was added. The tubes were mixed gently by inverting them several times. The samples were left to precipitate overnight at -20˚C. The next day, the precipitated RNA were pelleted by centrifugation at 12,000 x *g* (Sigma 1-14K) at 4˚C for 15 min. The supernatants were removed by aspirating gently and the pellets were air dried for ~15 min at RT. The pellets were re-suspended in 100 µL of RNase-free water, vortexed briefly and incubated for 3 h on ice and then for 30 min at RT. Samples were vortexed briefly after each incubation. 1 µL of SUPERase•In™ (Thermo Fisher Scientific) was added to each of the RNA samples to avoid RNA degradation and were stored at -80˚C until further use.

2.3.1.3 DNA and ribosomal RNA removal

The RNA samples of 18 samples (3 biological replicates each of WT and Δ*spoVG* strains in 3 growth conditions) were diluted 1 in 10 and 2 µL was quantified using Qubit RNA BR Assay Kit (Thermo Fisher Scientific) with Qubit® 2.0 Fluorometer. Genomic DNA contamination was removed by treating 30 µg of RNA samples with TURBO™ DNase (Thermo Fisher Scientific) according to the manufacturer's protocol. 1 µL of SUPERase•In™ (Thermo Fisher Scientific) was added per 50 µL of RNA samples to avoid RNA degradation. Post DNase treatment, RNA samples were cleaned and concentrated using the RNA Clean and concentrator™ -5 kit (ZYMO

RESEARCH) according to the manufacturer's protocol. Removal of genomic DNA was confirmed by PCR using the bacterial forward primer Bac27F and the universal reverse primer Univ1492R against the 16S rRNA gene using GoTaq Green master mix (Promega).

After confirming samples are free of genomic DNA contamination, RNA integrity and residual DNA contamination was measured by running the RNA samples on the Agilent 2100 Bioanalyzer RNA 6000 Pico kit (Agilent, UK) according to the manufacturer's protocol. RNA samples with RNA integrity number (RIN) > 7 were used for the down-stream processing.

RNA of each sample was diluted to a concentration of 26.6 ng/µL and fragmented by heating at 94˚C for 3 min on a thermal cycler (BioRad DNA Engine Tetrad® 2 PCR Thermal Cycler). The residual DNA was treated with DNases and the RNA samples were purified on Agencourt RNAClean XP bead according to the manufacturer's protocol.

The 18 RNA samples were split into two batches of 9 samples each. Each of the 9 RNA samples in a batch were ligated to a unique 8 bp barcoded adapter to the 3' end of the RNA samples. The 9 samples of each batch were pooled together and cleaned using RLT buffer and Zymo Clean & Concentrator™-5 column (200nt cut off).

Further, the ribosomal RNA (rRNA) of each batch was depleted using the 'Experienced user's protocol' of the Ribo-Zero™ kit. The rRNA depleted RNA samples were stored at -20˚C until further steps.

2.3.1.4 cDNA library preparation using RNAtag-Seq

The complementary DNA (cDNA) library preparation was adapted from Shishkin *et al*. (2015). Using the rRNA depleted RNA, the first strand cDNA was synthesized using the AR2 primer (53% GC, 19 bp) at the 5' end and cleaned (Table 2.3). The second strand was synthesized using the 3Tr3 adapter primer (55% GC, 22 bp) at the 3' end (Table 2.3). The excess adapters were cleaned, and PCR enrichment was carried out. The linkers P7 and P5 were ligated to the 3' end and the 5' end of the cDNA

respectively. Unique P7 (P7.1 and P7.4) was used for each pool and amplified for 12 cycles. All the PCR products of the 2 batches were pooled and cleaned using AMPure beads. The cDNA concentrations were measured using Qubit HS DNA kit and fragment size of the library were tested using an Agilent DNA HS chip. The library was diluted to 2 nM.

For the NextSeq run, a pool of 66 samples of 2 nM were loaded together. Of these, 36 were of *B. subtilis* origin that included 18 samples from this study. The rest belonged to *Campylobacter jejuni*. The samples were sequenced on NextSeq 550 system on a 150 cycle High-Output cartridge (2 x 75).

2.3.1.5 Transcriptomics data analysis

The binary base call (BCL) files generated from the NextSeq were converted to fastq files manually according to the i7 index using the software bcl2fastq (Conversion v2.19 User Guide) (Illumina). For demultiplexing the samples, the no-lane-splitting option was chosen before processing the fastq files. The de-multiplexing was carried out using the unique tags/barcodes as a part of Read 1 (R1). The sequences were trimmed to remove the tags and the sequences were pre-processed for DESeq2 as follows.

The reference genome of *B. subtilis* (NC_000964.3) was downloaded from NCBI in gff and fna format. The S number annotations for the non-coding RNAs were obtained from Nicolas *et al.,* (2012). Using bowtie2 v2.3.4.2 the reference genome was built and the R1 was flipped to match the R2 by using Seqtk. Both the reads were mapped to the reference genome using bowtie2 and the output was obtained in a sam file format (--very-sensitive-local --ff). The sam files were converted to bam files, sorted and indexed to the reference genome. Only the reads that mapped to the reference genome were subsequently used and counted using the coverageBed tool (v2.25.0). A modified PrepareDESeq_gff3_locus.py script (written by

Richard Brown) was used to combine the locus tag with its corresponding read counts and the output was saved as a txt file.

The transcriptomic data was analysed using the DESeq2 package in R adapted from Love, Huber and Anders, (2014). The txt file was used as the input file for the DESeq2 analysis. The data was inputted in the 'DESeqDataSetFromMatrix' format. For pre-filtering, 'keep' was set up rowSums(counts(dds)) >= 20, where only the rows that have at least 20 reads total were retained. Comparisons of Δ*spoVG* were made in contrast to the WT using the 'relevel' method. For visualization, log fold shrinkage 'lfcShrink' using the apeglm method was used (Zhu, Ibrahim and Love, 2019). For statistical significance, adjusted *p*-value (padj) < 0.05 was considered as significant. Figures were made using packages in R.

2.3.1.6 Identification of ncRNAs

Putative ncRNAs were identified across all the samples of WT and Δ*spoVG* reads in all the growth conditions using the 'toRNAdo' pipeline as described in Hermansen *et al.*, (2018). The version 2 script was used for Bedtools 2.24.0 and above. The script generated a csv output file that contained the start and stop positions, strand information, and type of putative ncRNA found. The names of the putative ncRNAs were preceded with 'BSU_NCR' followed by the start site position number for each unique transcript. The putative ncRNAs identified were combined with the gff file for *B. subtilis*, normalised and analysed with other transcriptomic reads using DESeq2.

2.3.2 Total Proteomic Analysis

2.3.2.1 Bacterial cultures and sample harvesting

168, *trp*+ (BSB1) strains of the WT and the Δ*spoVG* were grown in biological triplicates according to the standard laboratory conditions in LB, M9 and

sporulation media. 20 OD_{600} units of each of the cell strain was collected for each time point as discussed in the transcriptomic data for the growth conditions M9, SH2 and SH5 (section 2.3.1.1). For growth in LB, cells were harvested at OD_{600} of 1.0.

The samples were collected by centrifugation at 12,000 x *g* for 7 min and washed three times with 5-10 mL cold PBS. The pellets were re-suspended in 200 µL urea buffer (8 M Urea, 50 mM Tris and 75 mM NaCl). The samples were bead beaten as described previously (section 2.2.4) and the lysates stored with the beads in the lysing matrix tubes at -20˚C until further use.

2.3.2.2 Sample preparation protein precipitation and trypsin digestion

To prepare samples for mass-spec analysis, the tubes were thawed at 4˚C and the lysates were cleared by centrifugation at 12,000 x *g* at 4˚C for 1 hour. The cleared lysate (~250 µL) was collected into a clean microfuge tube and quantified by Qubit Protein Assay Kit (Thermo Fisher Scientific) as per manufacturer's protocol. To 150 µg of protein, 50 mM ammonium bicarbonate (ABC) was added to make a final volume of 50 µL. To this, 1 µL of 50X of tris(2-carboxyethyl)phosphine hydrochloride (TCEP) (a reducing agent) and chloroacetamide (CAA) (50mM and 250 mM) was added to get a final TCEP (1 mM) and CAA (5 mM) concentration (Goodman *et al.*, 2018). Tubes were incubated at RT for 30 min in the dark. A trypsin stock solution of 40 µL (0.5 µg/µL) was prepared for the digestion of the protein at RT. Post 30 min incubation, the samples were diluted with 50 mM ABC to make a 1 in 4 dilution (2 M urea final conc.) i.e., for each 51 µL of solution, 150 µL of 50 mM ABC was added to get a final volume of 201 µL and mixed well with pipetting up and down gently. The proteins were digested with trypsin with a 1 in 50 ratios, i.e., 6 µg of trypsin (0.5 µg/µL) was added to 150 µg of proteins. The tubes were mixed well by inverting them and incubated overnight at RT in a dark place.

The following day, the samples were filtered and purified using C18-Stage tip method as described in (Rappsilber, Mann and Ishihama, 2007). The

samples were processed for mass-spectrometric quantification as described in the sections below.

2.3.2.3 Mass spectrometry

NOTE: The following analysis was carried out by Dr. Juan R. Hernandez-Fernaud at the Proteomics Research Technology Platform, University of Warwick, UK.

Reverse phase chromatography was used to separate tryptic peptides prior to mass spectrometric analysis. Two columns were utilized, an Acclaim PepMap μ-precolumn cartridge 300 μm i.d. x 5 mm 5 μm 100 Å and an Acclaim PepMap RSLC 75 μm x 50 cm 2 μm 100 Å (Thermo Scientific). The columns were installed on an Ultimate 3000 RSLCnano system (Dionex). Mobile phase buffer A was composed of 0.1% formic acid in water and mobile phase B 0.1 % formic acid in acetonitrile. Samples were loaded onto the μ-precolumn equilibrated in 2% aqueous acetonitrile containing 0.1% trifluoroacetic acid for 8 min at 10 μL min-1 after which peptides were eluted onto the analytical column at 300 nL min-1 by increasing the mobile phase B concentration from 3% B to 35% over 73 min and then to 80% B over 2 min, followed by a 15 min re-equilibration at 3% B.

Eluting peptides were converted to gas-phase ions by means of electrospray ionization and analysed on a Thermo Orbitrap Fusion (Q-OT-qIT, Thermo Scientific). Survey scans of peptide precursors from 350 to 1500 m/z were performed at 120K resolution (at 200 m/z) with a 4 x 105 ion count target. Tandem MS was performed by isolation at 1.6 Th using the quadrupole, HCD fragmentation with normalized collision energy of 35, and rapid scan MS analysis in the ion trap. The MS2 ion count target was set to 1x104 and the max. injection time was 200 ms. Precursors with charge state 2–7 were selected and sampled for MS2. The dynamic exclusion duration was set to 45 s with a 10 ppm tolerance around the selected precursor and its isotopes. Monoisotopic precursor selection was turned on. The instrument was run in top speed mode with 2 s cycles.

2.3.2.4 Proteomics data analysis

Peptide matching and database search

NOTE: The following search was carried out by Dr. Juan R. Hernandez-Fernaud at the Proteomics Research Technology Platform, University of Warwick, UK.

The MaxQuant engine was used to search the raw data against the *B. subtilis* database (GenBank accession number = AL009126.3, UniProtKB proteome ID = UP000001570) and the common contaminant database from MaxQuant. Peptides were generated from a tryptic digestion with up to two missed cleavages, carbamidomethylation of cysteines as fixed modifications, and oxidation of methionines as variable modifications. Precursor mass tolerance was 10 ppm and product ions were searched at 0.8 Da tolerances.

Data filtering using Perseus

Perseus is generally a follow up step after MaxQuant and helps in the interpretation of the protein quantification data (Tyanova, Temu and Cox, 2016; Tyanova and Cox, 2018). The “proteinGroups.txt” output file containing the protein quantification data was used to perform the downstream processing using Perseus (version 1.5.5.3). The data preparation was adopted from the standard filtering method as described for the Label-free interaction data in the Perseus documentation (Tyanova *et al.*, 2016). Each growth condition data set was analysed individually from the “proteinGroups.txt” file as per Data preparation methods (Tyanova *et al.*, 2016).

2.3.2.5 Differential protein production analysis

The data were processed in RStudio version 3.5.0. The log transformed LFQ intensities were imputated as follows (i) if out of three biological

replicates of a sample, there was one missing value (NA) then the NA was replaced by the average of the two other LFQ intensities, and (ii) if there were more than two NA values, then they were replaced by the minimum LFQ intensity generated from the entire LC-MS/MS data set. The imputated values were fit into a linear model using the lmFit (Linear Model for Series of Arrays) function of the Limma package. The differential production of each protein was quantified by the eBayes (Empirical Bayes Statistics for Differential Expression) function of the same package. A contrast was set by setting the values for WT as reference values compared to the Δ*spoVG* values. To obtain a list of proteins that are most likely to be differentially produced in this comparison, topTable function was used. Volcano plots of differentially produced proteins were made using ggplot2 package.

2.4 RNA-Protein Interactions in *B. subtilis*

NOTE: The following experiments were carried out by Dr. Pedro Arede Rei in the laboratory of Dr. Sander Granneman, University of Edinburgh, UK.

2.4.1 Ultraviolet (UV) Cross-linking and Immunoprecipitation (CLIP)

The WT, 168, *trp*+ and the SpoVG-FLAG tagged strain of (supplemented with tetracycline) *B. subtilis* cells were grown in LB to OD_{600} = 0.5 and zapped with 20 mJ (~18 seconds) of UV in Vari-X-linker. The details of the experiment are described in (Urdaneta *et al.*, 2019).

2.4.2 UV Cross-linking and Analysis of cDNA (CRAC)

Three biological replicates each of the WT, 168, *trp*+ and the SpoVG-HTF tagged strain was used. The protocol for cross-linking and sequencing was adapted as described in Van Nues *et al.,* (2017). In summary, the RNA-protein complexes were cross-linked using the Vari-X-linker. The RNA was

radio-labelled and eluted from the RNA-protein complex as described in the kinetic CRAC (χCRAC) protocol (Van Nues *et al.*, 2017). The RNA-sequencing was not performed.

2.5 RNA-RNA Interactions by CLASH

NOTE: The following experiments were carried out by Dr. Pedro Arede Rei in the laboratory of Dr. Sander Granneman, University of Edinburgh, UK.

Three biological replicates of the WT 168, *trp*+ and SpoVG-HTF tagged strain were used for the cross-linking, ligation, and sequencing of hybrids (CLASH) in *B. subtilis* as described in (Kudla *et al.*, 2011). The data was analysed by Dr. Sander Granneman as described in (Iosub *et al.*, 2020).

2.6 Protein-Protein Crosslinking and Immunoprecipitation (PPI)

2.6.1 Cell Culturing

The protein-crosslinking experiment was adapted from the SPINE protocol for *B. subtilis* (Herzberg *et al.*, 2007). A 4% formaldehyde solution was prepared by dissolving paraformaldehyde (white powder) in 1 X PBS pH 6.5 at 65-70 °C for 20-30 min. Three biological replicates of WT, 168, *trp*+ (BSB1) and SpoVG-HTF tagged strains were grown in 500 mL LB under standard laboratory conditions for 8 hours (~OD_{600} of 2.3) until they reached stationary phase.

2.6.2 Crosslinking with Formaldehyde

The samples were split into two technical replicate of 250 mL each. One replicate each of the WT and the SpoVG-HTF were crossed-linked with 0.3% formaldehyde (final concentration), whereas another technical

replicated was left untreated. Both the cross-linking and non-crosslinking flasks were incubated for 20 min at 37˚C with shaking. The cells were immediately harvested and pelleted by centrifugation at 18,600 x *g* for 15 min at 18℃. The cells were washed once with 20 mL 1X PBS (pH 6.5) at RT and pelleted in a 50 mL falcon by centrifugation at 3,220 x *g* for 20 min at 4˚C. The pellets were stored at -20˚C.

2.6.3 Cell Lysis and Immunoprecipitation

The pelleted cells were resuspended in 1 in 100 volume of the initial culture volume in 1X Tris-HCl pH 7.6, i.e., resuspend the pellet in 2.5 mL (from a culture volume of 250 mL) on ice. The cells were lysed using lysing matrix B containing 0.1mm silica beads. To lyse the cells, bead-beating tubes were prepared by adding 150 µL of lysing matrix B slurry in 1X Tris-HCl pH 7.6. The resuspended cell pellets were split into equal volumes (~833 µL each for every 2500 µL of total cell resuspension) into tubes containing the bead slurry. Cells were bead-beat using a Precellys® Evolution Homogenizer (Bertin, VWR) for two complete cycles with the following settings : Tube: 0.5 mL, Cryolys: Off, Speed: 6,500 rpm, Temperature: 4˚C, Cycle: 3 X 30 s, Mode: Manual, Pause: 30 s. Clear lysates were collected by centrifugation at 5,000 x *g* for 10 min at 4˚C. The lysates were pelleted by centrifugation at 15,000 x *g* for 10 min at 4˚C to clear all the cells debris. For the final clearing step, the lysates were filtered through a 0.45 µm filter and marked as “total input”. 100 µL of the total input was stored at -20˚C for SDS-PAGE and Western blotting. The rest of the total lysate was further processed. To pre-clear the total cell lysates for non-specific protein binding before immunoprecipitation, 60 µL of Dynabeads Protein A magnetic beads (Thermo Fischer Scientific 10001D) were used for every ~2 mL of total input. Samples were mixed well and rotated with beads for one hour at RT. The clear lysates were collected, and beads were discarded. 50 µL of each of this lysate were collected for protein gels and Western blots and saved as “cleared lysate” at -20˚C.

Immunoprecipitation was carried out using ANTI-FLAG® M2 Magnetic Beads (SIGMA-ALDRICH Catalog Number M8823) manufacturer's instructions as a guideline. Briefly, to equilibrate the bead every 1 volume of the beads was re-suspended in 5 volumes of 1X PBS, i.e., every 200 µL of anti-FLAG magnetic beads was re-suspended in 1 mL PBS. The suspension was mixed thoroughly, and the supernatant was discarded. This was repeated two more times. For every ~1.2 mL of protein lysate to be immunoprecipitated, 40 µL of equilibrated anti-FLAG magnetic beads was used. Tubes were rotated at RT for one hour and the protein-protein complexes were immunoprecipitated on the anti-FLAG magnetic beads. After the completion of the IP, supernatant was collected and stored at -20˚C as "IP Supernatant". Each of the 40 µL of anti-FLAG magnetic beads were then washed with 200 µL 1X PBS for three times. The protein-protein complexes were eluted off the beads using 200 µL of 0.1 M Glycine (pH3) with continuous rotation for 20 min at RT. The eluted complexes (~200 µL) were stabilized with 20 µL of 1M Tris-HCl (pH 8) and collected into 1.5 mL microfuge tubes. This was collected as "IP-Glycine" and stored at -20˚C for gels and Western blots. The duplicate of the beads was eluted with 100 µL of 1X – SDS loading sample buffer (4X buffer diluted to 1X with 1X PBS) and stored at -20˚C.

2.6.4 Protein Gels and Western Blotting

To confirm crosslinking, IP and elution of the protein-protein samples, SDS-PAGE gels were run with samples from each of the control and test samples. Samples were heated for 10 min at 97˚C and 15 µL were loaded onto 4–15% Mini-PROTEAN® TGX™ Precast Protein Gels, 15-well, 15 µl (#4561086). Protein gels and Western blots were performed using standard laboratory protocols as described in section 2.2.5. Anti-FLAG (mouse) antibodies were used to detect proteins crosslinked with SpoVG-HTF.

2.6.5 Mass Spectrometry

NOTE: The following experiment was carried out by Dr. Liangcui Chu in the laboratory of Dr. Sander Granneman, University of Edinburgh, UK.

Proteins were denatured with 10 mM Dithiothreitol (DTT) at 55°C for 30 min then mixed with three volumes UA (8 M urea in 100 mM Tris-HCl, pH 8). Denatured proteins were cleaned with filter-aided sample preparation (FASP, Microcon-30kDa Centrifugal Filter Unit, Millipore MRCF0R030) (Wisniewski et al, 2009). After passing the mixture through the FASP column, it was washed with 200 µL UA. Proteins were alkylated with 100 µL 50 mM iodoacetamide (Sigma-Aldrich I6125) at room temperature for 20 min in dark, then washed twice with 100 µL UA and twice with 100 µL ABC (Ammonium bicarbonate, Sigma-Aldrich 09830). MS Grade Trypsin Protease (1 µg per sample, Thermo Scientific 90057) in 39 µL ABC was applied on the membrane and incubated at 37°C overnight. The peptides were collected by centrifugation and the elution was repeated with 40 µL ABC. Peptide concentrations were measured using the NanoDrop (Thermo Scientific ND-2000C). The peptides were acidified with Trifluoroacetic acid (TFA) to a pH<=3 before they were desalted using C18-StageTips (Rappsilber et al, 2003). Two pieces of C18 disks (Empore 2215) were stacked on the tip, and then activated by 15 µL methanol and equilibrated with 50 µL 0.1% TFA. Samples were loaded, passed, and washed with 50 µL 0.1% TFA on the tips.

The tryptic peptides eluted from StageTips (80% Acetonitrile (ACN), 0.1% TFA), lyophilised, re-suspended in 0.1% TFA and analysed on a Fusion Lumos mass spectrometer connected to an Ultimate Ultra3000 chromatography system (both Thermo Scientific, Germany) incorporating an autosampler. 5 µL of the tryptic peptides, for each sample, was loaded on an Aurora column (IonOptiks, Australia, 250 mm length), and separated by an increasing ACN gradient, using a 90-min reverse-phase gradient (from 3%–40% ACN) at a flow rate of 400 nL/min. The mass spectrometer

was operated in positive ion mode with a capillary temperature of 275°C, with a potential of 1300 V applied to the column. Data were acquired with the mass spectrometer operating in automatic data-dependent switching mode, MS 120k resolution in the Orbitrap, MS/MS obtained by HCD fragmentation (36 normalised collision energy), read out in the ion-trap with "rapid" resolution with a cycle-time of 2s. MaxQuant version 1.6.2.10/ 1.6.10.43 (Cox & Mann, 2008) was used for mass spectra analysis and peptide identification via Andromeda search engine (Cox et al, 2011). Match between runs and iBAQ (intensity-based absolute quantification) log fit were chosen. Trypsin was chosen as protease with minimum peptide length 7 and maximum two missed cleavage sites. Carbamidomethyl of cysteine was set as fixed modification, methionine oxidation and protein N-terminal acetylation as variable modifications. First search peptide tolerance was 20 ppm and main search peptide tolerance 4.5. Peptide spectrum match (PSM) was filtered to 1% false discovery rate (FDR).

2.6.6 Data Analysis

The LFQ intensities were analysed for differential analysis as discussed in section 2.3.2.5 except that the LFQ intensities were not transformed and the data was not imputated for the missing values.

2.7 Phenotypic Assays

2.7.1 Swimming and Swarming in *B. subtilis*

The cells were grown in LB supplemented with antibiotics wherever necessary overnight at 37˚C with continuous shaking at 200 rpm. Next day it was diluted 1 in 100 in fresh LB without antibiotics and grown for 3.5 hours. Cells were standardized to OD_{600} 0.1 and 1 µL was spotted on 20 mL M9 with 0.7% agar (swarming) or 0.26% (swimming). The M9 plates were

previously dried for 10 min before spotting and 5 min after spotting. The plates were incubated at 37˚C for 14-16 hours overnight in monolayers. The swarm plates were imaged using Canon G10.

2.7.2 Biofilms Formation on LBGM agar

Freshly streaked colonies of *B. subtilis* were inoculated in LB with appropriate antibiotics overnight. Next day, the overnight culture was diluted to 1/100 in fresh LB and grown for 3.5 hours without antibiotics. The cells were standardized to OD_{600} 0.1 and 1 µL was spotted on 20 mL LB supplemented with glycerol (final concentration of 1% vol./vol.) and manganese (final concentration of 1 mM $MnSO_4$) (LBGM) with 1.5% agar (Shemesh and Chai, 2013). The plates were dried in the laminar airflow hood for 10 min pre-spotting and 5 min after spotting. Plates were incubated at 30˚C in monolayer with face up for up to 5.5 days. The images were captured on Canon G10.

2.7.3 Biofilms Formation on MSgg agar

NOTE: The experiments were carried out by Tetyana Sukhodub in the laboratory of Prof. Nicola Stanley-Wall, University of Dundee, UK.

The cells for biofilms on MSgg agar were prepared as described in (Arnaouteli *et al.*, 2019).

2.7.4 TCA Precipitation

The protocol for trichloroacetic acid (TCA) precipitation of the secreted proteins was adapted from Huppert *et al*., (2014). In summary, *B. subtilis* strains were grown in LB to an OD_{600} of 1.0-1.3. The cells were pelleted by centrifugation at 16,000 x *g* for 2 min at 4˚C and supernatants were collected

in fresh tubes. The supernatants were filtered through 0.45 µm filter. 1 volume of ice-cold 20% TCA was added and mixed well. The supernatants were left to precipitate overnight at 4˚C. The cell pellets were frozen at -20˚C. The following day, the samples of supernatants precipitated with TCA were pelleted by centrifugation at 15,000 x *g* for 20 min at 4˚C to collect the secreted proteins. The liquid was poured off and the precipitated protein pellets were washed with ice-cold acetone twice and air dried for a few minutes. The dried pellets were re-suspended in 20 µL of 1X loading dye and processed for SDS-PAGE.

Additionally, the stored cell pellets were washed in ice-cold (0-4˚C) 1X PBS and pelleted by centrifugation at 15,000 x *g* for 2 min. The cells were lysed by bead-beating with 100 µL 1X PBS with 0.1% SDS and 1% Triton X-100. For beat-beating 30 µL slurry of Lysing Matrix B beads (0.1 mm silica spheres, MP Biomedicals) with 1X PBS (0.1% SDS and 1% Triton X-100) was added to the cell pellets. The beat-beating was performed on Precellys® Evolution Homogenizer. The supernatant was collected, and the spin was repeated. 18 µL of the pelleted protein were re-suspended with 1X loading dye. All the samples for SDS-PAGE were boiled for 15 min at 97˚C prior to loading on the gel. Post SDS-PAGE, the gels were stained with InstantBlue® (Expedeon) Coomassie stain overnight at RT with constant shaking. The following day the gels were washed with deionized water and imaged.

Chapter 3 Transcriptomic Analysis of Δ*spoVG*

3.1 Introduction

Post-transcriptional gene regulation mediated by small regulatory RNAs (sRNAs) in bacteria is an emerging field of RNA biology. Hfq is a well-known RNA-binding protein (RBP) and is known as a global regulator of post-transcriptional gene regulators that mediates interactions between sRNA-mRNA in *Escherichia coli* (reviewed in (Santiago-Frangos and Woodson, 2018)). Investigation of Hfq in the Gram-positive model organism *Bacillus subtilis* identified Hfq as an RBP but it was not recognised as a global regulator of sRNA-mRNA interactions (Dambach, Irnov and Winkler, 2013; Hämmerle *et al.*, 2014; Rochat *et al.*, 2015).

In *Listeria monocytogenes* and *Borrelia burgdorferi* SpoVG has been identified as a novel RNA-binding protein (Burke and Portnoy, 2016; Savage *et al.*, 2018). SpoVG is also known as a DNA-binding protein and gene regulator in the human pathogen *Staphylococcus aureus* and in *B. burgdorferi* (Jutras *et al.*, 2013; Liu, Zhang and Sun, 2016). In *S. aureus,* SpoVG is known to be involved in the resistance against methicillin and glycopeptides, and in the production of extracellular polysaccharides (Matsuno and Sonenshein, 1999; Schulthess *et al.*, 2009, 2011). Recently, SpoVG was also identified as an important protein for the sporulation in *Bacillus anthracis* (Chen *et al.*, 2020).

spoVG is a highly conserved gene (Gupta, Bhandari and Naushad, 2012), and was first identified in *B. subtilis*. Its protein product (SpoVG) is a negative regulator of asymmetric septum formation in the stage V of sporulation (Matsuno and Sonenshein, 1999). The expression of *spoVG* is induced at the early stages of sporulation and SpoVG is known to participate in the later stages of sporulation (Rosenbluh *et al.*, 1981; Sandman, Losick and Youngman, 1987). This earlier activation of *spoVG* expression (akin to the early expression of other sporulation gene *spoVC* (Piggot and Coote,

1976)), earlier than the protein product (SpoVG) function has been speculated as a result of either due to the developmental regulation of translation of *spoVG* or due to the early synthesis of protein before its function (Rosenbluh *et al.*, 1981). *spoVG* also participates in sporulation in *B. subtilis* synergistically with a sporulation gene *spoIIB* (Margolis, Driks and Losick, 1993). The double mutant of Δ*spoIIB*Δ*spoVG* is known to strongly block sporulation at stage II of sporulation (Margolis, Driks and Losick, 1993). The presence of a third mutation in the sporulation gene *spoVS* along with Δ*spoIIB*Δ*spoVG* allows the blocked sporulation to advance to a late stage (Resnekov, Driks and Losick, 1995). This highlights the complexity of the regulation of genes in sporulation. Little is known about the role of *spoVG* in *B. subtilis* beyond sporulation. The growing body of evidence of the regulatory roles of SpoVG and its RNA-binding properties in other bacteria suggest a potential global regulatory function of *spoVG* in *B. subtilis* beyond sporulation.

Therefore, to investigate the roles and functions of *spoVG* a global transcriptomic profiling experiment using RNAtag sequencing (RNAtag-Seq) was designed (Shishkin *et al.*, 2015). Here, the Δ*spoVG* strain was tested in minimal media supplemented with glucose (M9) and sporulation conditions; sporulation stage II (sporulation hour 2 = SH2) and sporulation stage V (sporulation hour 5 = SH5). The output of the sequencing data was used to identify the biologically significantly enriched gene sets and biological pathways perturbed in Δ*spoVG* which will be discussed in this chapter. The chapter also explored the effects of the deletion of *spoVG* in the sigma factor regulons to identify its novel regulatory roles in *B. subtilis*.

3.2 Results and Discussion

3.2.1 Growth of Δ*spoVG* in M9

The growth of a gene deletion mutant for *spoVG* by replacement with a phleomycin antibiotic resistance cassette (construction described in

Chapter 2, section 2.1.2.1) was compared to the WT strain in minimal media supplemented with glucose (M9). A strain where *spoVG* was complemented (Δ*spoVG*::*spoVG*) at the *amyE* locus (construction described in Chapter 2, section 2.1.2.2) was also included. The Δ*spoVG* strain showed a very similar growth pattern to the WT (Figure 3.1). A slight increase in growth was present in the Δ*spoVG* and Δ*spoVG*::*spoVG* strains. The complementation strain was able to restore other phenotypes as described in Chapter 7. This observation suggested a possible mild polar effect on the downstream gene during the M9 growth condition.

3.2.2 Cell Culture and Sample Preparation for RNAtag-Seq

To compare the expression profiles of the WT and mutant, three biological replicates of the WT and Δ*spoVG* strains were grown in M9 until mid-exponential growth and in sporulation media till sporulation hour 2 (SH2) and sporulation hour 5 (SH5) (Section 2.3.1.1). RNA extraction and preparation of cDNA libraries for RNAtag-Seq was carried out as described in Section 2.3.1.2 to Section 2.3.1.43.2.. The cDNA libraries were paired-end sequenced on a NextSeq 550 using a high-output cartridge (150 cycles – 2 x 75).

3.2.3 RNAtag-Seq Pre-Data Processing and Quality Check

In this experiment, RNA samples were tagged using the barcodes from the original RNAtag-Seq study (Shishkin *et al.*, 2015). In total, 18 cDNA libraries were sequenced on the NextSeq platform along with cDNA libraries from different experiments, which comprised of 30 samples of *Campylobacter jejuni* and 18 samples of *B. subtilis.*

The reads from the samples were aligned to the *B. subtilis* 168 reference genome (NC_000964.3) and the percentage of alignment was calculated. The numbers of mapped reads ranged between 0.93 to 6.3 million reads

per sample with alignment rates ranging from 95-99% (Table 3.1). Since, all the replicates had a coverage rate >94%, the data was considered to be of good quality.

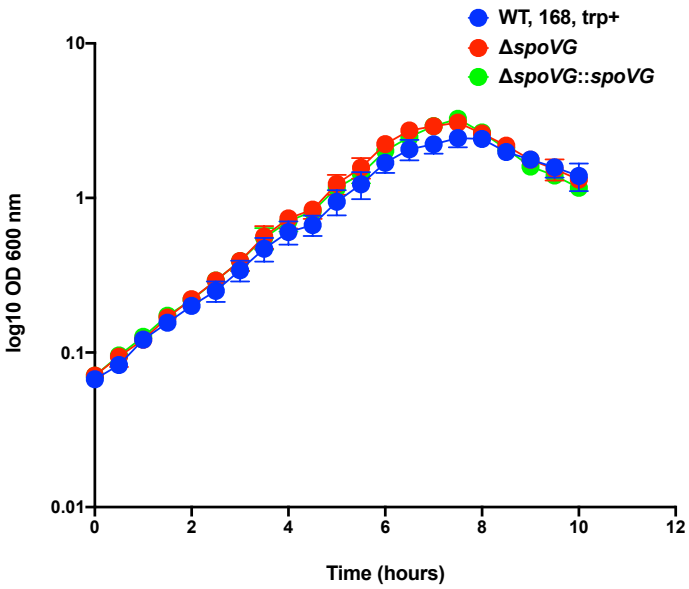


Figure 3.1: Δ*spoVG* shows similar growth to the WT in M9 supplemented with glucose

Growth of WT (blue line), Δ*spoVG* (red line) and the complemented strain Δ*spoVG*::*spoVG* (green line) in shake flasks. Each line on the growth curve represents the average of 5 biological replicates (4 biological replicates for Δ*spoVG*::*spoVG*) ± standard errors of the means (SEM).

3.2.3.1 Data quality analysis

For the differential gene expression analyses, the read counts from Δ*spoVG* were compared to the WT in each growth condition using the DESeq2 package (Love, Huber and Anders, 2014). The DEGs identified with adjusted *p*-value (padj) less than 0.05 irrespective of the $\log_2$ fold change (LFC) values were considered statistically significant. Figure 3.2 outlines the major steps adopted in the RNAtag-Seq data analysis.

Table 3.1: Depth and coverage outputs from RNAtag-Seq

M9 - M9 minimal media, SH2 - sporulation hour 2 and SH5 - sporulation hour 5. W - WT, DEL - Δ*spoVG*. Biological replicate numbers of sample are shown from 1-3.

Samples	NextSeq Outputs	
Condition_strain_replicate	Percentage of total reads aligned to *B. subtilis* NC_000964.3 (%)	Number of reads mapped
M9_DEL_1	98.40	4,985,717
M9_DEL_2	98.03	3,107,568
M9_DEL_3	98.62	3,906,328
M9_W_1	98.62	6,243,659
M9_W_2	98.57	4,018,904
M9_W_3	98.57	3,162,411
SH2_DEL_1	96.64	3,223,257
SH2_DEL_2	97.88	3,651,905
SH2_DEL_3	97.99	4,227,724
SH2_W_1	94.86	1,504,197
SH2_W_2	97.29	1,479,861
SH2_W_3	97.03	932,054
SH5_DEL_1	95.36	4,184,141
SH5_DEL_2	97.86	5,609,482
SH5_DEL_3	96.8	2,769,349
SH5_W_1	98.57	6,119,117
SH5_W_2	96.47	2,856,831
SH5_W_3	97.71	3,831,957

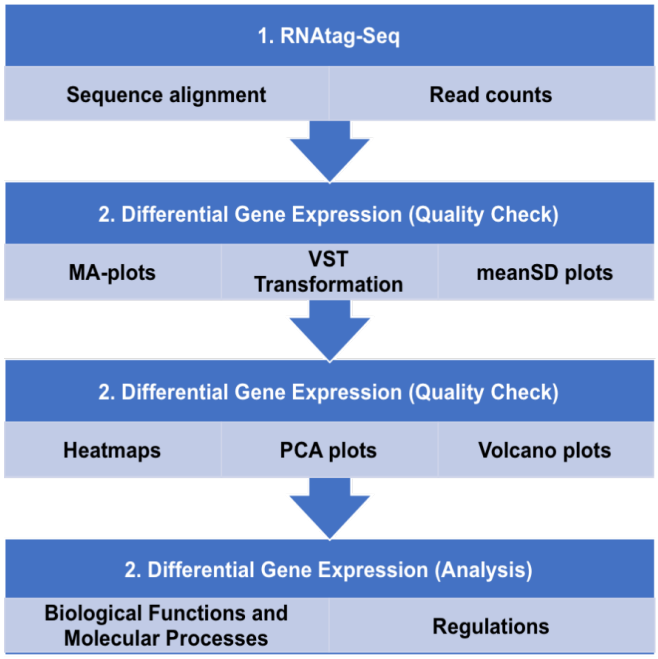


Figure 3.2: Schematic of the steps taken in the differential gene expression analysis of Δ*spoVG*.

The flow chart shows the major steps taken in the differential gene expression analysis of the RNAtag-Seq data analysis.

To visualise the changes in the gene expression in the absence of *spoVG*, MA-plots were constructed. MA-plots are commonly used to visualize the $\log_2$ fold change values of a given gene versus the mean of normalised counts of all the genes in the input data. The MA-plot analysis showed that several genes were significantly differentially expressed (padj < 0.05) in Δ*spoVG* in each of the growth conditions tested (Figure 3.3). The visualisation of the plots showed M9 to have the greatest number of differentially expressed genes compared to the sporulation conditions (Figure 3.3). Additionally, more DEGs were observed in SH5 compared to SH2 among the sporulation conditions. These observations confirmed the changes in the transcriptome of Δ*spoVG* in M9, SH2 and SH5.

3.2.3.2 PCA clustering

A principal component analysis (PCA) of the differential gene expression identified three distinct clusters corresponding to the different growth media tested (Figure 3.4). The tight grouping of samples in each cluster showed that the transcriptome of the samples in each growth condition behaved similarly to a large extent and that the experiment was successful in the media tested. Results from the individual PCA identified two distinct clustered in each growth condition (Figure 3.5). Examination of the principal component 1 (PC1) of each condition showed that major source of variation in these clustering was derived due to the transcriptomic variations between Δ*spoVG* and WT (Figure 3.5). This confirmed that the deletion of *spoVG* indeed affects the gene expression profile (transcriptome) of the cells in each of the M9, SH2 and SH5 conditions. Whereas the variation arising among the biological replicates in principal component 2 (PC2) could be attributed to the batch effect incorporated during sample preparation.

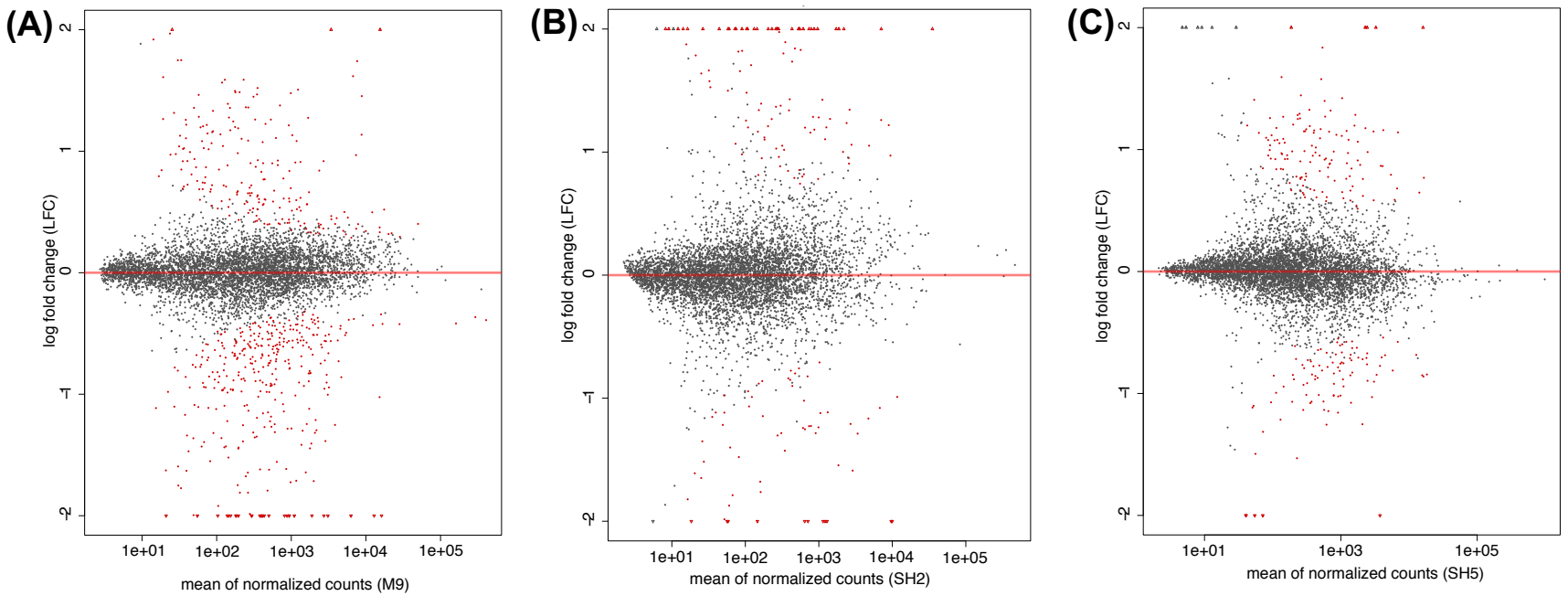


Figure 3.3: Deletion of *spoVG* results in differential expression of genes

The MA-plots of differentially expressed genes in Δ*spoVG* vs WT in **(A)** M9, **(B)** SH2 and **(C)** SH5 growth conditions. The red dots on the plots represent statistically significantly DEGs (padj < 0.05). The points that fall out of the plot window are shown as open red triangles pointing up or down depending on the directionality of the DEGs.

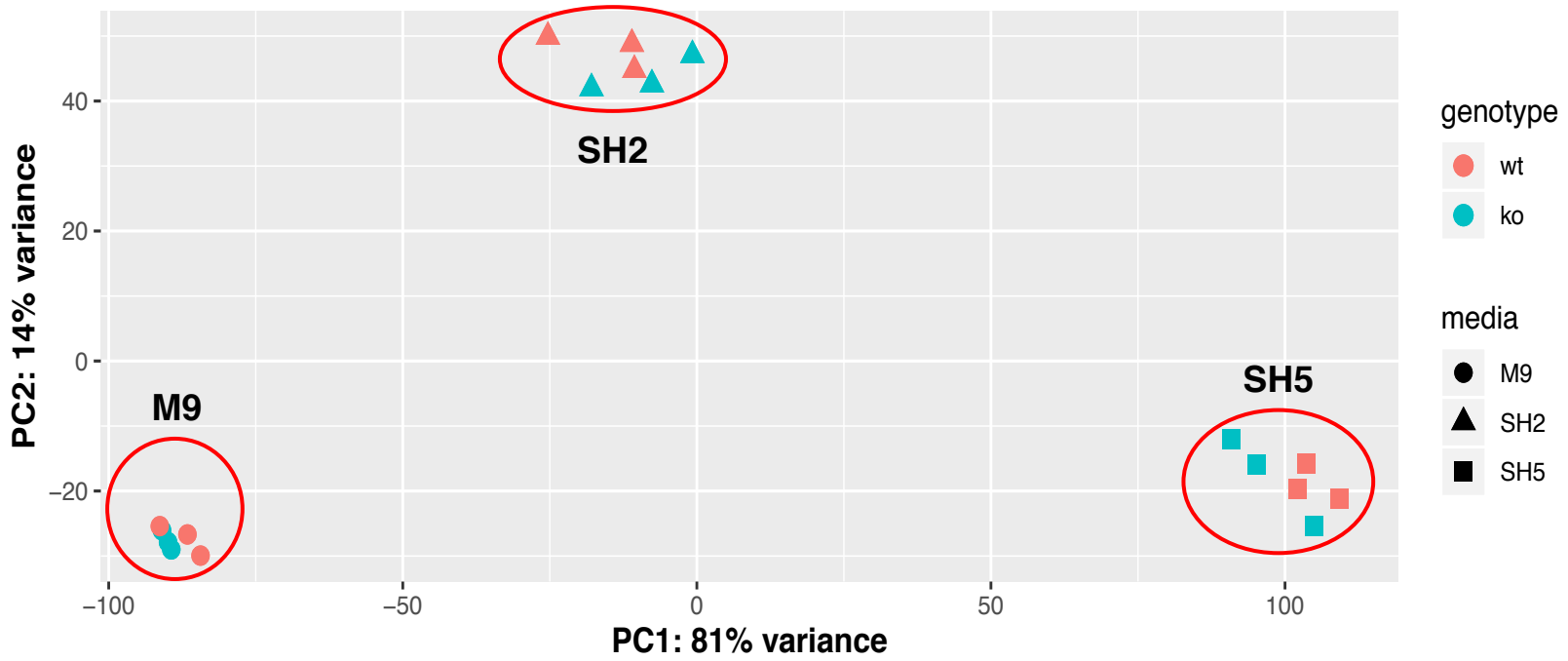
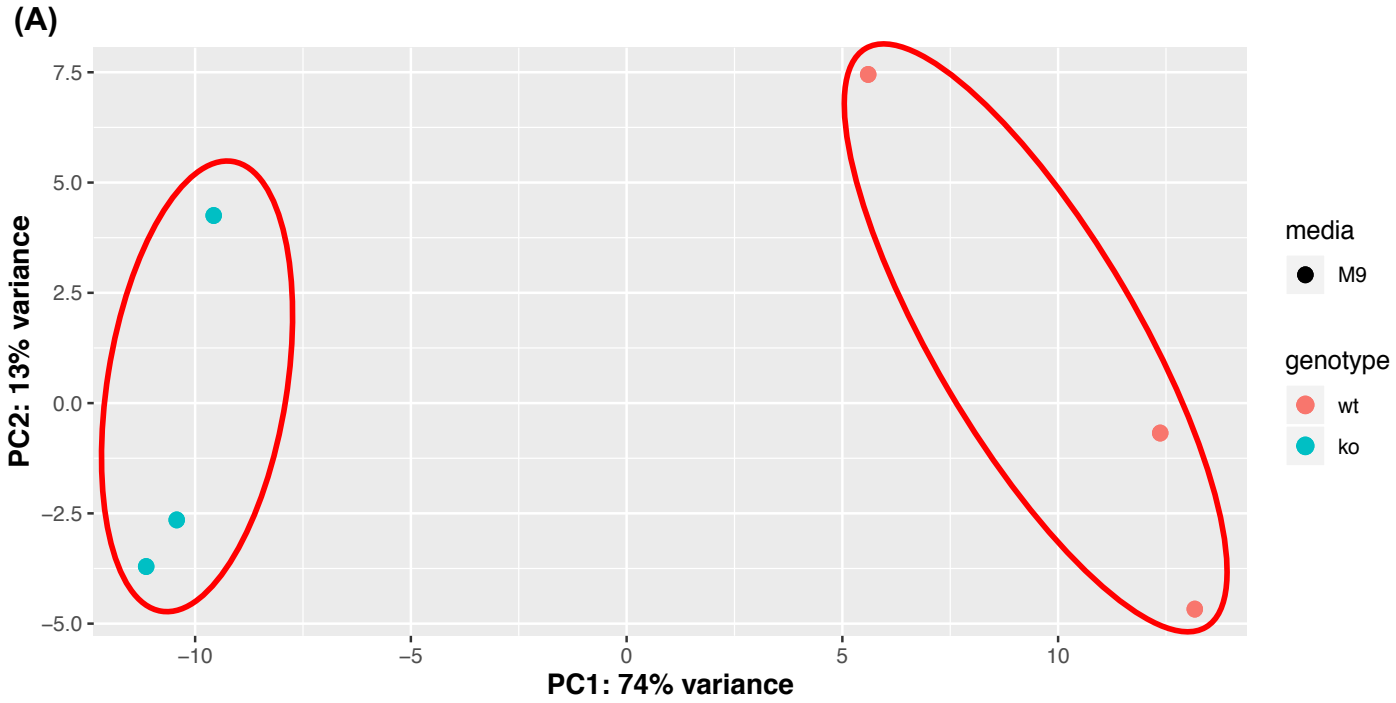


Figure 3.4: The biological replicates of the WT and Δ*spoVG* cluster differently based on the type of growth media used

PCA plot of the biological replicates of the WT (genotype = wt; coloured in red) and the *spoVG* knockout mutants (genotype = ko; coloured in green) growth in 3 different growth media (M9 in circles; SH2 in triangles and SH5 in squares). Each point on the PCA plot represents the transformed count values of all the genes (including protein-coding and non-coding RNAs) identified in the study. The samples grown in different media are shown by different shapes; the different genotypes are shown by different colours.


(A)
7.5
5.0
2.5
0.0
−2.5
−5.0
PC2: 13% variance
−10
−5
0
5
10
PC1: 74% variance
media
M9
genotype
wt
ko

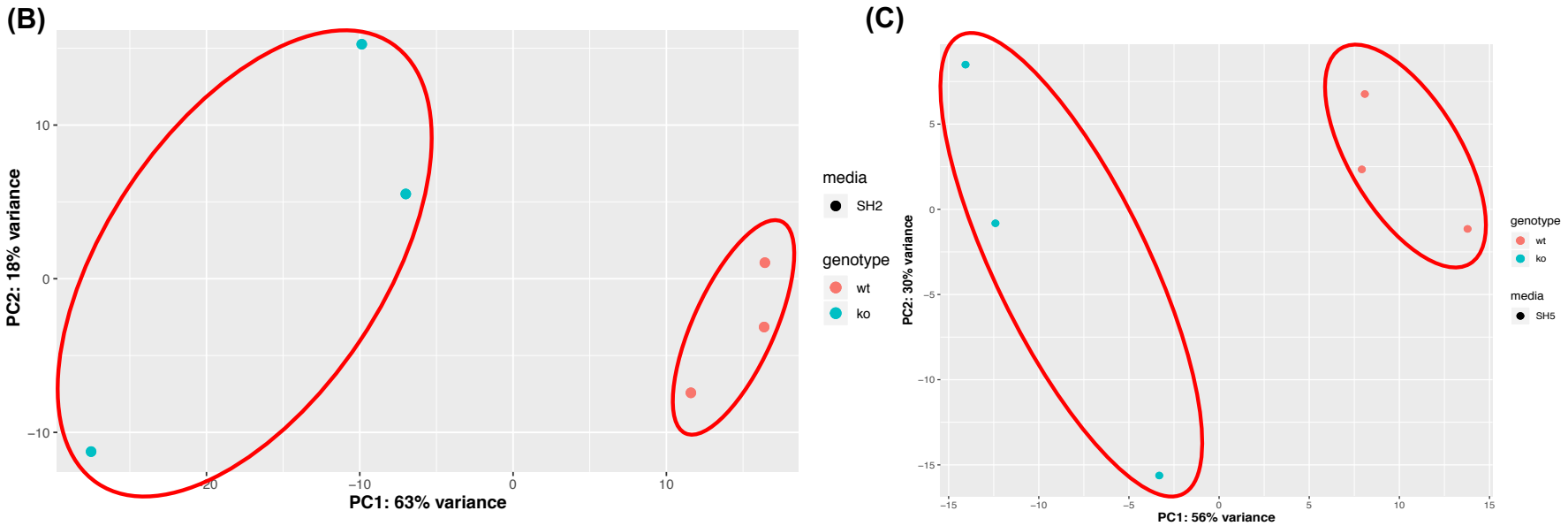


Figure 3.5: The replicates of WT and Δ*spoVG* transcripts cluster separately based on their genotypes in each growth media

PCA plots represent clustering of WT (genotype = wt; coloured in red) and the *spoVG* knockout mutants (genotype = ko; coloured in green) based on the growth media **(A)** M9 **(B)** SH2 **(C)** SH5. Each point on the PCA plot represents the VST transformed count values of all the genes (including protein-coding and non-coding RNAs) identified in the study.

Among the growth conditions, the M9 condition generated the maximum variance of 87% (sum of PC1 and PC2) as a result of *spoVG* deletion (Figure 3.5 (A)). Whereas the SH5 condition generated 86% variance with PC1 and PC2 (Figure 3.5 (C)). The samples of SH2 showed the least variance of 81% due to the *spoVG* deletion (Figure 3.5 (B)). This result suggested that the effect of *spoVG* deletion was maximal in the M9 condition compared to the two sporulation conditions. Also, amongst the sporulation conditions the effect of *spoVG* deletion was larger in SH5 compared to the SH2.

3.2.4 Identification of Putative ncRNAs

To identify any putative ncRNAs arising in the study, the toRNAdo pipeline was used (Chapter 2, section 2.3.1.6). The identified ncRNAs were intergraded into the input files of the DESeq2 analysis. From the differential expression analysis 32, 7 and 19 significantly DEGs were identified in the Δ*spoVG* strain in the M9, SH2 and SH5 conditions respectively. Table 3.2 summarises the differentially expressed putative ncRNAs identified in this study.

Table 3.2: Summary of differentially expressed putative non-coding RNAs (ncRNAs) identified in different growth conditions

Putative ncRNAs identified	M9	SH2	SH5
Total no. of putative ncRNAs	32	7	19
Total no. of up-regulated putative ncRNAs	10	2	6
Total no. of down-regulated putative ncRNAs	22	5	13

3.2.5 Analysis of the Differentially Expressed Genes

To identify the genes differentially expressed in Δ*spoVG*, the ncRNAs identified in this study (section 3.2.4) were also incorporated in the analysis. The following section discusses the cumulative output of the protein-coding and the ncRNAs identified in this study. The distributions of the identified genes were visualised on a Volcano plot where the -$\log_{10}$(adjusted *p*-value) values were plotted against respective LFC values for each gene. The patterns showed that Δ*spoVG* resulted in up-regulated and down-regulated genes in each of the growth conditions tested (Figure 3.7).

From the differential expression analysis 444, 161 and 227 significantly DEGs were identified in the Δ*spoVG* strain in the M9, SH2 and SH5 conditions respectively (Figure 3.6). The largest number of significantly DEGs were identified in the M9 condition. Between the SH2 and SH5 condition, the SH5 condition expressed more significantly DEGs. The details of the significantly DEGs in each condition will be described in the following sections.

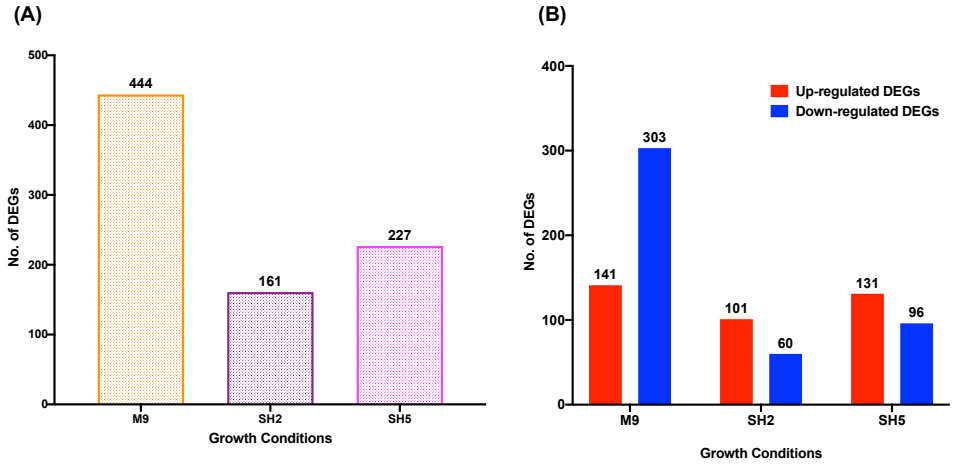
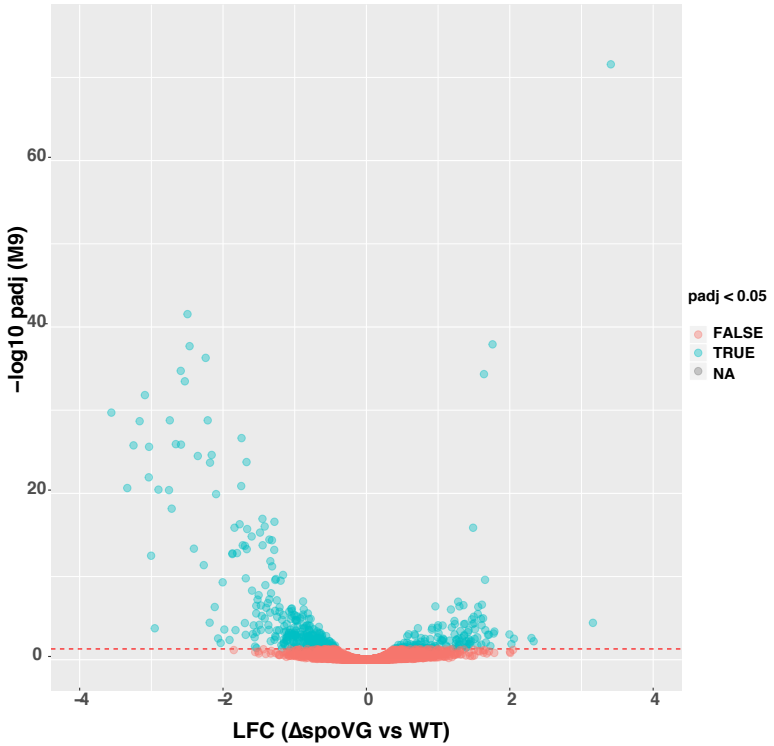


Figure 3.6: Deletion of *spoVG* produced statistically significantly DEGs in each growth media

(A) The cumulative number of significant DEGs (padj < 0.05) including all the coding and non-coding RNAs and the putative non-coding RNAs in the M9, SH2 and SH5 conditions **(B)** The directionality of the significantly DEGs in each growth condition indicated. The red bars represent the up-regulated genes, and the blue bars represent the down-regulated genes.

(A)


60
40
20
0
−log10 padj (M9)
−4
−2
0
2
4
LFC (ΔspoVG vs WT)
padj < 0.05
FALSE
TRUE
NA

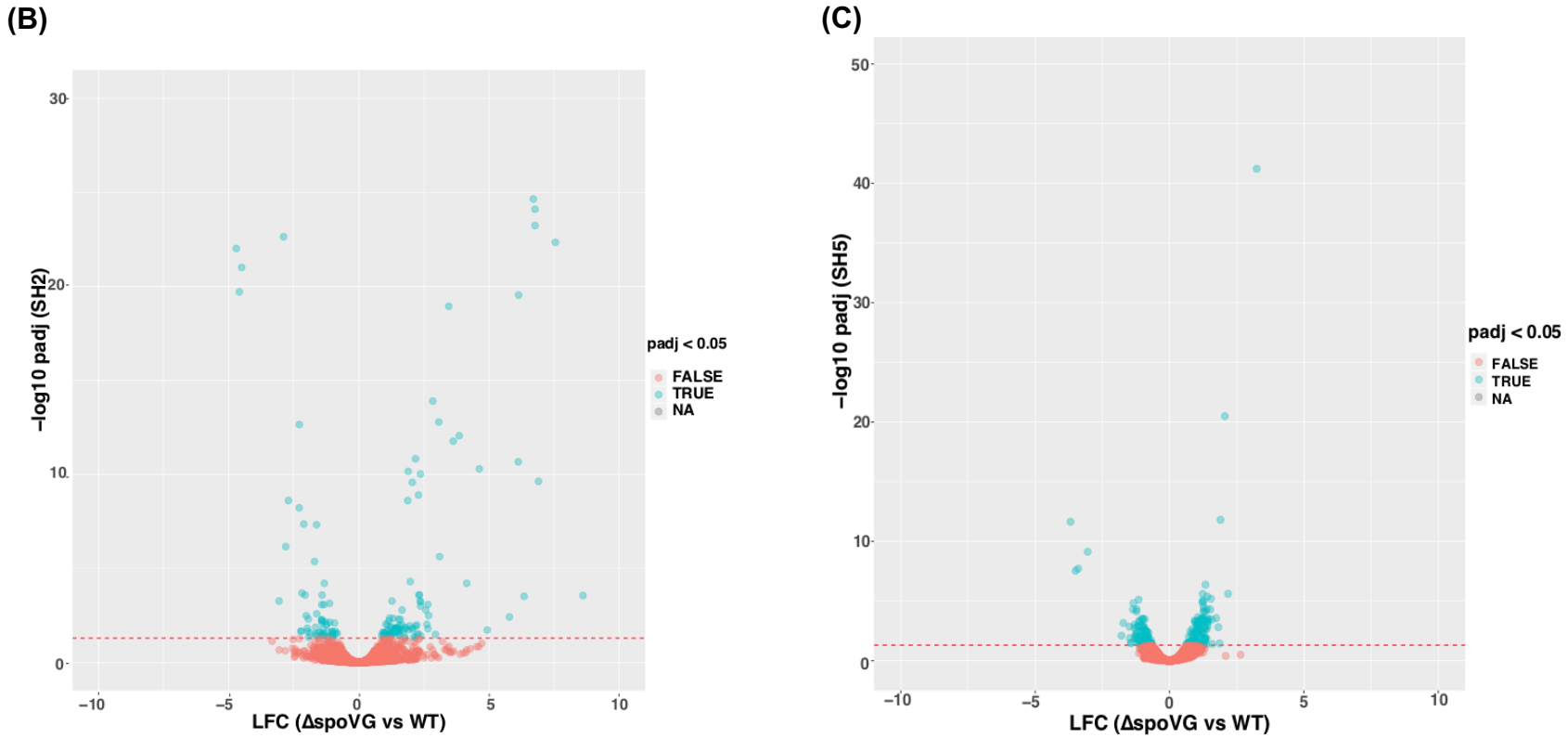


Figure 3.7: Volcano plots of show the up-regulated and down-regulated genes in Δ*spoVG* in each of the (A) M9, (B) SH2 and (C) SH5

3.2.5.1 DEGs in M9

Among the 444 DEGs in M9, 141 were up-regulated and 303 were down-regulated in Δ*spoVG* (Figure 3.6 (B)). The number of DEGs identified in M9 were the highest amongst all the growth conditions studied. Among the most increased DEGs in M9 was the ncRNA S1427 which is flanked upstream by *upp* and downstream by sRNA S1426 (Table 3.3 (A)). The *upp* gene codes for uracil phosphoribosyltransferase which is important for the uridine monophosphate (UMP) synthesis (Presecan *et al.*, 1997). S1426 forms the 5'UTR of *atpI* which encodes for the ATP synthase (subunit I) (Presecan *et al.*, 1997). Table 3.3 lists the top 10 up-regulated and down-regulated genes in Δ*spoVG* in M9.

In the M9 dataset there were 303 significantly down-regulated DEGs, of which 210 were protein-coding genes, 71 ncRNAs and 22 putative new ncRNAs. The putative new ncRNAs were identified according to as described in (Hermansen *et al.*, 2018). In brief, a transcript was characterised a novel putative ncRNA if it fulfilled the following hallmarks – (i) the normalised base-pair expression value was above a defined expression cut-off; (ii) the transcript showed a distinctive 'peak' with a minimum of 5:1 expression ratio between its highest to lowest points of the 'peak'; (iii) the transcript length was greater than 50 bp; (iv) the same transcript was identified across all the replicates for each growth condition (Hermansen *et al.*, 2018). Among the functional genes, the most significantly down-regulated DEG was the protein-coding gene *pksI* (Table 3.3 (B)). The gene *pksI* belongs to the *pks* operon that encodes for a megacomplex responsible for the biosynthesis of bacillaene; a linear non-ribosomal peptide/polyketide (NRP/PK) with antimicrobial properties (Butcher *et al.*, 2007; Straight *et al.*, 2007; Nagler *et al.*, 2016). Many of the other genes in the *pks* operon were also down-regulated in Δ*spoVG* in M9.

Table 3.3: Significantly DEGs in Δ*spoVG* in M9

Top 10 up-regulated **(A)** and down-regulated **(B)** DEGs in *ΔspoVG* in M9 are arranged in a decreasing and ascending order of LFC values respectively.

(A)

Locus Tag	Gene	Product	Description	Function	LFC	padj
new_3788276_3788397_c	S1427	intergenic region	new RNA feature, intergenic region	NA	3.410	2.66E-72
BSU20350	*yorK*	single-strand DNA-specific exonuclease	single-strand DNA-specific exonuclease	unknown	3.160	3.66E-05
BSU21290	*yomN*	unknown	unknown	unknown	2.333	6.08E-03
BSU21300	*yomM*	unknown	unknown	unknown	2.303	2.45E-03
BSU24740	*corA*	unknown	general stress protein, similar to magnesium transporter	unknown	2.058	2.91E-03
new_2246149_2246571	S804	3'UTR of *yomM*	new RNA feature, 3'UTR of *yomM*	NA	2.023	1.28E-02
BSU21190	*yomX*	unknown	unknown	unknown	1.995	9.06E-04
new_1116857_1117712_c	S365	intergenic region	new RNA feature, intergenic region	NA	1.786	4.04E-04
BSUB_NCR1117109	NA	NA	NA	NA	1.775	6.98E-04
BSU00500	*gcaD*	bifunctional N-acetylglucosamine-1-phosphate	bifunctional N-acetylglucosamine-1-phosphate uridyltransferase/glucosamine-1-phosphate acetyltransferase	cell wall metabolism	1.760	1.22E-38

(B)

Locus Tag	Gene	Product	Description	Function	LFC	padj
BSUB_NCR2273724	NA	NA	NA	NA	-3.558	1.95E-30
BSU17170	*pksI*	unknown	involved in polyketide synthesis	polyketide synthesis	-3.336	2.28E-21
new_2273515_2273748	*bsrG*	hydrophobic class I toxin	hydrophobic toxin; interferes with cell envelope biosynthesis, causes membrane invaginations together with delocalization of the cell wall synthesis machinery, triggers autolysis, and inhibits protein biosynthesis	stability of prophage SPβ	-3.249	1.70E-26
BSU17160	*pksH*	unknown	involved in polyketide synthesis	polyketide synthesis	-3.163	2.11E-29
BSU17140	*pksF*	unknown	involved in polyketide synthesis	polyketide synthesis	-3.090	1.53E-32
BSUB_NCR2069983	*NA*	NA	NA	NA	-3.036	1.18E-22
BSU17150	*pksG*	hydroxymethylglutaryl CoA synthase	involved in polyketide synthesis	polyketide synthesis	-3.032	2.53E-26
BSU17130	*acpK*	acyl carrier protein	acyl carrier protein	polyketide biosynthesis	-3.004	3.11E-13
new_820648_820823	S254	independent transcript	new RNA feature, independent transcript	osmoprotection	-2.952	1.65E-04
new_2069824_2069977	SR5	antisense RNA to *bsrE*	antisense RNA, controls expression of the BsrE toxin	Control of BsrE toxin expression	-2.901	3.55E-21

3.2.5.2 DEGs in SH2

In SH2 161 DEGs were affected in response to Δ*spoVG*, of which 101 were up-regulated and 60 were down-regulated (Figure 3.6 (B)). The most up-regulated gene in SH2 was *pyrC* with LFC = 10.83 (Table 3.4 (A)). *pyrC* is a member of the *pyr* operon in *B. subtilis* that encodes for PyrC (a dihydroorotase) (Liang *et al.*, 2005). PyrC is crucial for the *de novo* biosynthesis of UMP, is an integral membrane protein and a regulator of protein PyrR (Turner, Lu and Switzer, 1994; Hahne *et al.*, 2008). In addition to *pyrC* many other genes of this operon were also up-regulated in Δ*spoVG* in SH2. S811 was the most down-regulated gene in SH2 in Δ*spoVG* (Table 3.4 (B)). S811 is transcribed as an independent transcript and appears to be a 3' untranslated region (UTR) of *bsrG* (Zhu and Stülke, 2018). *bsrG* encodes for a hydrophobic toxin that forms a part of the type I toxin-antitoxin system with ncRNA SR4 (Jahn *et al.*, 2012). S811 is also antisense to the *yokL* and *asrG* genes.

3.2.5.3 DEGs in SH5

The deletion of *spoVG* in SH5 resulted in 227 significantly DEGs of which 131 were up-regulated and 96 were down-regulated (Figure 3.6 (B)). Table 3.5 lists the top 10 up-regulated and down-regulated genes identified in SH2 in Δ*spoVG*. In SH5, S1427 was the most up-regulated gene Table 3.5 ((A)). Whereas, SR5 was the most down-regulated gene in SH5. SR5 is a small regulatory RNA (sRNA) that acts antisense to *bsrE* that produces the BsrE toxin (Meißner, Jahn and Brantl, 2016).

Table 3.4: Top 10 significantly DEGs in Δ*spoVG* in SH2

Top 10 up-regulated **(A)** and down-regulated **(B)** DEGs in Δ*spoVG* in SH2 arranged in a decreasing and ascending order of LFC values respectively.

NA = not available

(A)

Locus Tag	Gene	Product	Description	Function	LFC	padj
BSU15500	*pyrC*	dihydroorotase	dihydroorotase	pyrimidine biosynthesis	10.834	1.63E-08
new_1630017_1630089	S568	intergenic region	new RNA feature, intergenic region	NA	8.611	2.73E-04
new_3788276_3788397_c	S1427	intergenic region	new RNA feature, intergenic region	NA	8.517	9.26E-134
BSU15560	*pyrE*	orotate phosphoribosyltransferase	orotate phosphoribosyltransferase	pyrimidine biosynthesis	7.551	4.60E-23
BSU15520	*pyrAB*	carbamoyl-phosphate synthetase	carbamoyl-phosphate synthetase	pyrimidine biosynthesis	6.905	2.32E-10
BSU15540	*pyrD*	dihydroorotic acid dehydrogenase	dihydroorotic acid dehydrogenase	pyrimidine biosynthesis	6.773	8.03E-25
BSU15550	*pyrF*	orotidine-5-phosphate decarboxylase	orotidine 5-phosphate decarboxylase	pyrimidine biosynthesis	6.772	5.88E-24
BSU15530	*pyrK*	dihydroorotic acid dehydrogenase	dihydroorotic acid dehydrogenase	pyrimidine biosynthesis	6.704	2.30E-25
new_3535821_3536011	S1314	5'UTR of *sacB*	new RNA feature, 5'UTR of *sacB*		6.351	2.98E-04
BSU15510	*pyrAA*	carbamoyl-phosphate synthetase	carbamoyl-phosphate synthetase	pyrimidine biosynthesis	6.138	2.93E-20

(B)

Locus Tag	Gene	Product	Description	Function	LFC	padj
BSUB_NCR2273724	NA	NA	NA	NA	-4.704	9.83E-23
new_2273801_2274302	S811	independent transcript	new RNA feature, independent transcript	NA	-4.589	1.98E-20
new_2273515_2273748	*bsrG*	hydrophobic class I toxin	hydrophobic toxin; interferes with cell envelope biosynthesis, causes membrane invaginations together with delocalization of the cell wall synthesis machinery, triggers autolysis, and inhibits protein biosynthesis	stability of prophage SP-β	-4.498	9.93E-22
new_1386984_1387035_c	S472	5'UTR of *ispA*	new RNA feature, 5'UTR of *ispA*	NA	-3.064	5.33E-04
new_2100447_2100579	S744	5'UTR of *dhaS*	new RNA feature, 5'UTR of *dhaS*	NA	-2.890	2.33E-23
BSU37350	*sboA*	subtilosin A	subtilosin A	antimicrobial peptide	-2.806	6.86E-07
new_3239640_3239929	S1208	5'UTR of *nupN*	new RNA feature, 5'UTR of *nupN*		-2.704	2.44E-09
BSU13190	*ispA*	intracellular serine protease	intracellular serine protease	protein degradation	-2.289	5.98E-09
BSU19310	*dhaS*	aldehyde dehydrogenase (NAD)	3-hydroxypropionaldehyde-specific aldehyde dehydrogenase (NAD)	unknown	-2.286	2.23E-13
BSU39320	*yxiB*	unknown	unknown	unknown	-2.234	2.13E-02

Table 3.5: Top 10 significantly DEGs in Δ*spoVG* in SH5

Top 10 up-regulated **(A)** and down-regulated **(B)** DEGs in *ΔspoVG* in SH5 are arranged in a decreasing and ascending order of LFC values respectively. (NA = not available)

(A)

Locus Tag	Gene	Product	Description	Function	LFC	padj
new_3788276_3788397_c	S1427	intergenic region	new RNA feature, intergenic region	NA	7.256	7.80E-135
BSU00500	*gcaD*	bifunctional N-acetylglucosamine-1-phosphate	bifunctional N-acetylglucosamine-1-phosphate uridyltransferase/glucosamine-1-phosphate acetyltransferase	cell wall metabolism	3.244	6.05E-42
BSU24620	*tasA*	major component of biofilm matrix	major component of biofilm matrix, forms amyloid fibers	biofilm formation	2.180	2.45E-06
BSU00510	*prs*	phosphoribosylpyrophosphate synthetase	phosphoribosylpyrophosphate synthetase, universally conserved protein	biosynthesis of histidine	2.058	3.24E-21
BSU06230	*iolT*	major transporter for inositol	major transporter for inositol	myo-inositol uptake	1.895	1.57E-12
BSU34300	*epsH*	undecaprenyl (UnDP) priming UDP-N-acetyl-glucosamine transferase	undecaprenyl priming UDP-N-acetyl-glucosamine transferase	synthesis of extracellular poly-N-acetylglucosamine	1.860	3.55E-02
new_4007659_4007722_c	S1504	5'UTR of *yxiT*/2	new RNA feature, 5'UTR of *yxiT*/2		1.815	1.59E-03
BSU39030	*yxiT*	unknown	unknown	unknown	1.737	2.71E-04
BSU24670	*comGG*	minor pseudopilin	DNA transport machinery, minor pseudopilin	genetic competence, DNA uptake	1.586	3.99E-02
BSU39029	*yxiT/1*	unknown	part of the *yxiT* pseudogene	unknown	1.576	3.67E-04

(B)

Locus Tag	Gene	Product	Description	Function	LFC	padj
BSUB_NCR2273724	NA	NA	NA	NA	-3.685	2.29E-12
BSUB_NCR2069983	NA	NA	NA	NA	-3.499	2.90E-08
new_2069824_2069977	SR5	antisense RNA to *bsrE*	antisense RNA, controls expression of the BsrE toxin	control of BsrE toxin expression	-3.399	1.95E-08
new_2273515_2273748	*bsrG*	hydrophobic class I toxin	hydrophobic toxin; interferes with cell envelope biosynthesis, causes membrane invaginations together with delocalization of the cell wall synthesis machinery, triggers autolysis, and inhibits protein biosynthesis	stability of prophage SP-beta	-3.047	7.45E-10
new_3510674_3510779	S1309	5'UTR of *lutP*	new RNA feature, 5'UTR of *lutP*	NA	-1.791	8.06E-03
BSU28099	*yszA*	unknown	unknown	unknown	-1.720	6.99E-04
new_2273801_2274302	S811	independent transcript	new RNA feature, independent transcript	NA	-1.490	1.41E-03
BSU36940	*ywlD*	unknown	unknown	unknown	-1.477	1.30E-02
new_1383187_1383294_c	S469	intergenic region	new RNA feature, intergenic region	NA	-1.451	3.45E-02
new_1571801_1571902_c	S551	5'UTR of *ylbJ*	new RNA feature, 5'UTR of *ylbJ*	NA	-1.413	3.03E-02

3.2.5.4 Common DEGs in all the conditions

To identify any commonly identified DEGs in Δ*spoVG* an overlap of the significantly DEGs (padj < 0.05) in each condition was determined. The data showed that 20 DEGs were commonly affected across all three conditions (Figure 3.8). Whereas 28 DEGs were common between M9 and SH5, and 17 DEGs between M9 and SH2. Only 9 DEGs were found common between the SH2 and SH5 conditions. Table 3.6 lists the commonly affected DEGs with their LFC and padj values.

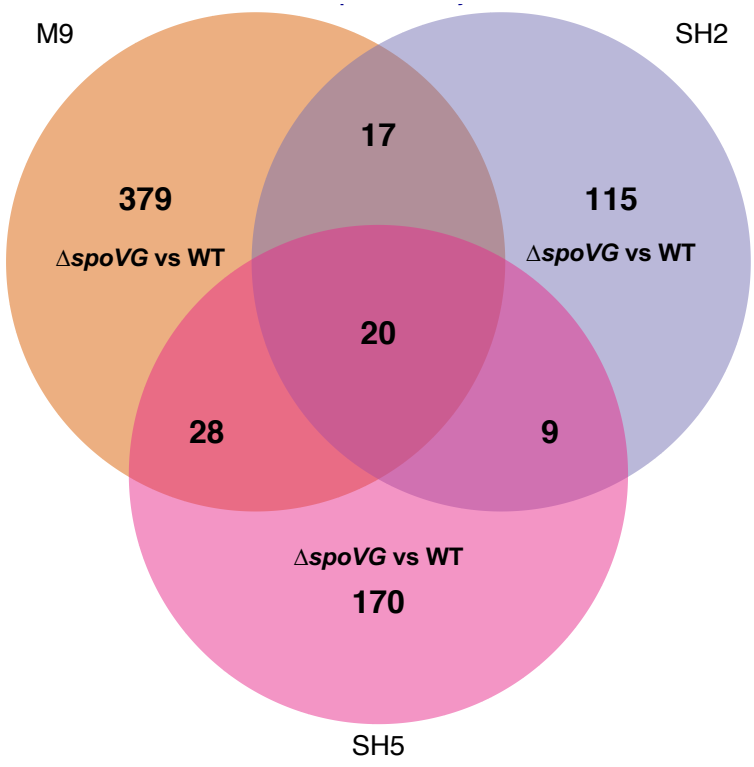


Figure 3.8: The transcriptomic analysis of the Δ*spoVG* strain identified commonly altered significantly DEGs across all the growth conditions.

Venn diagram represents the overlap of the significantly differentially expressed genes (padj < 0.05) in the Δ*spoVG* mutant compared to the WT in different growth conditions (M9, SH2 and SH5).

Table 3.6: Significantly differentially expressed genes (DEGs) in Δ*spoVG* vs WT found commonly in M9, SH2 and SH5

NA = not available

Locus Tag	Gene	Product	Description	Function	M9		SH2		SH5	
					LFC	padj	LFC	padj	LFC	padj
Up-regulated in all the conditions										
new_3788276_3788397_c	S1427	intergenic region	new RNA feature, intergenic region, flanked by S1426 downstream and *upp* upstream	NA	3.410	2.66E-72	8.517	9.26E-134	7.256	7.80E-135
BSU00500	*gcaD*	bifunctional N-acetylglucosamine-1-phosphate	bifunctional N-acetylglucosamine-1-phosphate uridyltransferase/glucosamine-1-phosphate acetyltransferase	cell wall metabolism	1.760	1.22E-38	3.459	1.16E-19	3.244	6.05E-42
BSU00510	*prs*	phosphoribosylpyrophosphate synthetase	phosphoribosylpyrophosphate synthetase, universally conserved protein	phosphoribosylpyrophosphate synthesis (biosynthesis of histidine)	1.640	4.62E-35	3.076	1.64E-13	2.058	3.24E-21
BSU38400	*epr*	minor extracellular serine protease	minor extracellular serine protease, involved in control of swarming motility	protein degradation	1.063	7.58E-05	0.942	4.36E-02	0.814	4.60E-02

Down-regulated in all the conditions										
BSUB_NCR2273724	NA	NA	NA	NA	-3.558	1.95E-30	-4.704	9.83E-23	-3.685	2.29E-12
new_2273515_2273748	*bsrG*	hydrophobic class I toxin	hydrophobic toxin; interferes with cell envelope biosynthesis, causes membrane invaginations together with delocalization of the cell wall synthesis machinery, triggers autolysis, and inhibits protein biosynthesis	stability of prophage SPß	-3.249	1.70E-26	-4.498	9.93E-22	-3.047	7.45E-10
BSUB_NCR2069983	NA	NA	NA	NA	-3.036	1.18E-22	-1.936	4.84E-03	-3.499	2.90E-08
new_2069824_2069977	SR5	antisense RNA to *bsrE*	antisense RNA, controls expression of the BsrE toxin	control of BsrE toxin expression	-2.901	3.55E-21	-2.016	3.22E-03	-3.399	1.95E-08
new_2273801_2274302	S811	independent transcript	new RNA feature, independent transcript, potential 3'UTR of *bsrG,* antisense to *yokL* and *asrG*	NA	-2.753	4.07E-21	-4.589	1.98E-20	-1.490	1.41E-03
BSU31550	*nupO*	ABC transporter for guanosine	ABC transporter for guanosine (ATP-binding protein)	uptake of guanosine	-1.601	1.51E-15	-0.890	2.88E-02	-0.666	3.47E-02

Mixed LFC										
new_3740106_3740205_c	S1399	3'UTR of *ssbB*	new RNA feature, 3'UTR of *ssbB*	NA	-1.447	1.73E-14	1.881	2.44E-09	1.232	2.54E-06
BSU36310_copy2	NA	NA	NA	NA	-1.419	9.53E-17	1.899	6.89E-11	1.221	9.86E-06
BSU24690	*comGE*	minor pseudopilin	DNA transport machinery, minor pseudopilin	genetic competence, DNA uptake	-0.880	3.68E-02	1.805	1.33E-02	1.160	7.86E-03
BSU24710	*comGC*	major pseudopilin	DNA uptake, major component of the pseudopilus	genetic competence	-0.858	8.17E-03	2.310	2.56E-04	1.247	2.71E-04
BSU24730	*comGA*	traffic ATPase	late competence protein, traffic ATPase, binds and transports transforming DNA, required for growth arrest of competent cells	genetic competence	-0.804	1.38E-03	2.369	9.50E-11	1.236	3.89E-04
BSU24720	*comGB*	DNA transport machinery	polytopic membrane protein, DNA transport machinery	genetic competence, DNA uptake	-0.767	1.06E-03	2.055	2.63E-10	1.387	4.58E-05
BSU24700	*comGD*	minor pseudopilin	DNA transport machinery, minor pseudopilin	genetic competence, DNA uptake	-0.751	4.70E-02	2.370	6.00E-04	1.065	1.27E-03
BSU35480	*degV*	fatty acid binding protein	fatty acid binding protein	phosphorylation of fatty acids	-0.638	5.52E-03	1.287	1.41E-02	1.269	3.34E-04
new_368894_371728_c	S122	intergenic region	new RNA feature, intergenic region, flanked by *yckB* downstream and *nin* upstream, antisense to BSU_03385, *yckC*, *yckD* and *yckE*	NA	-0.583	1.07E-02	2.179	1.48E-11	1.102	1.11E-03
BSU12440	*phrA*	phosphatase (RapA) inhibitor	response regulator aspartate phosphatase (RapA) inhibitor, control of the phosphorelay	control of sporulation initiation	-0.522	4.94E-02	-1.246	1.00E-02	0.994	4.63E-03

3.2.6 Gene Set Enrichment Analysis

To understand the biological changes in the transcriptome of Δ*spoVG* in each growth condition, a gene set enrichment analysis (GSEA) was carried out. The GSEA was based on the gene classification system of the *B. subtilis* functional gene classification presented in SubtiWiki (Mäder *et al.*, 2012). In this classification, each gene and its product are assigned to functional categories. These categories are divided into 4-5 sub-categories, where Category 1 (Cat1) is the broadest classification and Category 5 (Cat5) is the most specific classification. For this analysis, the Category 3 (Cat3) (geneCategories) was chosen. The gene sets with padj < 0.05 were considered as statistically significant.

From this analysis, significantly enriched gene sets were identified in each of the M9, SH2 and SH5 growth conditions (Figure 3.9). Particularly, 12 gene sets were significantly enriched in M9, 3 in SH2 and 8 in SH5. In M9, *Biosynthesis of antibacterial compounds* was the most enriched gene set amongst the 12 significantly enriched gene sets. Whereas, in SH2 *Biosynthesis/ acquisition of nucleotides* was the most significantly enriched gene set. In SH5, *Genetic competence* was the most significantly enriched gene set. Table 3.7 summarises all the significantly enriched gene sets in detail.

When analysed for commonly enriched gene sets across M9, SH2 and SH5, although no gene set was commonly enriched across all the growth conditions, one gene set - *Biosynthesis/ acquisition of nucleotides* was common between M9 and SH2, *Genetic competence* was common between SH2 and SH5, and 2 gene sets – *General stress proteins (controlled by SigB)* and *Quorum sensing* were common between M9 and SH5 (Figure 3.10).

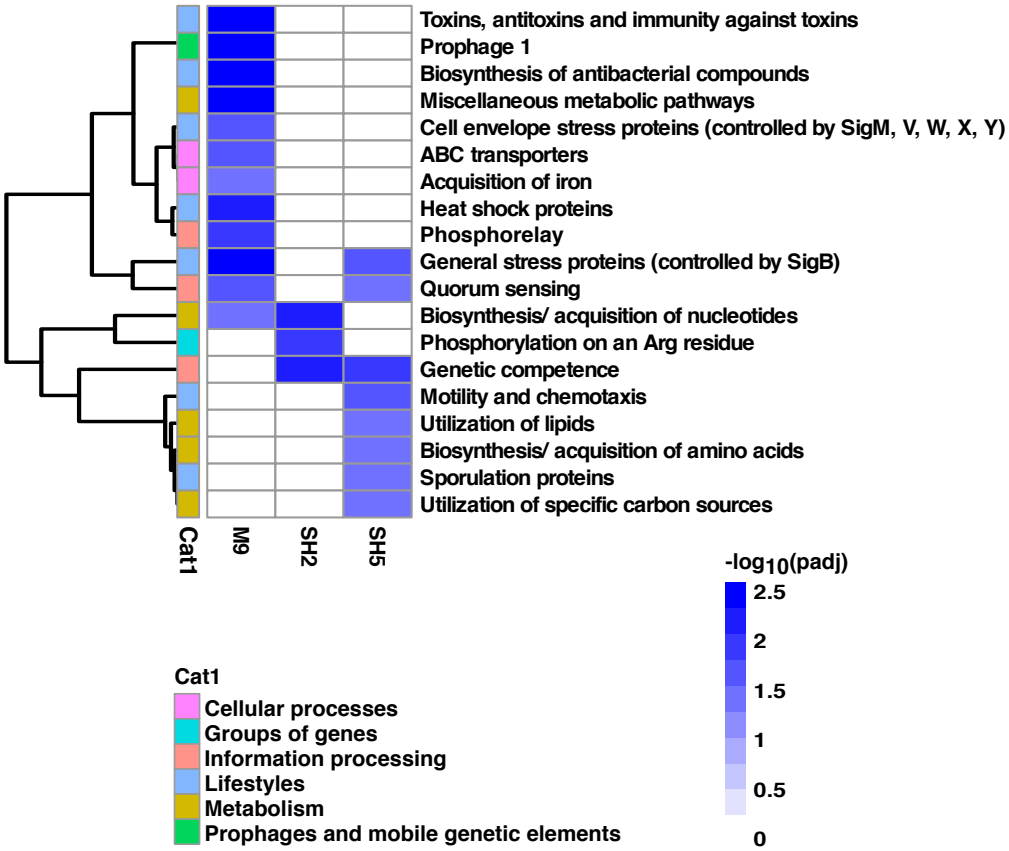


Figure 3.9: GSEA of the Δ*spoVG* transcriptomic data identifies significantly enriched gene sets in each growth condition

The pathways altered in each growth condition is indicated in as shaded box based on their adjusted *p*-value (padj). The colour scale is set from on a negative logarithm of the padj ranging from 0-maximum value form the data. The pathways that have -log(0.05) close to zero have been left unshaded to focus only on the pathways that are significantly changed. The darkest colour indicates the most significantly altered pathway.

Table 3.7: Significantly enriched gene sets identified in Δ*spoVG* from the transcriptomic data

GSEA - Category 3	padj	Total genes in pathway	Genes up-regulated	Genes down-regulated
M9				
Biosynthesis of antibacterial compounds	2.42E-03	50	1	49
General stress proteins (controlled by SigB)	2.42E-03	151	127	24
Miscellaneous metabolic pathways	2.42E-03	73	15	58
Prophage 1	2.42E-03	15	6	9
Toxins, antitoxins and immunity against toxins	2.42E-03	17	0	17
Heat shock proteins	8.07E-03	27	21	6
Phosphorelay	1.21E-02	44	17	27
Cell envelope stress proteins (controlled by SigM, V, W, X, Y)	1.97E-02	133	65	68
ABC transporters	2.55E-02	170	65	105
Quorum sensing	2.66E-02	22	7	15
Acquisition of iron	2.75E-02	44	16	28
Biosynthesis/ acquisition of nucleotides	3.63E-02	63	51	12
SH2				
Biosynthesis/ acquisition of nucleotides	6.10E-03	56	35	21
Genetic competence	6.10E-03	57	40	17
Phosphorylation on an Arg residue	1.22E-02	102	75	27
SH5				
Genetic competence	1.22E-02	55	48	7
General stress proteins (controlled by SigB)	2.03E-02	154	127	27
Motility and chemotaxis	2.03E-02	68	64	4
Biosynthesis/ acquisition of amino acids	2.74E-02	145	41	104
Sporulation proteins	3.17E-02	485	148	337
Quorum sensing	3.66E-02	21	18	3
Utilization of specific carbon sources	3.66E-02	219	142	77
Utilization of lipids	4.42E-02	30	20	10

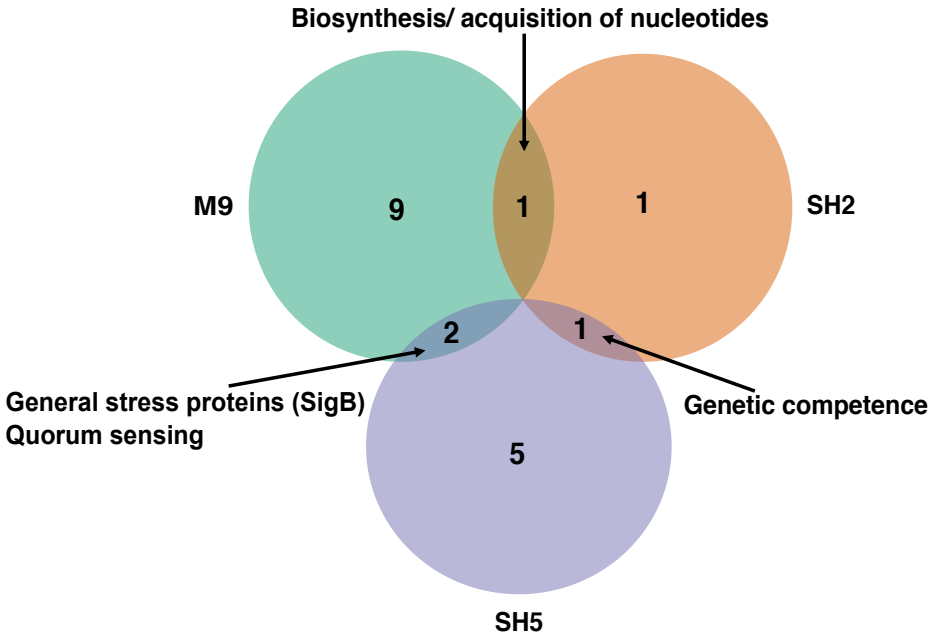


Figure 3.10: Commonly identified gene sets in Δ*spoVG* under different growth conditions

The overlap analysis of the enriched category 3 pathways in all the growth conditions.

3.2.6.1 Biosynthesis of antimicrobial compounds

The GSEA in Δ*spoVG* in M9 identified *Biosynthesis of antibacterial compounds* as the most significantly enriched gene set (Figure 3.9). In total, 50 genes were identified in this gene set out of which all, but one was down-regulated in the mutant (Table 3.7). Some of the prominently down-regulated genes in this gene set belong to operons such as *pks*, *sunA*, *pps*, *skf* and *asnH*. All of these operons in *B. subtilis* are known to produce antimicrobial compounds for example, the *pks* operon encodes for a megacomplex that participates in the biosynthesis of bacillaene – a linear polyketide/ non-ribosomal peptide (PK/NRP) (Butcher *et al.*, 2007; Straight *et al.*, 2007; Nagler *et al.*, 2016). *sunA* encodes for protein sublancin - a highly stable and potent glycopeptide with antibacterial properties (Oman *et al.*, 2011). Whereas, the *pps* operon (*ppsABCDE*) is associated with the

production of the NRPs plipastatin which is one of the most commonly produced NRPs by *B. subtilis* (Nishikiori *et al.*, 1986; Tsuge *et al.*, 1999; Tsuge, Matsui and Itaya, 2007).

B. subtilis produces a broad spectrum of antimicrobial substances based on its ecological niche or the physiology of the cells (Abriouel *et al.*, 2011; De Leersnyder *et al.*, 2018). These antimicrobial substances are generally composed of compounds such as lipopeptides, surfactin, fengycins, and bacteriocins. Most of these compounds are peptides that are either ribosomally synthesised (such as lantibiotics) or are synthesised as NRPs (Stein, 2005). Additionally, *B. subtilis* also produces other compounds such as polyketides (PK) which also have antimicrobial properties (Stein, 2005). NRPs belong to the diverse class of secondary metabolites that have a range of properties such as antibiotics, anti-cancer agents, cytostatics, immunosuppressants, pigments, siderophores or toxins (Martínez-Núñez and López, 2016). The down-regulation of genes involved in the *Biosynthesis of antimicrobial compounds* in Δ*spoVG* suggested a role of *spoVG* in M9 in their production.

3.2.6.2 Toxins, antitoxins and immunity against toxins

The GSEA in M9 also identified *Toxins, antitoxins and immunity against toxins* as one of the top enriched gene sets in Δ*spoVG* (Figure 3.9). In this gene set, 17 genes were identified in Δ*spoVG* (Table 3.7). Noticeably, all of the affected toxin/antitoxin genes were down-regulated in Δ*spoVG*. Some of these genes belong to the *bsrG*/SR4 class of toxins, the *sdpABC* and the *skf* operon. In this data, *bsrG* was down-regulated in Δ*spoVG* however SR4 was not significantly differentially expressed in the mutant. BsrG is a hydrophobic toxin and is controlled by the antitoxin sRNA SR4 (Jahn *et al.*, 2012). The SR4 antitoxin RNA inhibits the translation of *bsrG* toxin gene and destabilizes the *bsrG* mRNA (Saito, Kakeshita and Nakamura, 2008; Jahn and Brantl, 2013). The BsrG toxin acts to kill its host by interfering with the biosynthesis of the cell envelope, causing membrane invaginations

together with delocalization of the cell wall synthesis machinery, triggering autolysis and inhibition of protein biosynthesis (Jahn *et al.*, 2012).

The *sdpABC* operon encodes for the cannibalistic toxin SdpC and proteins assisting in its maturation (González-Pastor, Hobbs and Losick, 2003). The SdpC toxin kills non-sporulating *B. subtilis* that are sensitive to the toxin by disrupting their proton motive force (Lamsa *et al.*, 2012). The *skf* operon encoded for the sporulation killing factor SkfA toxin and its transporter (Allenby *et al.*, 2006). SkfA induces the degradation of the sibling cells that do not enter the sporulation pathway which provides nutrition to the sporulating cells (González-Pastor, 2011). The significantly decreased expression of the toxin genes in the Δ*spoVG* strain in M9 was intriguing. Probably, *spoVG* helps in the production of the toxins in the cells growing in minimal conditions that in turn helps them survive in nature by killing the non-sporulating cells and providing energy for the sporulating cells. The deletion of *spoVG* directly or indirectly reduces the production of toxins in *B. subtilis*.

In SH2 and SH5 several toxin/antitoxin genes were also down-regulated in Δ*spoVG*. In SH2, the genes *sdpB*, *sdpC* and S1297 (3' UTR of *sdpC*), *bsrG* and antitoxin SR5 and in SH5, *bsrG*, S811 (potential 3’ UTR of *bsrG*) and SR5 were down-regulated. The down-regulation of the toxin genes in the sporulation conditions re-enforced the hypothesis that *spoVG* may help in the biosynthesis of the toxins in the sporulating *B. subtilis* and hence facilitating the killing of the non-sporulating cells during sporulation. Also, since the toxin gene *bsrG*, ncRNA S811 and antitoxin SR5 were found to be commonly down-regulated in all the growth conditions, *spoVG* is speculated to have a role in the signalling cascade of toxin/antitoxin pathways in *B. subtilis* as well. However, more direct and experimental evidence would be needed to confirm this hypothesis.

3.2.6.3 General stress proteins (controlled by SigB)

The GSEA identified *General stress proteins (controlled by SigB)* gene set was one of the most significantly enriched gene sets in M9 and SH5 (Figure 3.9). In total, 151 genes in M9 and 154 genes in SH5 were identified in this gene set (Table 3.7). In both the growth conditions, 127 genes were up-regulated which comprised most of the affected genes in this gene set (Table 3.7). In M9, the expression of *sigB* gene was up-regulated by LFC = 1. In addition to the up-regulation of *sigB*, many other genes of the general stress response (GSR) regulon were also up-regulated in Δ*spoVG* in M9. Although, the reasons for the up-regulation of *sigB* and the other general stress response genes remain unknown, it could be speculated that perhaps deletion of *spoVG* results in the sigma factor competition in the genes affected due to the deletion (Mauri and Klumpp, 2014; Ayala, Bartolini and Grau, 2020). Nonetheless, further investigations would be needed to ascertain these changes.

Some GSR genes such as *yycD* and *yjbC* (LFC values 1.32, and 1.26 respectively) in SH5, were also up-regulated in the mutant. The *yycD* gene is involved in the survival of ethanol stress (Nicolas *et al.*, 2012) and *yjbC* is a part of the paraquat and salt stress survival which is co-expressed with the transcriptional regulator *spx* (Antelmann, Scharf and Hecker, 2000; Nicolas *et al.*, 2012). In this analysis, *spx* was not differentially expressed in the Δ*spoVG* in SH5.

In the natural habitat, *B. subtilis* is constantly challenged by a range of stress and nutritional starvation conditions. One of the strongest and most noticeable responses to such conditions is the dramatic induction of general stress proteins (GSP) by sigma factor SigB. The expression of the transcription factor SigB and its associated regulatory proteins are known to be induced under a variety of stress conditions (Boylan *et al.*, 1993) or in stationary phase (Kalman *et al.*, 1990). The SigB controlled general response mechanism is complex and comprises of more than 200 genes as members of the regulon (Helmann *et al.*, 2001; Price *et al.*, 2001;

Nannapaneni *et al.*, 2012). Since M9 and SH5 are both nutritionally limiting for the cell, it would not be surprising for the cell to activate the general stress response.

3.2.6.4 Genetic competence

The GSEA in both SH2 and SH5 identified *Genetic competence* as one of the significantly enriched gene sets (Figure 3.9). In SH2, 57 genes were identified in this gene set out of which 40 were up-regulated in Δ*spoVG* (Table 3.7). Whereas, 55 genes were identified in SH5 out of which 40 were up-regulated in the mutant (Table 3.7). Comparison of the results from SH2 and SH5 identified genes belonging to the *comGABCDEFG-spoIIIL* and *nin-nucA* operons in both the conditions. Additionally, in SH2 *comK* was up-regulated (LFC ~2). However, *comK* was not significantly differentially expressed in SH5.

Genetic competence is one of the principal developmental pathways of *B. subtilis*. The state of genetic competence is the ability of a cell to take up exogenous DNA fragments from their immediate environment into the cell (Dubnau, 1991). The state of competence can be advantageous as it allows for up-take of DNA and exchange of genetic material, which could be useful for the survival *B. subtilis* in adverse conditions such as starvation. ComK is a master transcriptional regulator of competence in *B. subtilis* (van Sinderen *et al.*, 1994). Whereas, the *comGABCDEFG-spoIIIL* operon encodes for proteins that are essential for the binding of DNA to the surface of the competent cells, and allows the uptake of the exogenous DNA during *B. subtilis* transformation (Chung and Dubnau, 1998). The gene *nin* is co-expressed with a deoxyribonuclease (DNase) NucA and forms the *nucA-nin* operon (Provvedi, Chen and Dubnau, 2001). Nin is an inhibitor of the NucA, which is a membrane-associated nuclease that catalyses DNA cleavage during transformation (Provvedi, Chen and Dubnau, 2001). Both *nucA* and *nin* are regulated by the transcription factor ComK, which is known to regulate the late competence genes (Ogura *et al.*, 2002).

During early stationary phase, a significant proportion of the cell population (1-10%) become competent as an alternative strategy of nutrient deprivation and gain survival advantages over the other sporulating cells (Schultz *et al.*, 2009). In the domesticated *B. subtilis* 168 strain, approximately 50-70% of cells commit to sporulation while the rest enter into competence as a survival strategy in extreme nutrient deprivation conditions (Schultz *et al.*, 2009). Hence, the cells that commit to the sporulation pathway have reduced competency compared to the non-sporulating cells (Schultz *et al.*, 2009). The cascade of signal transduction between sporulation and the competence state of the cell is one of the most complex pathways understood in *B. subtilis* (Schultz *et al.*, 2009).

Increased expression of competence genes in Δ*spoVG* in both of the sporulation conditions in this study indicated a possible decreased sporulation efficiency in the mutant. This, in turn, indicated a role of *spoVG* in mediating sporulation in *B. subtilis*. Nonetheless, targeted studies would be needed to elucidate the role of *spoVG* in the sporulation-competence state of *B. subtilis* under these growth conditions.

3.2.6.5 Acquisition of iron

Another significantly enriched gene set in M9 was the *Acquisition of iron* (Figure 3.9). Out of the 44 genes identified in GSEA, 28 genes were down-regulated (Table 3.7). Among these were the genes of *dhb* operon and the *yvmC-cypX* operon. The *dhb* operon encodes for the siderophore 2,3-dihydroxybenzoate (DHB) (Rowland *et al.*, 1996; May, Wendrich and Marahiel, 2001) whereas, the *yvmC-cypX* operon encodes for the biosynthesis of the chelator pulcherrimin (Tang *et al.*, 2006; Arnaouteli *et al.*, 2019).

For bacteria, efficient uptake of iron in the form of oxidised ferric ions (Fe^{3+}) is critical for their growth and viability. Iron forms an integral component of the central metabolic processes such as the tricarboxylic acid (TCA) cycle, electron transport, amino acid and nucleotide biosynthesis, nitrogen

fixation, oxidative stress response, DNA replication, and photosynthesis. However, the Fe^{3+} has low-solubility which needs specialised mobilization strategies and high-affinity transportation for its uptake. Siderophores are the bacterial secreted extracellular iron-chelating molecules which bind to the Fe^{3+} ions with high affinity and help in the transportation of the Fe^{3+} ions into the bacterial cells (Andrews, Robinson and Rodríguez-Quiñones, 2003; Miethke and Marahiel, 2007; Hider and Kong, 2010). The decreased expression of genes in the *Acquisition of iron* indicated a role of *spoVG* in iron metabolism in *B. subtilis*. Further experimental evidence would be needed to establish these observations.

3.2.6.6 Biosynthesis/ acquisition of nucleotides

The GSEA also identified the *Biosynthesis/ acquisition of nucleotides* gene set in the M9 and SH2 data (Figure 3.9). In M9, 63 genes were identified in this gene set out of which 51 were up-regulated whereas, in SH2 56 genes were identified out of which 35 were up-regulated (Table 3.7).

In both the M9 and SH2 conditions the *nupNOPQ* operon and the gene *prs* were affected. While the genes of the *nupNOPQ* operon were down-regulated, *prs* was strongly induced in both M9 and SH2. The *nupNOPQ* operon encodes for the guanosine transporters that are essential for the uptake of the nucleoside guanosine (Belitsky and Sonenshein, 2011). *prs* is expressed from the *gcaD-prs-ctc* operon which is located down-stream to *spoVG* in the *B. subtilis* genome (Hilden, Krath and Hove-Jensen, 1995). The *gcaD-prs-ctc* operon is essential for the growth of *B. subtilis* and forms a part of the stringent response (Eymann *et al.*, 2002; Kobayashi *et al.*, 2003). In the operon, expression of *gcaD* and *prs* is essential whereas the expression of *ctc* is not. *ctc* can also be independently expressed with its own promoter P_{gcaD} (Nicolas *et al.*, 2012). *prs* is involved in the purine nucleoside production in *Bacillus amyloliquefaciens* and *B. subtilis* (Zakataeva *et al.*, 2012). In SH2, the *pyr* operon was drastically up-regulated in the Δ*spoVG*. The *pyr* operon is essential in the de novo

pyrimidine biosynthesis in *B. subtilis* (Turner, Lu and Switzer, 1994). The perturbations observed in this study in the biosynthesis of nucleotides suggest a role of *spoVG* in nucleotide metabolism. Experimental evidence would be needed to decipher these roles of *spoVG* in *B. subtilis*.

3.2.7 KEGG Pathway Analysis of the Differentially Expressed Genes

GSEA identifies the functionally associated genes in a biological process or molecular function based on their functional annotations. However, it does not calculate the frequency of occurrence of the genes in the entire set of genes involved in a specific pathway. Therefore, to identify the enrichment of the biological pathways solely due to the significantly differentially expressed genes identified in Δ*spoVG*, a Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis was carried out using hypergeometric statistical test. KEGG pathways enriched with padj < 0.05 were considered statistically significant.

The KEGG pathway analysis identified significantly enriched pathways in each of the growth conditions tested (Figure 3.11). Three KEGG pathways – viz., *Biosynthesis of siderophore group nonribosomal peptides*, *Nonribosomal peptide structures* and *Biosynthesis of secondary metabolites* - unclassified were identified in M9. In the sporulation conditions ten KEGG pathways in SH2 and only one pathway - *Inositol phosphate metabolism* was significantly enriched in SH5 (Figure 3.11). No KEGG pathways were common in all three conditions. Table 3.8 lists the identified KEGG pathways in detail.

The *Biosynthesis of siderophore group nonribosomal peptides* in M9 comprised of the *dhb* operon which is involved in the biosynthesis of the catecholate siderophore bacillibactin (May, Wendrich and Marahiel, 2001). All of the genes in this pathway were down-regulated in Δ*spoVG*. These results were in line with the down-regulation of genes identified in the gene set *Acquisition of iron* (section 3.2.6.5) where most of the genes were down-regulated in the mutant.

Table 3.8: Summary of KEGG pathways enriched in Δ*spoVG* in *B. subtilis*

The pathways are arranged in an ascending order of adjusted *p*-value (padj).

KEGG Pathway – M9	padj	No. of gene in pathway (k)	No. of genes enriched (x)
Biosynthesis of siderophore group nonribosomal peptides	1.22E-03	6	5
Nonribosomal peptide structures	2.26E-03	9	6
Biosynthesis of secondary metabolites - unclassified	2.61E-02	10	5
KEGG Pathway – SH2	**padj**	**No. of gene in pathway (k)**	**No. of genes enriched (x)**
Pyrimidine metabolism	9.09E-06	49	9
Ribosome	3.26E-05	57	9
Quorum sensing	3.39E-03	68	7
ABC transporters	7.33E-03	140	10
Histidine metabolism	1.43E-02	18	3
beta-Alanine metabolism	2.65E-02	9	2
Valine, leucine and isoleucine degradation	2.79E-02	23	3
Arginine and proline metabolism	3.48E-02	25	3
Metabolic pathways	3.88E-02	598	25
Lysine degradation	3.89E-02	11	2
KEGG Pathway – SH5	**padj**	**No. of gene in pathway (k)**	**No. of genes enriched (x)**
Inositol phosphate metabolism	1.72E-02	15	8

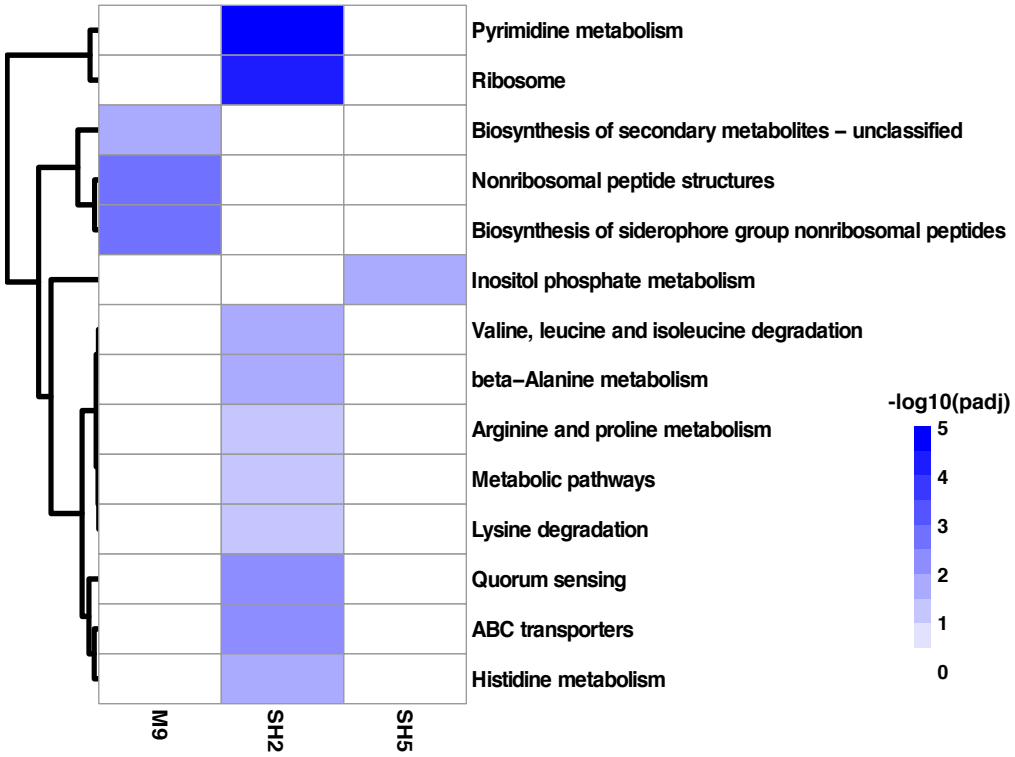


Figure 3.11: Significantly enriched KEGG pathways identified in the transcriptomic analysis of Δ*spoVG* in *B. subtilis*

Only the significantly enriched pathways are shown with shaded colours. The intensity of the shaded is proportional to the significance ($-\log_{10}(padj)$) of the enrichment analysis. Deepest shade indicates the most significant pathway.

Nonribosomal peptide structures in M9 comprised the genes of *pps* operon and all the genes were down-regulated. This result was comparable to the down-regulation and enrichment of genes involved in the *Biosynthesis of antimicrobial compounds* (section 3.2.6.1). All the genes in *Pyrimidine metabolism* in SH2 were up-regulated in Δ*spoVG* and were comparable to the *Biosynthesis/ acquisition of nucleotides* (section 3.2.6.6) gene set in SH2 (Figure 3.9). Several amino acid metabolic pathways were also enriched in the KEGG pathway analysis of Δ*spoVG* in SH2 (Figure 3.11). These observations suggested a role of *spoVG* in nucleotide and amino acid metabolism in *B. subtilis*. Many metabolic genes were also significantly

up-regulated and enriched in the KEGG pathway analysis that included the carbon and nitrogen metabolism in *B. subtilis*.

In SH5, the KEGG pathway *Inositol phosphate metabolism* comprised of the *iol* operon. All the genes identified in this pathway were up-regulated in the mutant. The *iol* operon encodes for proteins that are involved in the catabolism of the myo-inositol (MI) (Yoshida *et al.*, 1997, 2008). Myo-inositol is utilized by *B. subtilis* as an alternate carbon source. This pathway was comparable to the identification of gene set *Utilization of specific carbon sources* in SH5 (Figure 3.9).

3.2.8 Affected Sigma Factor Regulons due to Δ*spoVG*

To identify the effect of Δ*spoVG* on the regulons, the significantly DEGs were matched to the sigma factor regulons. The analysis identified several regulons affected in each of the growth conditions (Figure 3.12). In M9 most of the affected sigma factor regulons decreased in the absence of Δ*spoVG* (Figure 3.12 (A)). The most affected regulons included the SigI (decreased), SigB (increased) and SigW (decreased) sigma factor regulons. The identification of up-regulation of SigB sigma factor regulons was in line with the observation of enriched of the general stress gene set (section 3.2.6.3). SigW is an extra cytoplasmic function (ECF) sigma factor which is generally involved in the cell wall metabolism and resistance to antimicrobial compounds (Helmann, 2016). *Biosynthesis of antimicrobial compounds* was one of the most significantly enriched gene sets in M9 where majority of the genes were down-regulated in the absence of *spoVG* (section 3.2.6.1). Hence, it was speculated that *spoVG* has a role in enabling the transcription of genes involved in the synthesis of antimicrobial compounds in *B. subtilis*.

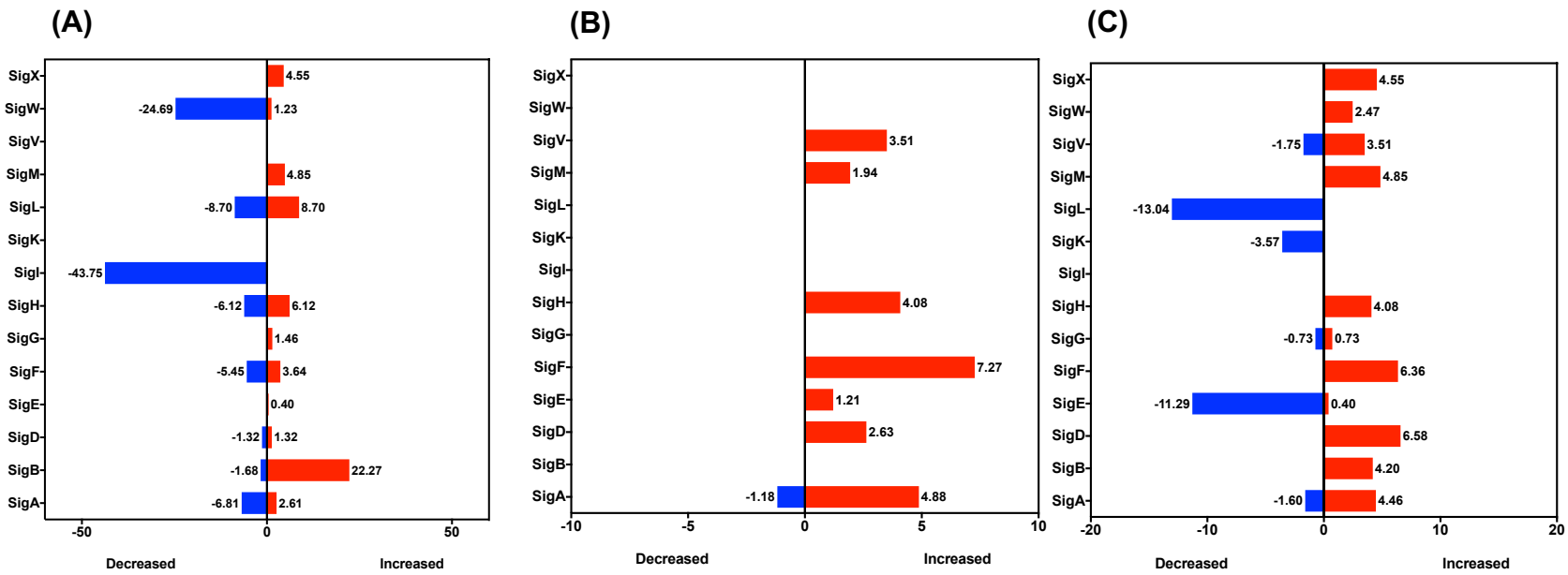


Figure 3.12: Sigma factor regulons affected in Δ*spoVG* in the transcriptomic analysis

The bar graph shows the percentage of change (X-axis) of regulons (Y-axis) in each growth condition **(A)** M9, **(B)** SH2 and **(C)** SH5. The negative bars on x-axis represent the percentage of change in decreased regulons.

Also, the genes of the SigI sigma factors are involved in iron metabolism (Ramaniuk *et al.*, 2018). The *Acquisition of iron* gene set was also one of the significantly enriched gene sets in M9, where most of the affected genes in Δ*spoVG* were down-regulated (section 3.2.6.5). These observations therefore taken together suggested a role of *spoVG* is in a positive regulation of genes involved in the Biosynthesis of antimicrobial compounds and iron metabolism and in the suppression of general stress response genes.

Among the sporulation conditions, SH2 showed a distinct overall increase in the sigma factor regulons which indicated a suppressive role of *spoVG* in SH2 (Figure 3.12 (B)). These changes in SH2 were however only minute. A maximal change in SH2 was observed in the SigF sigma factor regulon. SigF is an early forespore-specific sporulation specific sigma factor which regulates the activation of genes involved in the production of germination protease and enzymes involved in DNA repair (Wang *et al.*, 2006). In vegetative cells, SigF regulates the expression of genes involved in fatty acid metabolism and anabolism (Overkamp and Kuipers, 2015). An up-regulation of SigF regulon genes indicated a suppressive role of *spoVG* in the expression of SigF regulon genes in SH2.

In SH5, while most of the sigma factor regulons increased, some also decreased in response to Δ*spoVG* (Figure 3.12 (C)). The most affected sigma factor regulons in SH5 included SigE (decreased), SigF (increased) and SigL (decreased). SigE is an early mother-cell specific sigma factor (Sun, Cabrera-Martinez and Setlow, 1991) and encompasses 253 genes in the SigE regulon in *B. subtilis* (Eichenberger *et al.*, 2003). The up-regulation of the SigF regulon in SH5 in Δ*spoVG* indicated a suppressor role of *spoVG* on the fore-spore specific genes in SH5. The down-regulation of the SigE regulon genes suggested that *spoVG* facilitates the expression of SigE regulon genes in SH5.

The sigma factor SigL is an important link between the central carbon and nitrogen metabolism in *B. subtilis* (Choi and Saier, 2005) and regulates the acetoin catabolic pathway (Ali *et al.*, 2001). In this analysis, most of the SigL

sigma factor regulon genes were down-regulated in Δ*spoVG* in SH5. The knowledge about the connection between the carbon and nitrogen metabolism in *B. subtilis* is still far from complete, especially in the sporulating cells (Wacker *et al.*, 2003; Choi and Saier, 2005). The down-regulation of SigL regulon in Δ*spoVG* suggested a role of *spoVG* in regulating the carbon and nitrogen metabolism in *B. subtilis* in SH5. Further experimental evidence would be needed to understand these changes in detail. In summary, major changes in Δ*spoVG* were recorded in the early mother-cell (SigE regulon) and fore-spore specific genes (SigF regulon) and the link between the carbon and nitrogen metabolism (SigL regulon) were recorded in SH5.

3.3 Conclusion

This chapter presents the first global transcriptomic analysis of the Δ*spoVG* strain in three different growth conditions i.e., minimal media supplemented with glucose (M9), sporulation hour 2 (SH2) and sporulation hour 5 (SH5) in *B. subtilis*. The data analysis focused on identifying the differentially expressed genes (DEGs) in each condition and find the putative ncRNAs that were included in the differential analysis of gene expression. The differential analysis identified hundreds of DEGs in each of M9, SH2 and SH5. A maximal change in gene expression was identified in M9 with 444 DEGs. Among the sporulation conditions, most changes were observed in SH5 with 227 DEGs where most genes were up-regulated in the mutant. In SH2 161 gene were significantly differentially expressed. The significant changes in the gene expression profile of Δ*spoVG* each of the growth conditions tested indicated a functional role of *spoVG* in each of M9, SH2 and SH5. However, no changes were observed in the growth of Δ*spoVG* in in M9 when compared with the WT. A slight increase in the growth of Δ*spoVG* and Δ*spoVG*::*spoVG* was also observed in M9 which is speculated to be due to a possible mild polar effect on the downstream genes as a result of the deletion.

Further, to identify the biological relevance of the changes observed in the transcriptomic data, a GSEA of all the genes identified in this study was carried out. The results showed significant enrichment of biological gene sets in each of the growth conditions tested. The detailed analysis identified genes sets such as *Biosynthesis of antimicrobial compounds*, Toxins, antitoxins and immunity against toxins and *General stress proteins (controlled by SigB)* in M9. Some commonly enriched gene sets were also identified in different growth conditions such as *Biosynthesis/ acquisition of nucleotides* was common between M9 and SH2, *Genetic competence* between SH2 and SH5, *General stress proteins (controlled by SigB)* and *Quorum sensing* between M9 and SH2. The overlap between different gene sets indicated a potential common role of *spoVG* in different growth conditions. The GSEA also identified *Acquisition of iron* as a significantly enriched gene set where most of the identified genes were down-regulated.

In addition to GSEA, a KEGG pathway analysis was also carried out to identify the KEGG pathways affected only due to the DEGs in each growth condition. The KEGG enrichment analysis also identified significantly enriched pathways in each of M9, SH2 and SH5. Similarities between GSEA and KEGG pathways were identified in some of the enriched KEGG pathways. For example, enrichment of Nonribosomal peptides structures, and *Biosynthesis of siderophore group nonribosomal peptides* in M9 was akin to the *Biosynthesis of antimicrobial compounds* and *Acquisition of iron*, respectively that were identified in the GSEA of M9. These observations indicated a role of *spoVG* in the biosynthesis of antimicrobial compounds and iron metabolism in *B. subtilis*. Similarly, changes were also observed in the amino acid metabolism, carbon metabolism and nitrogen metabolism which indicated a potential role of *spoVG* in these biological processes.

In further analysis of the sigma factors regulons, major changes were observed in the sigma factors regulating sporulation in both the SH2 and SH5. These results pointed towards a potential suppressor role of *spoVG* in SH2 and enhancer of SH5. A change in SigL sigma factor regulon genes

was also observed that indicated a role of *spoVG* in the crucial link between the carbon and nitrogen metabolism in sporulation conditions in *B. subtilis*.

Taken together, the finding of this study support a global effect of *spoVG* in the gene expression profile of *B. subtilis*. SpoVG is known as a pleiotropic regulator of gene regulation in *S. aureus*, *L. monocytogenes* and *B. burgdorferi* (Burke and Portnoy, 2016; Liu, Zhang and Sun, 2016; Savage *et al.*, 2018). In a recent study in *Bacillus anthracis,* SpoVG has been shown to have an important role in the sporulation of the cells (Chen *et al.*, 2020). These observations collectively show the multifunctionality of *spoVG* in different bacteria. Multifunctionality is a known phenotype of global gene regulators. For example, as a global regulator of gene expression, the RNA-binding protein Hfq also shows pleiotropic changes in bacteria such as *E. coli* (reviewed in (Vogel and Luisi, 2011)) and *Clostridium difficile* (Boudry *et al.*, 2014). Thus, observation of global changes in the total transcriptome of Δ*spoVG* indicated a potential role of SpoVG as a novel global regulator of gene expression in *B. subtilis*. Nonetheless, more experimental evidence would be needed to validate the changes reported in this chapter, some of which will be discussed in Chapter 7 of this thesis.

Chapter 4 Proteomic Analysis of Δ*spoVG*

4.1 Introduction

To investigate the role of SpoVG in *Bacillus subtilis*, the impact of loss of SpoVG at a protein level was examined by global quantitative proteomics. The study of SpoVG at both the transcriptomic and the proteomic level would facilitate the understanding of the global role of the gene and its protein product in *B. subtilis*, which are otherwise missed by the transcriptomic investigations alone. Therefore, to monitor and have a comparative analysis of the transcriptomic and the proteomic analyses, the Δ*spoVG* and WT were grown and sampled in the same growth media as in the transcriptomic study (Chapter 3) i.e., minimal media supplemented with glucose (M9), sporulation hour 2 (SH2) and sporulation hour 5 (SH5). Additionally, the commonly used nutrient rich media lysogeny broth (LB) was also incorporated.

For quantifying the total protein content of the lysates, a label-free quantification (LFQ) approach using liquid chromatography–mass spectrometry (LC-MS/MS) was adopted. Unlike the standard methods of quantitative mass-spectrometry, the LFQ approach does not use a stable isotope to label the peptides. Instead, the method uses the relative quantification of amount of proteins between two or more biological samples (Bantscheff *et al.*, 2007; Asara *et al.*, 2008). Also, the use of LFQ is considerably more cost effective, time efficient and extremely sensitive approach to identify the peptides out of a complex mixture compared to traditional protein quantification methods (Boehmer *et al.*, 2010).

Via this study I present the first global proteome analysis of the Δ*spoVG* mutation in the *B. subtilis* 168, *trp*+ strain. This chapter discusses the identification of the differentially produced proteins (DPPs) in Δ*spoVG* strain across all the growth conditions. Further, the proteomic data was functionally characterised via Gene Set Enrichment Analysis (GSEA) and

KEGG pathway enrichment analysis to identify the downstream effect of the deletion of *spoVG* on the proteome of *B. subtilis*. The chapter also presents the changes in the protein levels of the sigma factor regulons as a result of *spoVG* deletion. Thus, the following sections will discuss the results of the experiments carried out in this study and elaborate on their significance in the biology of *B. subtilis*. The output of the analysis will also be helpful as a model study to understand the role and function of SpoVG in pathogenic bacteria such as *Listeria monocytogenes*, *Staphylococcus aureus*, *Borrelia burgdorferi* and *Bacillus anthracis* where SpoVG is being actively investigated.

4.2 Results and Discussion

4.2.1 Growth of Δ*spoVG* in LB

To study the effect of Δ*spoVG* on the growth of the bacteria, the mutant strain was grown in LB and compared to the WT. A complemented strain (Δ*spoVG*::*spoVG*) was also included. The growth curves showed no growth defect in Δ*spoVG* (Figure 4.1). However, a mild growth defect in Δ*spoVG* was observed in M9 (Chapter 3). Therefore, this result suggests that perhaps SpoVG has some role in M9 but not in LB.

4.2.2 Cell Culture and Sample Preparation for LC-MS/MS

The WT and Δ*spoVG* were grown in LB, M9 and sporulation media in biological triplicates. The samples were collected at mid-exponential phase for LB and M9, whereas for the sporulation conditions the samples were collected at hour 2 (SH2) and hour 5 (SH5) after mid-exponential growth in rich medium (Chapter 2, section 2.1.1.3). Total protein extraction and trypsin digestion of the proteins were carried out as described in Chapter 2 (section 2.3.2.2). The peptides were purified via High-Performance Liquid

Chromatography (HPLC) and quantified for tandem mass spectra (MS^2) on a Thermo Scientific Fisher Orbitrap (Chapter 2). This study did not account for the differences in the secreted or membrane bound proteins due to the absence of SpoVG.

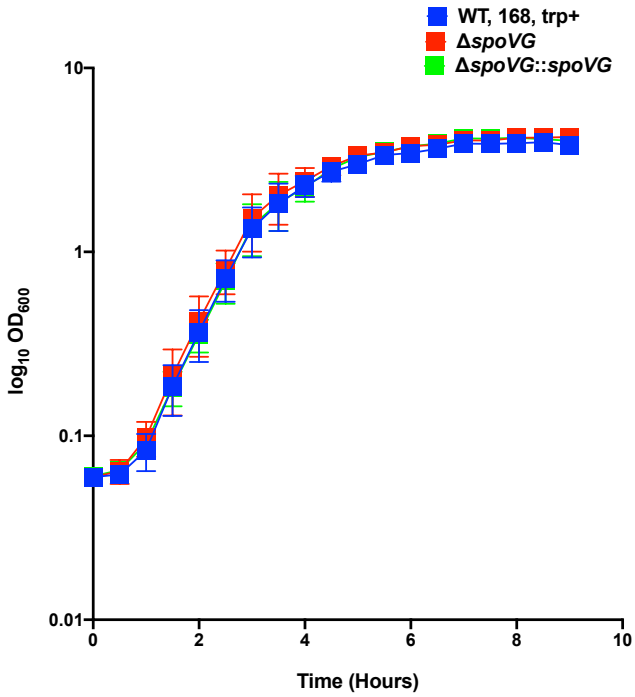


Figure 4.1: Δ*spoVG* shows identical growth to the WT in LB broth

The growth curves represent the growth of WT, Δ*spoVG* and Δ*spoVG*::*spoVG* in LB broth. Each line on the growth curve represents the average of 3 biological replicates ± standard errors of the means (SEM).

4.2.3 LC-MS/MS Data Pre-Processing and Quality Analysis

The LFQ quantification of the proteins on a standard LC-MS/MS mass-spectrometry generates the abundance or the peptide ion-intensities of the fragmented proteins. For this study, the MS^2 spectra known as the LFQ intensities were used to identify the *B. subtilis* proteins (Proteome ID: UP000001570; Genome Accession: AL009126)) using the MaxQuant tool (Chapter 2). Figure 4.2 illustrates the major steps in the proteomic data filtering and quality analysis (QA).

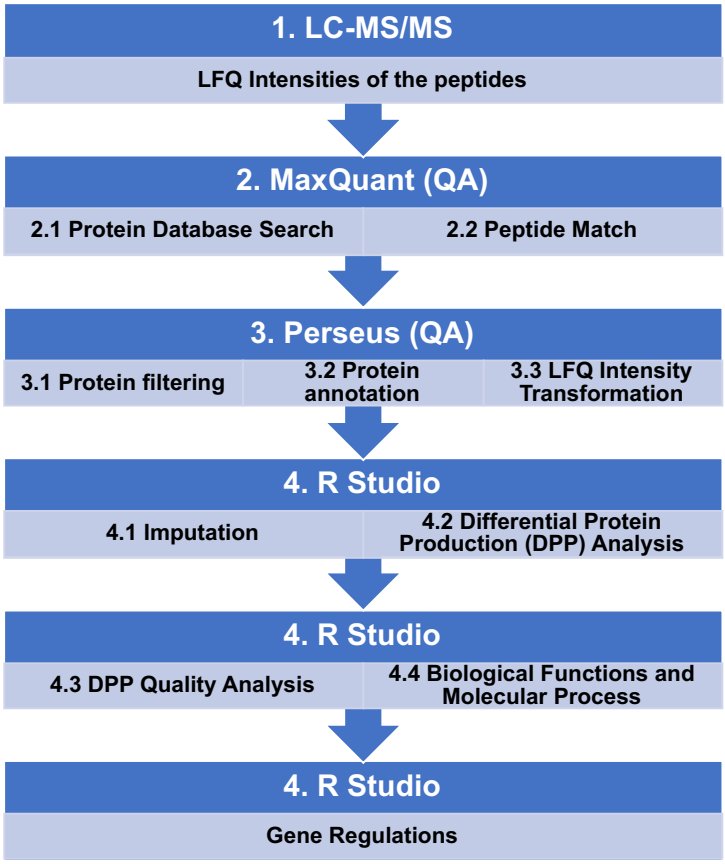


Figure 4.2: Schematic of the proteomic analysis of the DPPs in Δ*spoVG*

4.2.3.1 Data filtering and transformation

The MS^2 identified 2,866 proteins for each of the growth conditions for both the WT and Δ*spoVG* samples. Before comparison of the differential protein abundance, the potential contaminants were filtered out using the Perseus software, which resulted in 2,804 proteins for all the growth conditions. To verify if the obtained LFQ intensities followed a normal distribution, the $\log_2$ transformed LFQ intensities of the proteins were plotted against their count values. The results showed a normal distribution for all samples for each of the biological replicates (Figure 4.3). This indicated a good quality data set.

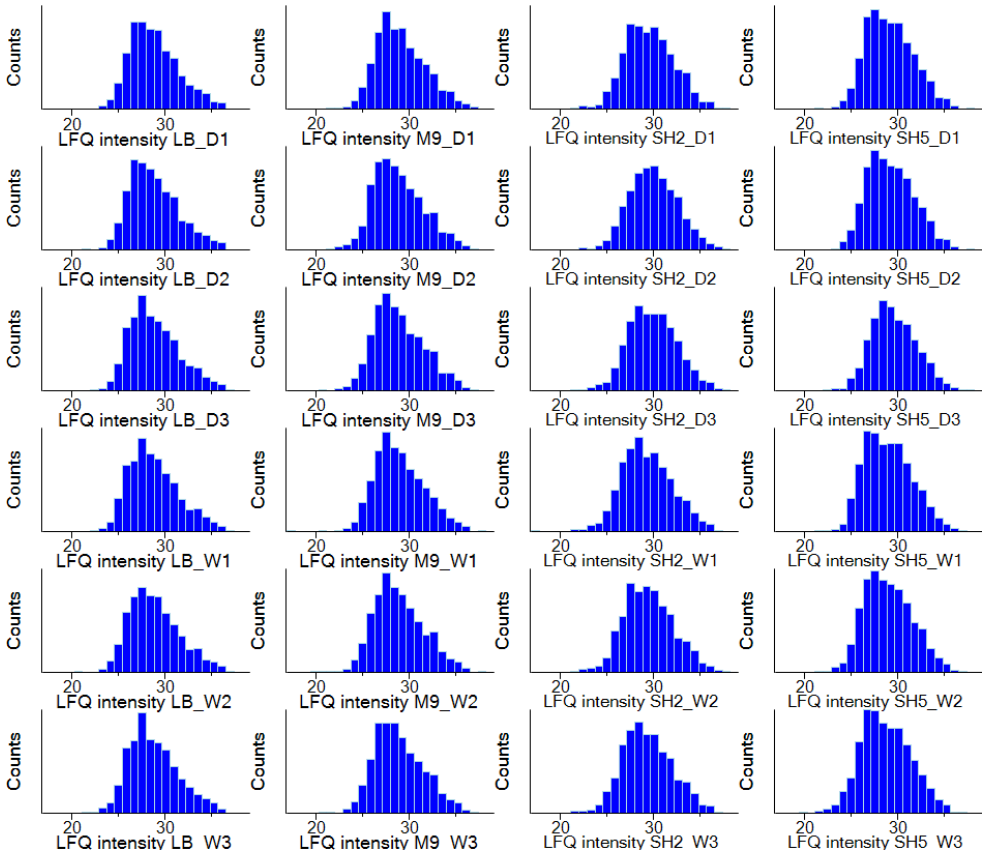


Figure 4.3: Normal distribution of the log_2 transformed protein LFQ intensities identified in LC-MS/MS experiment

The bell-curves represent the log_2 transformed intensities of the 2,804 proteins in each of the biological replicates of WT and Δ*spoVG* in each growth conditions (LB, M9, SH2 and SH5) before the cleaning steps applied on Perseus. The counts represent the number of matching peptides found for each of the proteins.

4.2.3.2 Experimental quality check

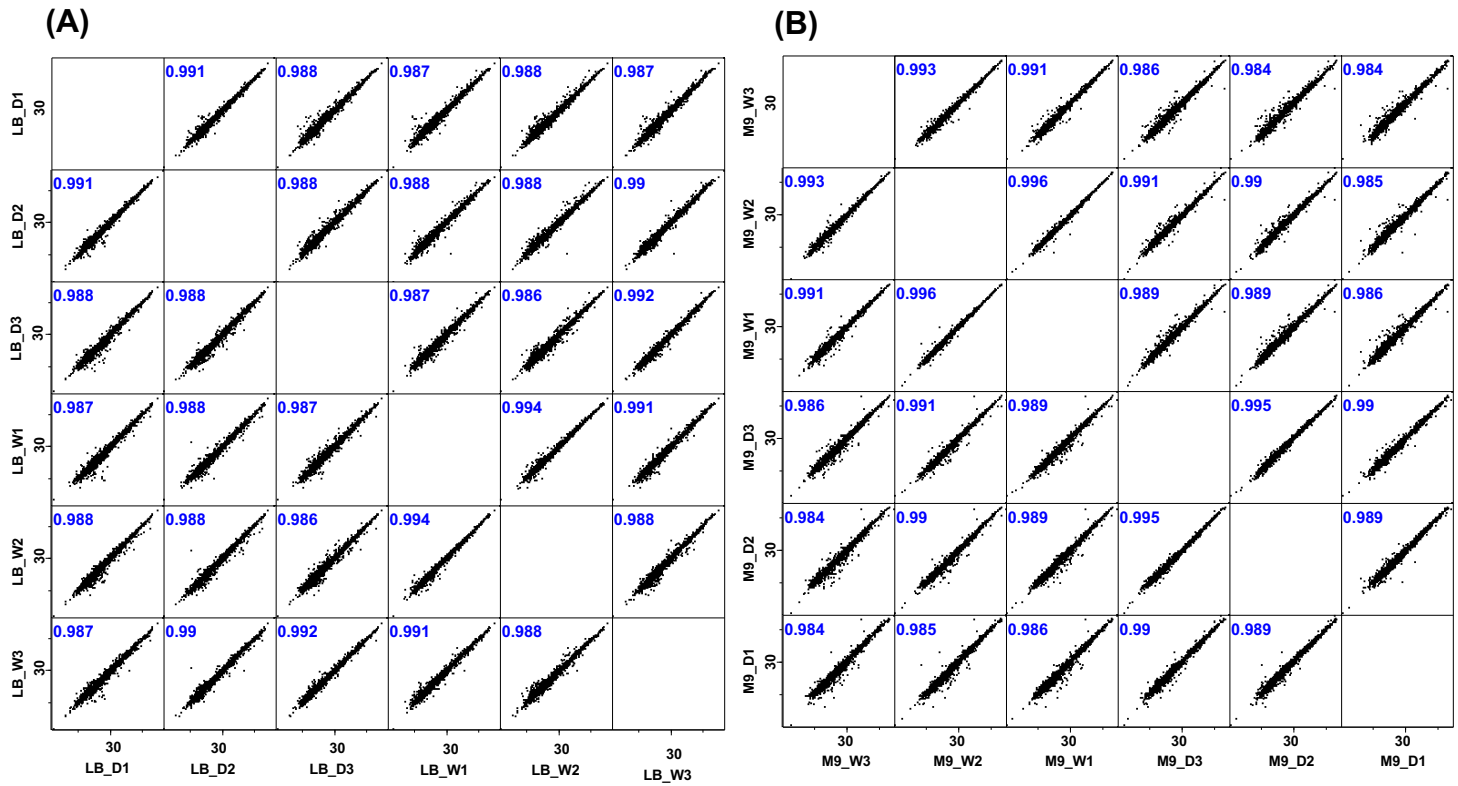
After the initial filtration, the proteins were further filtered for valid values and at least two unique peptides for each identified protein (Tyanova *et al.*, 2016). This filtration resulted in 2,422 proteins across each sample. To verify the similarity between the biological replicates and the reproducibility of the experiment, the transformed LFQ intensities were analysed by

Pearson correlation. The results were plotted on a multi-scatter plot. The growth condition-wise comparison of Δ*spoVG* vs WT showed a very high correlation (R > 0.98) between the two strains grown in LB and M9 (Figure 4.4 (A and B)). This suggested that the deletion of *spoVG* does not alter the global proteome of the cells dramatically in response to growth in nutrient rich and minimal media. However, the correlation (R > 0.98) between the biological replicates of each strain of the WT and Δ*spoVG* suggested that the experiments were highly reproducible.

For the sporulation conditions, the strains of the Δ*spoVG* and WT showed R values between 0.91-0.96 (Figure 4.4 (C and D)), indicating some changes in the proteome of Δ*spoVG* in both the sporulation conditions tested. Further, the correlation among the biological replicates of the WT and the Δ*spoVG* remained high (R > 0.98) for both the sporulation conditions (Figure 4.4 (C and D)). This indicated highly reproducible biological replicates of both WT and Δ*spoVG*.

4.2.3.3 Data Imputation and PCA clustering

Some of the proteins in the 2,422 protein thus far had duplicated Ensembl gene IDs. Therefore, only unique entries were retained for each protein identified. This resulted in retrieval of 2,414 proteins for all the samples in the study which were used for clustering analyses and further downstream differential protein production analyses. Mass-spectrometric quantification of peptides generates missing values (NA) for some proteins. The reasons for missing values in the MS^2 data could be either biological or technical such as (1) the peptides are missing in the protein sample, (2) peptide concentrations might be too low to be detected by the mass-spec instrument, and (3) peptides may be present but not detected (Bijlsma *et al.*, 2006; Hrydziuszko and Viant, 2012; Karpievitch, Dabney and Smith, 2012; O'Brien *et al.*, 2018; Wei *et al.*, 2018).


(A)
LB_D1
LB_D2
LB_D3
LB_W1
LB_W2
LB_W3
0.991 0.988 0.987 0.988 0.987
0.991 0.988 0.988 0.988 0.99
0.988 0.988 0.987 0.986 0.992
0.987 0.988 0.987 0.994 0.991
0.988 0.988 0.986 0.994 0.988
0.987 0.99 0.992 0.991 0.988
(B)
M9_W3
M9_W2
M9_W1
M9_D3
M9_D2
M9_D1
0.993 0.991 0.986 0.984 0.984
0.993 0.996 0.991 0.99 0.985
0.991 0.996 0.989 0.989 0.986
0.986 0.991 0.989 0.995 0.99
0.984 0.99 0.989 0.995 0.989
0.984 0.985 0.986 0.99 0.989

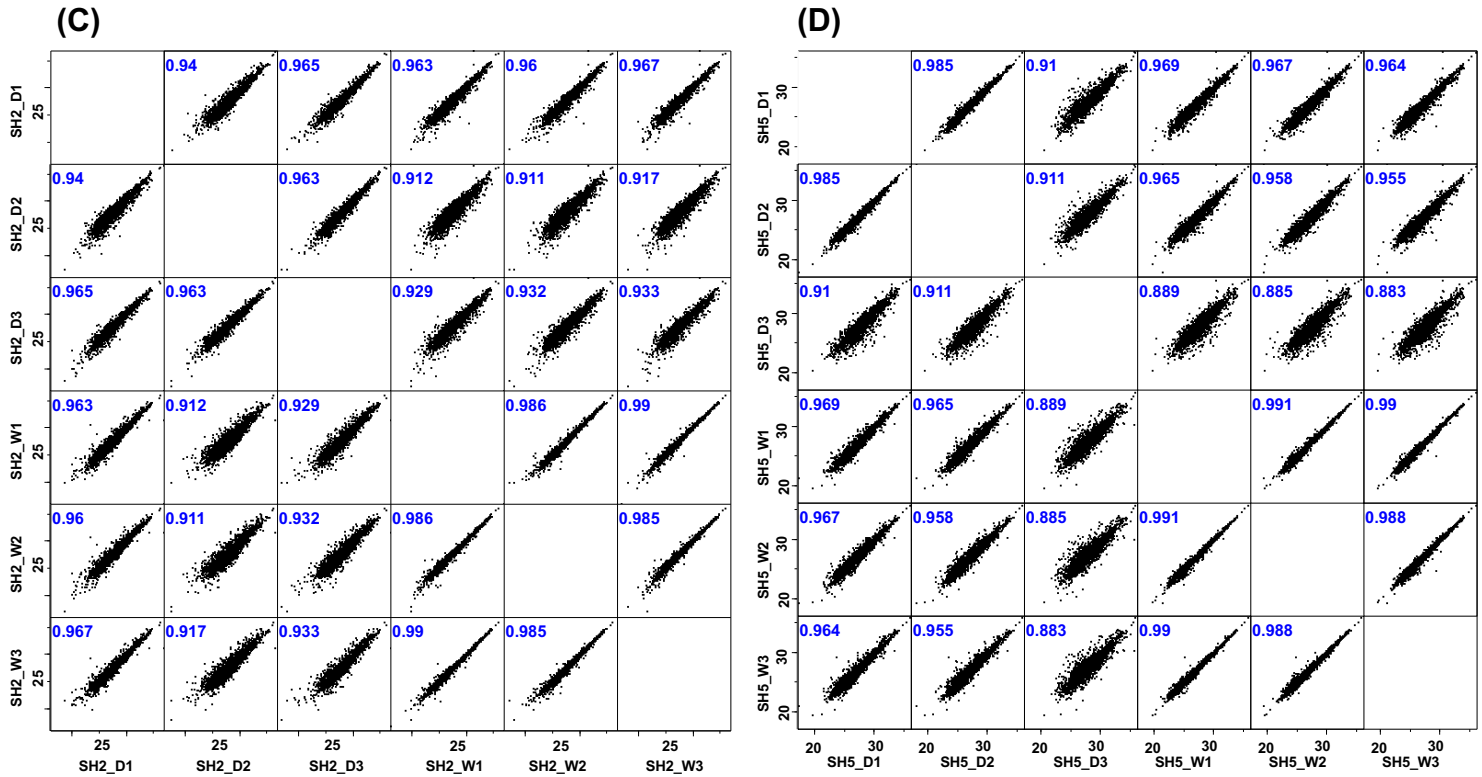


Figure 4.4: Pearson correlation analysis of the proteomes from the WT and Δ*spoVG* samples compared condition wise showed a good correlation amongst the samples

Correlation between log_2 transformed LFQ intensities were measured using Pearson correlation function on Perseus between 2,422 proteins in each sample. The scatter plot shows all the biological replicates condition-wise used in the experiment. Panel A to D represent the samples assays in the growth medium LB, M9, SH2 and SH5 respectively. Strains: D = Δ*spoVG*; W = WT.

The presence of NA values hinders in the statistical analysis of the data set since, ignoring proteins with NA values could result in losing information which might be important for the data analyses. Several methods of handling missing values exist, such as removing the data with NA values or imputation. Nonetheless, there is no single solution to handle the problem of missing values. The method chosen to deal with this problem is very specific to each study.

In this study, the missing values were dealt with imputation, as described in Chapter 2 (section 2.3.5). The imputated values were used for clustering and differential analysis of the proteins. To identify the variances and clustering within the samples, principal component analysis (PCA) was carried out. The results of the PCA showed that the replicates of the WT and the Δ*spoVG* clustered separately based on the growth media used (Figure 4.5) where, the principal component 1 (PC1) accounted for 53% of the variance and principal component 2 (PC2) for 18% variance. The separate clustering of the samples based on the growth media was expected and was a proof of concept of different proteomic profiles generated based on the different growth conditions. Additionally, analysis of the PC1 revealed that major source of variation in the samples was due to the sporulation conditions versus the LB and M9; wherein all the samples of SH2 and SH5 clustered on the positive x-axis (Figure 4.5). While the variations in PC2 was largely from the M9 and SH2 samples, which was similar to the observation in the transcriptomic data (Figure 3.4 and Figure 4.5).

Replicates in each growth condition were also clustered individually to visualise any changes arising due to the genotype. The analysis of PC1 showed distinct clustering of the WT and Δ*spoVG* samples in each growth condition (Figure 4.6). Tight clustering was observed between the WT replicates in SH2 and SH5 conditions (Figure 4.6 (C) and (D)). However, the replicates of WT in LB and M9 (Figure 4.6 (A) and (B)); and Δ*spoVG* replicates in each condition showed rather lose clustering (Figure 4.6). This

could be attributed to the batch effect incorporated during sample preparation.

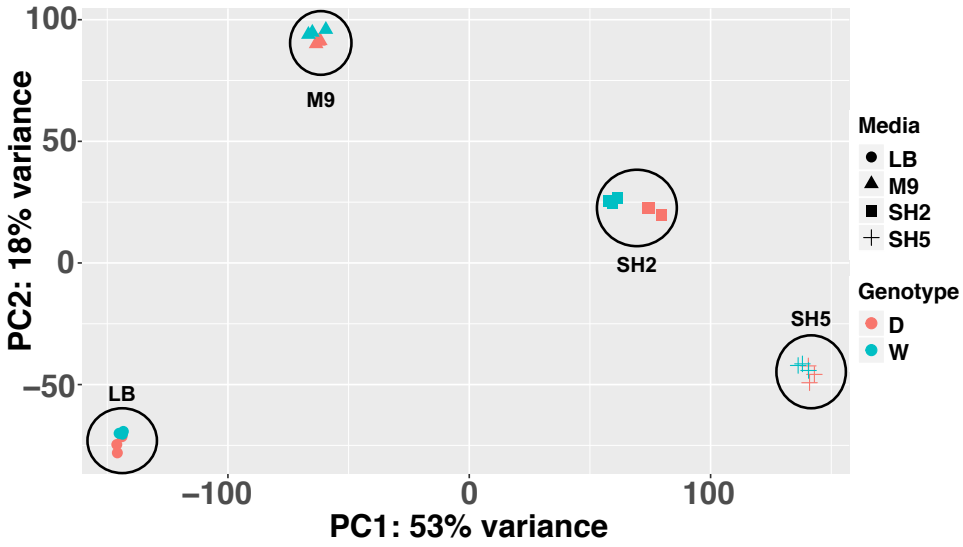


Figure 4.5: The biological replicates of the WT and Δ*spoVG* cluster differently based on the growth media used

The PCA plot of the transformed LFQ intensities of the 2,414 proteins identified in each of the strains. Genotype: W = WT; D = Δ*spoVG*

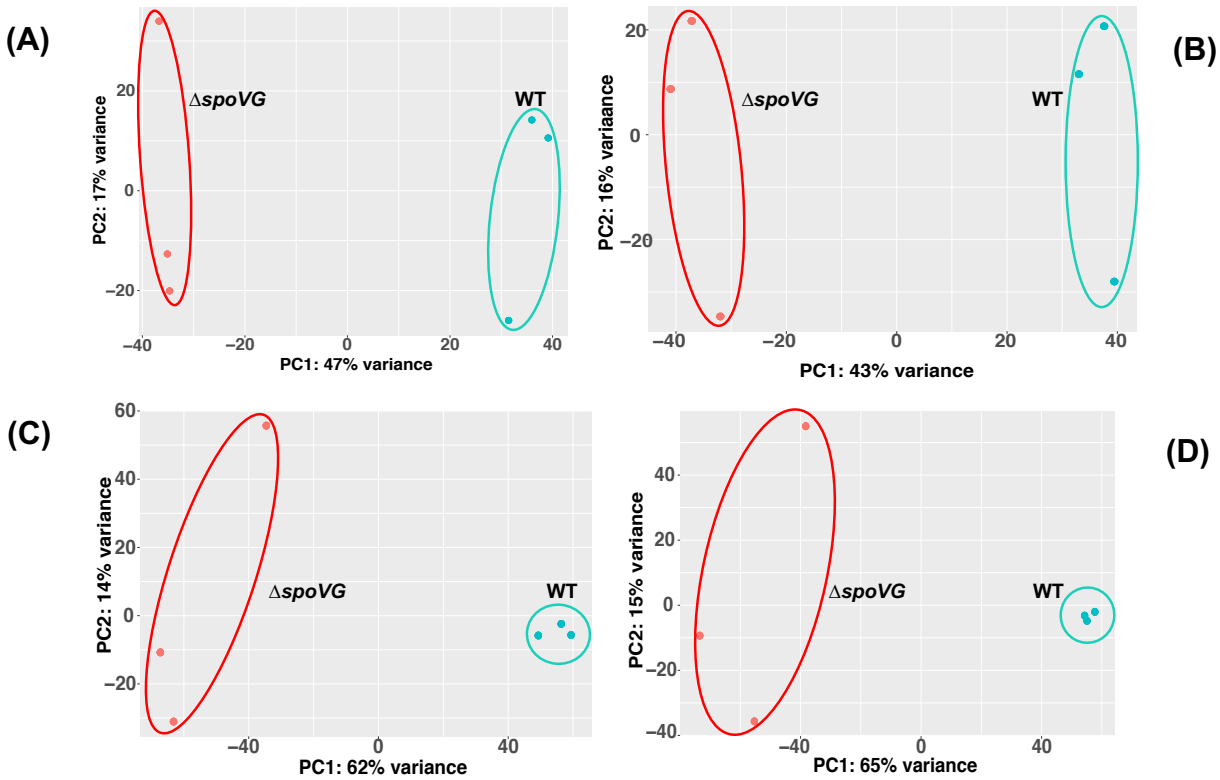


Figure 4.6: The replicates of WT and Δ*spoVG* proteome cluster separately based on their genotypes

Clustering of proteomes of the WT and the Δ*spoVG* in **(A)** LB **(B)** M9 **(C)** SH2 **(D)** SH5. The plots show the transformed LFQ intensities of the 2,414 proteins identified in each of the strains. Green dots (●) = WT; Red dots (●) = Δ*spoVG*.

4.2.4 Analysis of the Differentially Produced Proteins (DPPs)

To identify the DPPs in Δ*spoVG*, comparisons were made between the protein intensities of Δ*spoVG* vs the WT control. DPPs with adjusted *p*-value (padj) < 0.05 were considered statistically significant irrespective of their $\log_2$ fold change (logFC) values. To visualise the distribution of the proteins in each growth condition, their $-\log_{10}$(padj) were plotted against respective logFC values on Volcano plots (Figure 4.7). The analysis identified both significantly increased and decreased differentially produced proteins in Δ*spoVG* across each of the growth conditions studied.

In total 92, 110, 710 and 556 significantly differentially produced proteins were identified, which accounted for 3.8%, 4.6%, 29.4% and 23% of the 2,414 proteins identified in LB, M9, SH2 and SH5 respectively. These changes were largest for the sporulation conditions compared to LB and M9. Especially, most changes were identified in the SH2. Table 4.1 lists the number of DPPs with logFC > 2 and logFC < -2 in each growth condition.

Table 4.1: DPPs with logFC > 2 and logFC < -2 in Δ*spoVG*

Growth condition	Proteins with logFC < -2 (Decreased)	Proteins with logFC > 2 (Increased)
LB	36	33
M9	46	34
SH2	131	60
SH5	173	32

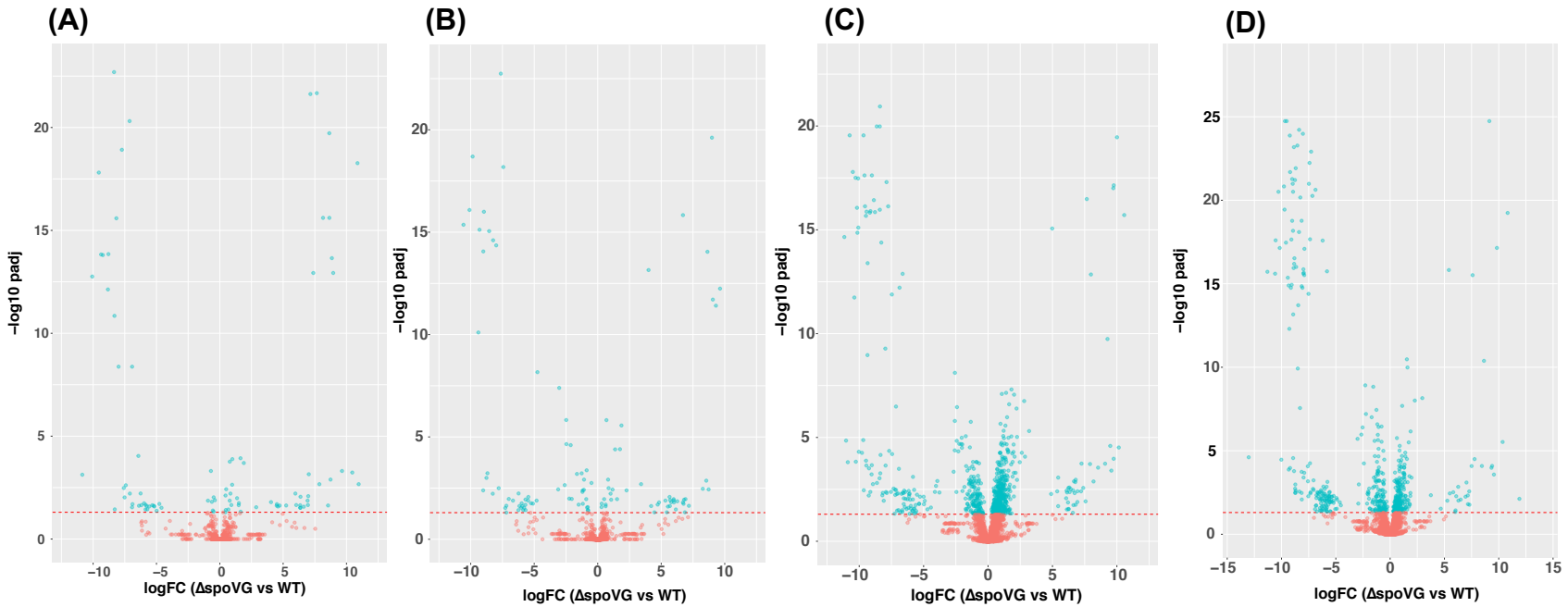


Figure 4.7: Distribution of the DPPs in all the growth conditions

The Volcano plots of -$\log_{10}$ padj values vs the logFC of the proteins identified in **(A)** LB, **(B)** M9, **(C)** SH2 and **(D)** SH5. Red dots show padj > 0.05 (non-significant); green dots show padj < 0.05 (significant).

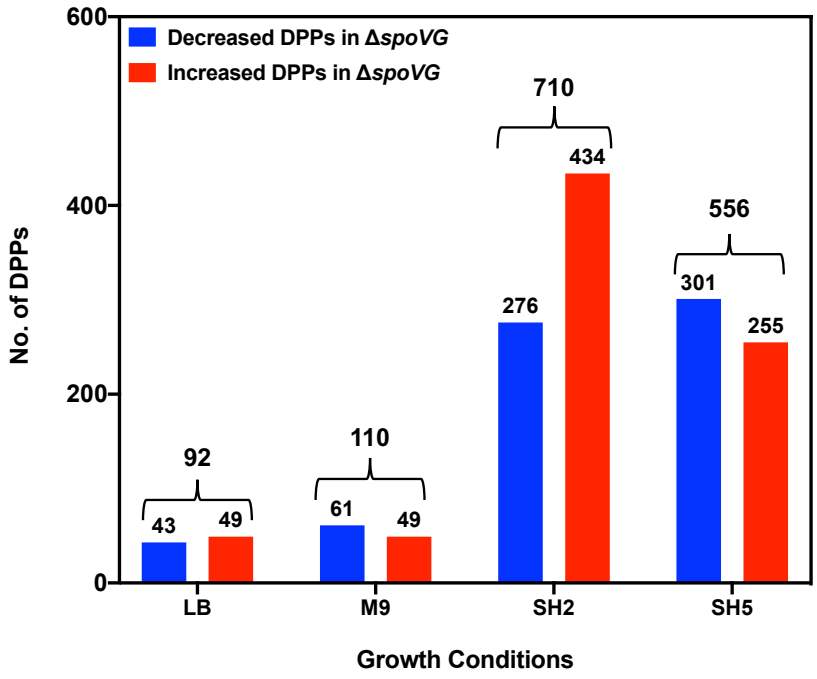


Figure 4.8: Overview of the DPPs (padj<0.05) in Δ*spoVG* in each condition investigated

4.2.4.1 DPPs in LB

In the LB condition, 92 proteins were differentially produced in Δ*spoVG,* among which 49 increased and 43 decreased in the mutant. The number of DPPs recorded in Δ*spoVG* in LB were the least amongst the rest of the growth conditions investigated in this study. Table 4.2 lists the top 10 increased and decreased proteins in the mutant in LB.

Among the most increased protein was AcoC (acetoin dehydrogenase E2 component), which increased (up to logFC 11) protein (Table 4.2 (A)). AcoC participates in the metabolism of acetoin, which is utilized as an alternate carbon source by *B. subtilis* (Ali *et al.*, 2001). Whereas, HisZ (histidyl-tRNA synthetase) was the most decreased protein in Δ*spoVG* (Table 4.2 (B)). HisZ is produced from the first gene of the *hisZGDBHAFI* operon which is involved in the biosynthesis of the amino acid histidine (Mäder *et al.*, 2002).

In addition to HisZ in this operon, only HisB was identified as a decreased DPP. Several of the increased DPPs in Δ*spoVG* also belonged to alternate carbon sources, carbon metabolism and electron transfer system.

4.2.4.2 DPPs in M9

In M9, 110 proteins were differentially produced out of which were 61 decreased and 49 increased in the mutant strain (Figure 4.8). The top 10 increased and decreased DPPs in Δ*spoVG* are listed in Table 4.3. The analysis revealed that GerR (Sporulation transcriptional regulator) was the most increased DPP which increased approx. by logFC 12.4 in Δ*spoVG* (Table 4.3 (A)). GerR is a sporulation transcription factor which is involved in the formation of spore coat and provided resistance to spores against lysozyme (Kuwana *et al.*, 2005; Cangiano *et al.*, 2010). Among the decreased DPPs, MmgC (acyl-CoA dehydrogenase) was the most DPP that decreased by approx. logFC -10.6 in the absence of SpoVG (Table 4.3 (B)). MmgC is a mother cell specific protein. MmgC is produced from the *mmgABCDEF* operon which is regulated by SigE, produce proteins similar to fatty-acid metabolising enzymes and is subjected to catabolite repression in a CcpA-dependent manner (Bryan, Beall and Moran, 1996; Steil *et al.*, 2005; Nagler *et al.*, 2016). Interestingly, no other Mmg protein was identified in the M9 data set. Many other decreased proteins belonged to antimicrobial compounds, sporulation, utilization of alternate carbon sources.

Table 4.2: Significant DPPs in LB in Δ*spoVG*

Top 10 increased **(A)** and decreased **(B)** DPPs in Δ*spoVG* in LB, where the proteins are arranged in a decreasing and ascending logFC values, respectively.

(A)

Locus Tag	Gene	Product	Description	Function	logFC	padj
BSU08080	*acoC*	acetoin dehydrogenase E2 component (dihydrolipoamide acetyltransferase)	acetoin dehydrogenase E2 component (dihydrolipoamide acetyltransferase)	acetoin utilization	10.966	2.15E-03
BSU10350	*yhfS*	unknown	similar to acetyl-CoA C-acetyltransferase	unknown	10.870	5.40E-19
BSU13819	*ykzS*	unknown	unknown	unknown	10.459	5.85E-04
BSU32770	*yusE*	unknown	similar to thioredoxin	unknown	9.649	4.92E-04
BSU05600	*ydgE*	unknown	putative N-acetyltransferase	unknown	8.954	1.17E-13
BSU15400	*ylmG*	unknown	unknown	unknown	8.849	2.23E-14
BSU31500	*yuxK*	unknown	unknown	unknown	8.741	1.29E-03
BSU39080	*licT*	transcriptional antiterminator (BglG family)	transcriptional antiterminator of the *bglP-bglH-yxiE* operon and *bglS*	control of beta-glucan and beta-glucoside utilization	8.648	2.46E-16
BSU20390	*yorG*	unknown	unknown	unknown	8.645	1.89E-20
BSU23040	*fer*	ferredoxin	ferredoxin	electron transfer	8.549	2.34E-02

(B)

Locus Tag	Gene	Product	Description	Function	logFC	padj
BSU34930	*hisZ*	histidyl-tRNA synthetase	histidyl-tRNA synthetase	translation	-10.070	1.75E-13
BSU21990	*ypdQ*	unknown	similar to RNase HI	unknown	-9.550	1.55E-18
BSU15630	*ylnF*	probably precorrin-2 dehydrogenase	probably precorrin-2 dehydrogenase	siroheme biosynthesis , sulfite reduction	-9.370	1.48E-14
BSU02990	*opuAB*	glycine betaine and arsenobetaine [ABC transporter] (permease)	glycine betaine and arsenobetaine [ABC transporter] (permease)	compatible solute transport	-9.228	1.58E-14
BSU02650	*pcp*	pyrrolidone-carboxylate peptidase	pyrrolidone-carboxylate peptidase	removal of the N-terminal pyroglutamyl group from peptides	-8.828	7.41E-13
BSU29690	*acuA*	Gcn5-related N-acetyltransferase	protein acetylase for the control of AcsA activity	control of AcsA activity	-8.791	1.42E-14
BSU35200	*yvkB*	unknown	similar to transcriptional regulator (TetR family)	unknown	-8.346	2.03E-23
BSU32610	*frlB*	fructoselysine-6-P-glycosidase	fructoselysine-6-P-glycosidase	metabolism of aminoacylated fructose	-8.314	1.42E-11
BSU31060	*gbsA*	glycine betaine-aldehyde dehydrogenase	glycine betaine-aldehyde dehydrogenase, glycine betaine synthesis	osmoprotection	-8.291	3.61E-02
BSU28800	*araA*	L-arabinose isomerase	L-arabinose isomerase	arabinose utilization	-8.164	2.58E-16

Table 4.3: Significant DPPs in M9 in Δ*spoVG*

Top 10 increased **(A)** and decreased **(B)** DPPs arranged in a decreasing and ascending logFC values, respectively.

(A)

Locus Tag	Gene	Product	Description	Function	logFC	padj
BSU15090	*gerR*	unknown	probably DNA-binding protein, regulates transcription of some spore coat genes	spore coat formation and resistance of spores to lysozyme	12.435	1.19E-09
BSU11450	*oppC*	oligopeptide ABC transporter (permease)	oligopeptide ABC transporter (permease)	initiation of sporulation, competence development	9.632	5.71E-13
BSU30940	*glgP*	glycogen phosphorylase	glycogen phosphorylase	glycogen biosynthesis	9.308	3.89E-12
BSU12610	*xkdG*	unknown	PBSX prophage	unknown	9.064	1.95E-12
BSU20390	*yorG*	unknown	unknown	unknown	8.992	2.41E-20
BSU37030	*racA*	cell division protein	cell division protein; attaches the chromosome to the cell pole at the onset of sporulation	unknown	8.733	3.65E-03
BSU19740	*yodT*	unknown	similar to adenosylmethionine-8-amino-7-oxononanoate aminotransferase	unknown	8.642	9.05E-15
BSU05610	*vmlR*	ABCF ATPase	F-type ATP-binding cassette protein, allosterically dissociates antibiotics (virginiamycin M, lincomycin) from their ribosomal binding sites	dissociation of antibiotics (virginiamycin M, lincomycin) from the ribosome	8.548	1.37E-03
BSU08140	*yfjD*	unknown	unknown	unknown	8.276	3.43E-03
BSU10810	*yisP*	farnesyl diphosphate phosphatase	farnesyl diphosphate phosphatase, production of farnesol	control of KinC activity	7.180	2.31E-02

(B)

Locus Tag	Gene	Product	Description	Function	logFC	padj
BSU24150	*mmgC*	acyl-CoA dehydrogenase	acyl-CoA dehydrogenase	mother cell metabolism	-10.568	4.33E-16
BSU17120	*pksE*	unknown	involved in polyketide synthesis	polyketide synthesis	-10.092	8.22E-17
BSU27260	*mccA*	O-acetylserine-thiol-lyase	O-acetylserine-thiol-lyase	methionine-to-cysteine conversion	-9.847	2.02E-19
BSU39680	*iolI*	inosose isomerase, converts 2KMI to 1-keto-D-chiro-inositol	inosose isomerase, converts 2KMI to 1-keto-D-chiro-inositol	myo-inositol catabolism	-9.397	7.91E-11
BSU01980	*skfH*	unknown	unknown	unknown	-9.307	7.60E-16
BSU29690	*acuA*	Gcn5-related N-acetyltransferase	protein acetylase for the control of AcsA activity	control of AcsA activity	-9.007	8.70E-15
BSU10430	*yhxD*	unknown	general stress protein, similar to alcohol dehydrogenase,	survival of salt and ethanol stresses	-8.959	1.00E-16
BSU22100	*kdgA*	2-dehydro-3-deoxy-phosphogluconate aldolase	2-dehydro-3-deoxy-phosphogluconate aldolase	utilization of galacturonic acid	-8.758	1.00E-03
BSU27930	*spo0B*	sporulation initiation phosphotransferase	sporulation initiation phosphotransferase of the phosphorelay	initiation of sporulation	-8.680	5.95E-04
BSU33490	*cadA*	cadmium transporting ATPase	cadmium transporting ATPase, resistance to cadmium	cadmium export	-8.546	8.87E-16

4.2.4.3 DPPs in SH2

Amongst all the growth conditions investigated in this study, SH2 produced the greatest number of DPPs in Δ*spoVG* (Figure 4.8). In total 710 DPPs were identified in SH2 out of which 434 were increased and 276 were decreased in the mutant (Figure 4.8). Table 4.4 lists the top 10 increased and decreased proteins in SH2. Among the increased DPPs, GlnL (two-component response regulator) was the most increased protein which increased by approx. logFC = 13.6 in the mutant (Table 4.4 (A)). GlnL is a part of the GlnK-GlnL two-component system and participates in the glutamine utilisation in *B. subtilis* (Fabret, Feher and Hoch, 1999; Satomura *et al.*, 2005). In the SH2 data, only GlnL was identified. Whereas, amongst the decreased DPPs, RpsU (ribosomal protein S21) was the most decreased proteins with logFC -12.511 in Δ*spoVG*. RpsU is the S21 subunit of the ribosomal protein which in combination with another ribosomal subunit RpsK (S11) are known to affect the cell motility and biofilm formation in *B. subtilis* (Akanuma *et al.*, 2012; Takada *et al.*, 2014). In this study, RpsK (S11) was also decreased in Δ*spoVG* in SH2 condition.

4.2.4.4 DPPs in SH5

The second largest number of DPPs were identified in SH5, where 556 DPPS were identified in Δ*spoVG* (Figure 4.8). In SH5, 255 DPPs were decreased and 301 DPPs were increased in Δ*spoVG*. The top 10 most decreased and increased DPPs in Δ*spoVG* are listed in (Table 4.5). Amongst the increased DPPs GlnL was the most increased DPP, which was similar to the observation in SH2 (Table 4.5 (A)). In SH5, GlnL was increased by approx. logFC 11.9 in the mutant, which was slightly lower to the fold change in SH2.

Table 4.4: Significant DPPs in SH2 in Δ*spoVG*

Top 10 increased **(A)** and decreased **(B)** DPPs arrange in a decreasing and ascending logFC values, respectively.

(A)

Locus Tag	Gene	Product	Description	Function	logFC	padj
BSU02450	*glnL*	two-component response regulator	two-component response regulator, regulation of the *glsA-glnT* operon	regulation of the *glsA-glnT* operon	13.645	2.61E-03
BSU24580	*yqhH*	unknown	similar to SNF2 helicase	unknown	10.572	1.95E-16
BSU17140	*pksF*	unknown	involved in polyketide synthesis	polyketide synthesis	10.150	3.05E-05
BSU24430	*spoIIIAA*	AAA protease	AAA protease, required for SigG activation	activation of SigG	10.010	3.49E-20
BSU19689	*yokU*	unknown	unknown	unknown	9.758	7.21E-18
BSU31500	*yuxK*	unknown	unknown	unknown	9.750	1.06E-04
BSU10430	*yhxD*	unknown	general stress protein, similar to alcohol dehydrogenase,	survival of salt and ethanol stresses	9.723	9.92E-18
BSU06230	*iolT*	major transporter for inositol	major transporter for inositol	myo-inositol uptake	9.572	3.95E-04
BSU11500	*spx*	transcriptional regulator Spx	transcriptional regulator Spx, involved in regulation of many genes, important for the prevention of protein aggregation during severe heat stress, required for protection against paraquat stress	negative and positive regulator of many genes	9.488	2.56E-05
BSU19730	*yodS*	unknown	similar to butyrate-acetoacetate CoA-transferase	unknown	9.274	1.81E-10

(B)

Locus Tag	Gene	Product	Description	Function	logFC	padj
BSU25410	*rpsU*	ribosomal protein S21	ribosomal protein	translation	-12.511	7.40E-17
BSU22010	*ypcP*	5'→3' exonuclease	5' → 3' exonuclease	protection of developing and dormant spores against UV DNA damage	-11.152	2.23E-15
BSU39430	*deoR*	transcriptional repressor	transcriptional repressor of the *dra-nupC-pdp* operon	regulation of deoxyribonucleotide utilization	-11.019	1.42E-05
BSU35410	*flgK*	flagellar hook-filament junction proteins	flagellar hook-filament junction proteins	motility and chemotaxis	-10.884	1.55E-04
BSU05810	*gmuB*	glucomannan-specific lichenan-specific Phosphotransferase system (PTS), EIIB component	glucomannan-specific permease of the PTS, EIIB of the PTS	glucomannan uptake and phosphorylation	-10.739	2.83E-20
BSU11660	*tenI*	thiazole tautomerase	thiazole tautomerase	biosynthesis of thiamine	-10.492	1.63E-18
BSU14150	*ykuN*	flavodoxin	flavodoxin, binds FMN, replaces ferredoxin under conditions of iron limitation, probably involved in electron transfer to nitric oxide synthase	electron transfer	-10.377	1.82E-12
BSU10310	*yhfO*	unknown	predicted acyltransferase	unknown	-10.285	3.13E-18
BSU04410	*ydbB*	unknown	unknown	unknown	-10.258	1.47E-04
BSU29580	*thiI*	sulfuryl transferase	sulfuryl transferase, biosynthesis of 4-thiouridine in tRNA	biosynthesis of 4-thiouridine in tRNA	-10.170	8.71E-17

Whereas, the most decreased DPP in SH5 was RpsU which was also the most decreased DPP in SH2 (Table 4.5 (A)). In SH5, RpsU was decreased by approx. logFC -12.5, which was close to half of the production in SH2.

The variation in the changes across different conditions indicated a diverse role of SpoVG in each of the growth conditions sampled. To determine if any DPPs were commonly changed in any of the four growth conditions, an overlap of all the DPPs was made. The results identified four DPPs (viz. Prs, GcaD, IolD and DhaS) common to all the growth conditions (Figure 4.9). It was observed that, the production of GcaD (bifunctional N-acetylglucosamine-1-phosphate) and Prs (phosphoribosylpyrophosphate synthetase) was increased in Δ*spoVG* across all the growth conditions (Table 4.6). Whereas the production of DhaS (aldehyde dehydrogenase (NAD)) was decreased in Δ*spoVG* in all the conditions (Table 4.6). The level of IolD (formation of 5-deoxy-D-glucuronic acid) was increased in all the conditions except in LB (Table 4.6).

GcaD and Prs are produced from the operon *gcaD-prs-ctc* in vegetatively growing cells (Hilden, Krath and Hove-Jensen, 1995), and are essential for the survival of *B. subtilis* (Kobayashi *et al.*, 2003). GcaD and Prs are also a part of the stringent response regulon (Eymann *et al.*, 2002). The operon *gcaD-prs-ctc* is located immediately downstream to the gene *spoVG* (Nicolas *et al.*, 2012). The protein Ctc which is encoded by the last gene of the operon was identified as a DPP on in LB. The protein DhaS is encoded from an independent gene *dhaS* and participates in glycerol metabolism where it converts 3-hydroxypropionaldehyde (3-HPA) into 3-hydroxypropionic acid (3-HP) (Su *et al.*, 2015). Whereas, IolD is encoded by the *iolD* gene which is a part of the *iolABCDEFGHIJ* operon (Yoshida *et al.*, 1997). IolD functions in the myo-inositol catabolism pathway by catalysing the third reaction of hydrolysing 3D-(3,5/4)-trihydroxycyclohexane-1,2-dione to 5-deoxy-D-glucuronic acid (Yoshida *et al.*, 2008).

Table 4.5: Significant DPPs in SH5 in Δ*spoVG*

Top 10 increased **(A)** and decreased **(B)** DPPs arrange in a decreasing and ascending logFC values, respectively.

(A)

Locus Tag	Gene	Product	Description	Function	logFC	padj
BSU02450	*glnL*	two-component response regulator	two-component response regulator, regulation of the *glsA-glnT* operon	regulation of the *glsA-glnT* operon	11.905	7.57E-03
BSU17080	*pksA*	transcriptional regulator (TetR family)	transcriptional regulator (TetR family), not involved in the regulation of the pks operon!	unknown	10.820	5.75E-20
BSU35030	*nagR*	transcriptional regulator (GntR family)	transcriptional regulator (GntR family)	regulator of the *nagA-nagB-nagR* operon	9.832	7.15E-18
BSU02290	*psd*	phosphatidylserine decarboxylase	phosphatidylserine decarboxylase	biosynthesis of phospholipids	9.557	2.66E-04
BSU34700	*psdA*	ABC transporter (ATP-binding protein)	ABC transporter (ATP-binding protein) for the export of lipid II-binding lantibiotics, such as nisin and gallidermin	export of toxic peptides	9.390	8.06E-05
BSU22510	*ypjC*	unknown	unknown	unknown	9.331	1.05E-04
BSU24730	*comGA*	traffic ATPase	late competence protein, traffic ATPase, binds and transports transforming DNA, required for growth arrest of competent cells	genetic competence	9.135	1.83E-25
BSU34800	*cwlO*	endopeptidase-type autolysin	D,L-endopeptidase-type autolysin, primary autolytic pathway for cell elongation	cell wall synthesis, cell elongation	8.641	4.13E-11
BSU06970	*rhiL*	rhamnose oligosaccharide ABC transporter	ABC transporter (sugar-binding protein) for rhamnose oligosaccharides	uptake of rhamnose oligosaccharides	8.449	8.16E-05

(B)

Locus Tag	Gene	Product	Description	Function	logFC	padj
BSU25410	*rpsU*	ribosomal protein S21	ribosomal protein	translation	-11.298	1.91E-16
BSU35750	*tagA*	UDP-N-acetyl-D-mannosamine transferase	UDP-N-acetyl-D-mannosamine transferase	biosynthesis of teichoic acid	-10.590	2.50E-16
BSU02600	*cwlJ*	spore coat protein, cell wall hydrolase	spore coat protein, cell wall hydrolase	spore germination	-10.543	2.51E-18
BSU26950	*yraG*	unknown	forespore-specific sporulation protein, similar to spore coat protein	unknown	-10.284	3.12E-21
BSU18410	*ggt*	gamma-glutamyltransferase	gamma-glutamyltransferase	degradation of poly-glutamate capsules	-10.162	7.21E-18
BSU39430	*deoR*	transcriptional repressor	transcriptional repressor of the *dra-nupC-pdp* operon	regulation of deoxyribonucleotide utilization	-10.016	3.43E-05
BSU36110	*chrS*	transcriptional repressor (Lrp family)	transcriptional repressor (Lrp family) of the *chrS-ywrB-ywrA* operon	regulation of chromate export	-9.775	1.49E-21
BSU05810	*gmuB*	glucomannan-specific lichenan-specific PTS, EIIB component	glucomannan-specific permease of the PTS, EIIB of the PTS	glucomannan uptake and phosphorylation	-9.722	3.64E-20
BSU17990	*sspO*	small acid-soluble spore protein (minor)	small acid-soluble spore protein (minor)	protection of spore DNA	-9.684	1.83E-25
BSU10310	*yhfO*	unknown	predicted acyltransferase	unknown	-9.589	3.46E-18

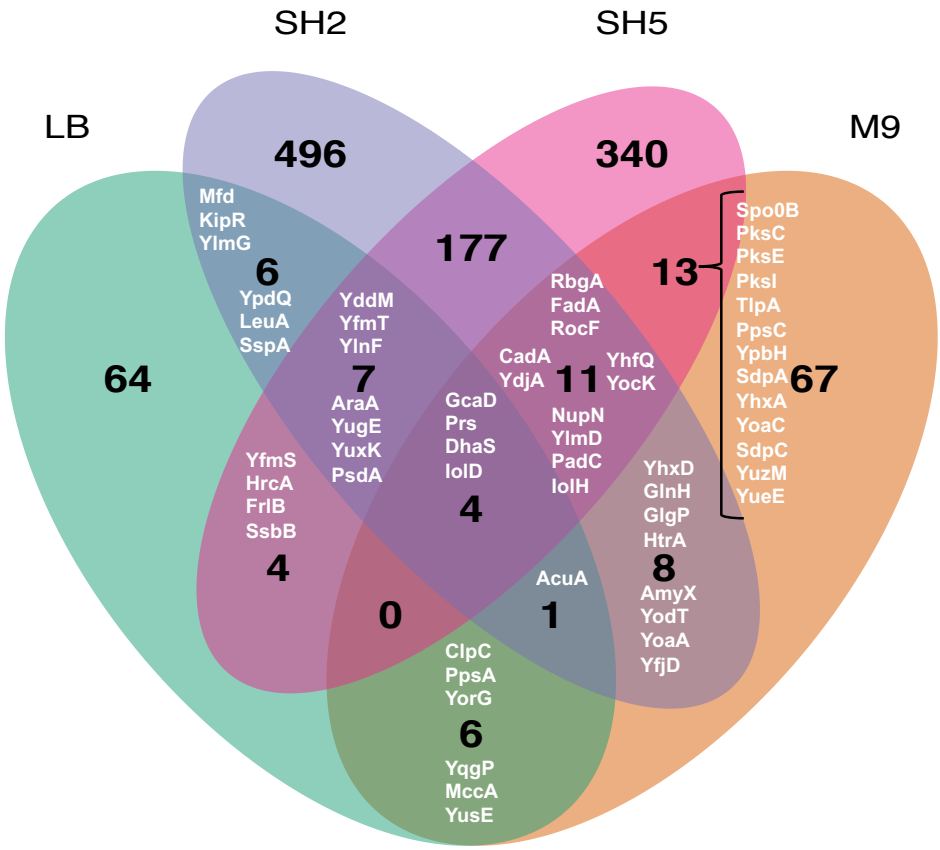


Figure 4.9: The Δ*spoVG* common DPPs across all the growth conditions

A Venn diagram of all the significantly DPPs in all the growth conditions.

Among the other unique overlaps, maximal overlap was observed between the sporulation conditions SH2 and SH5 (Figure 4.9). In total, 177 DPPs were uniquely common between SH2 and SH5. Among these, 16 proteins were essential for the survival of *B. subtilis* and belonged to processes such as biosynthesis of biotin, translation, biosynthesis of teichoic acid, fatty acid biosynthesis and spore coat formation. The function of the other common proteins between SH2 and SH5 ranged from acquisition of iron, DNA repair, biosynthesis of amino acids and co-factors, regulation of cell-wall metabolism and chemotaxis to utilization and uptake of alternate carbon sources. Most of the 177 common proteins in SH2 and SH5 showed a similar trend in their logFC values. However, many others changed in

Δ*spoVG* oppositely in SH2 and SH5 suggesting a potentially varied regulation in response to absence of SpoVG in the cells.

The second greatest number of unique overlaps were observed between M9 and SH5 with 13 DPPs common between them (Figure 4.9). Among the common overlaps between M9 and SH5 were the proteins associated with biosynthesis of antimicrobial compounds, toxins and chemotaxis (Appendix Table S.1). Most of these proteins had a similar trend of logFC values, suggesting a similar change in pathways in both M9 and SH5 in response to Δ*spoVG*. Additionally, 6 DPPs were uniquely common between LB and M9 and LB and SH5 (Appendix. Table S.3 and 4).

4.2.5 Gene Set Enrichment Analysis

To better understand the importance of changes in the proteome of Δ*spoVG* in different growth conditions and identify the biological processes affected due to the mutation, a Gene Set Enrichment Analysis (GSEA) was carried out. The GSEA was based on the *B. subtilis* database – SubtiWiki (Mäder *et al.*, 2012). In SubtiWiki, each gene or the gene product is assigned to one or more functional categories (geneCategories), which provides the overview of the functionally related genes or proteins (Mäder *et al.*, 2012). Each gene or its product, are assigned to four to five functional categories which present a broad classification of function to a very specific function, starting from Category 1 to Category 4. For this study, the Category 3 (Cat3) of the SubtiWiki classification system was chosen. The gene sets enriched with padj < 0.05 were considered statistically significantly enriched.

The results from the GSEA identified significantly enriched gene sets for each of the growth conditions (Figure 4.10). In LB, three gene sets were significantly enriched which belonged to the PBSX prophage, utilization of specific carbon sources and heat shock proteins (Figure 4.10). Three gene sets were also enriched in M9 which belonged to miscellaneous metabolic pathways, *Biosynthesis of antibacterial compounds* and *General stress proteins (controlled by SigB)*.

Table 4.6: Common DPPs in all the growth conditions examined in this study

The proteins are arranged in the ascending order of their location in the chromosome.

Locus tag	Gene	Product	LB		M9		SH2		SH5	
			logFC	padj	logFC	padj	logFC	padj	logFC	padj
BSU00500	*gcaD*	bifunctional N-acetylglucosamine-1-phosphate	0.97	2.29E-03	1.37	4.05E-05	2.02	8.70E-08	2.27	9.84E-09
BSU00510	*prs*	phosphoribosylpyrophosphate synthetase	1.63	1.18E-04	1.75	3.98E-05	1.41	1.51E-04	2.96	7.06E-09
BSU19310	*dhaS*	aldehyde dehydrogenase (NAD)	-6.92	4.21E-09	-2.59	2.26E-03	-2.66	4.49E-04	-2.08	3.74E-03
BSU39730	*iolD*	formation of 5-deoxy-D-glucuronic acid (3rd reaction)	-0.93	1.36E-02	1.01	5.93E-03	1.49	3.74E-05	0.96	2.47E-03

Table 4.7: Commonly changed DPPs with opposite log fold change (logFC) values in SH2 vs SH5

The proteins are organised in descending order of logFC of SH2.

Locus tag	Gene	Product	Function	SH2		SH5	
				logFC	padj	logFC	padj
BSU17330	*miaA*	tRNA isopentenylpyrophosphate transferase	tRNA modification	6.796	1.75E-02	-6.470	2.71E-02
BSU03200	*putB*	proline dehydrogenase (L-proline, NAD)	proline utilization	1.074	1.23E-03	-8.648	9.83E-17
BSU00150	*dgk*	deoxyguanosine kinase	purine salvage and interconversion	0.874	1.68E-02	-9.417	4.36E-16
BSU05390	*ydfF*	unknown	unknown	0.758	4.50E-02	-9.137	1.75E-15
BSU09710	*bmrC*	multidrug ABC transporter (ATP-binding protein)	multiple antibiotic resistance	0.516	3.65E-02	-9.090	2.24E-18
BSU38310	*ywbI*	transcriptional regulator (LysR family)	control of expression of the *ywbH-ywbG* operon	0.427	2.50E-02	-8.150	1.69E-19
BSU40950	*yyaC*	sporulation protein	unknown	-7.747	7.40E-17	1.008	5.00E-04
BSU06370	*pbuG*	hypoxanthin/ guanine permease	hypoxanthine and guanine uptake	-9.663	2.83E-20	0.498	1.09E-02

In the sporulation conditions, only one gene set was enriched in each of the conditions which belonged to translation in SH2 and sporulation proteins in SH5 (Figure 4.10). No gene sets were commonly enriched across all of the growth conditions. Table 4.8 lists the details of the enriched gene sets in each condition. The variety of changes observed thus far indicated a pleiotropic role of the SpoVG in the physiology of *B. subtilis*. The following sections will discuss these changes in detail.

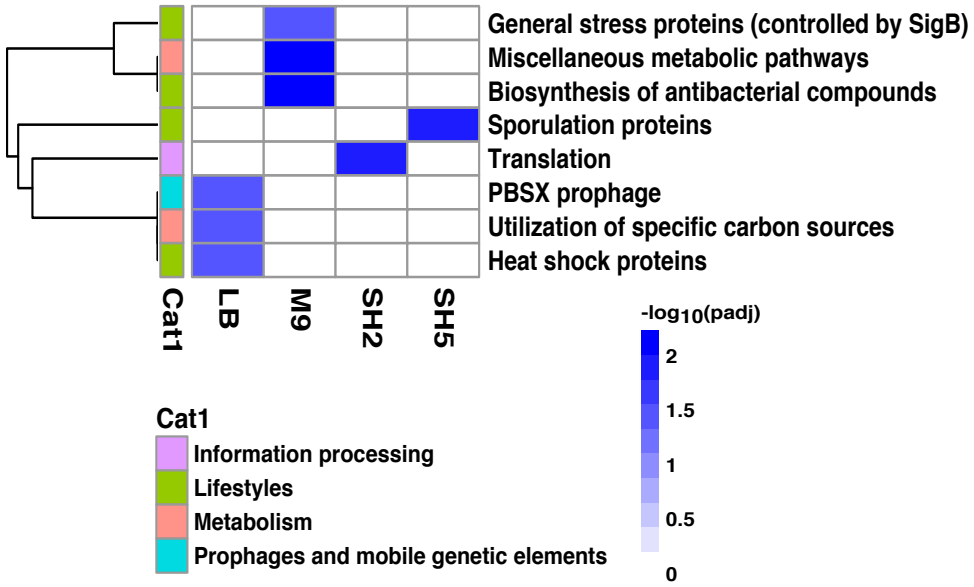


Figure 4.10: GSEA of the significantly DPPs in Δ*spoVG*

The significantly (adjusted *p*-value < 0.05) enriched gene sets are shown in blue shading. The shading indicates values of $-\log_{10}(padj) > 0$, where padj = 0.05. The $-\log_{10}(padj) = 0$ (white) indicates non-significant gene set and $-\log_{10}(padj) > 2$ (deepest shade) imply the most significantly enriched gene get.

4.2.5.1 Utilization of specific carbon sources

GSEA in LB identified Utilization of specific carbon sources as one of the significantly enriched gene sets in Δ*spoVG* (Figure 4.10). 154 proteins were associated with this gene set of which 49 proteins were increased and 69

were decreased in the mutant (Table 4.8). Some of the most affected DPPs in this gene set were the Aco proteins (AcoB and AcoC) and AraA.

In *B. subtilis*, AcoB and AcoC are encoded from the *acoABCL* operon and are involved in the metabolism of the four carbon compound acetoin as a part of the central carbon metabolism (Ali *et al.*, 2001). Acetoin is a major catabolic product of *B. subtilis* which is excreted out of the cells during aerobic growth in glucose and can serve as a carbon source for the bacteria (Ali *et al.*, 2001). Acetoin can also be produced as a result of simultaneous consumption of glucose, xylose and arabinose (Zhang *et al.*, 2016). Acetoin are transported back to the cell and further metabolised when the sugars are used up (Speck and Freese, 1973; Ali *et al.*, 2001). Interestingly, AcoB and AcoC were also one of the most increased DPPs in Δ*spoVG* in LB (Table 3.3 (A)). AraA is the first protein of the L-arabinose (*ara*) operon which is involved in the utilization of arabinose (alternate carbon source) in *B. subtilis* (Sd-nlogueira *et al.*, 1997). *B. subtilis* can grow on L-arabinose as the sole carbon and energy source (Mota, Tavares and Sá-Noguelra, 1999; Mota, Sarmento and Sa-Nogueira, 2001; Zhang *et al.*, 2016). In this study, AraA decreased by logFC ~ -8.2 in Δ*spoVG* in LB (Table 3.3 (B)).

Among the other increased proteins in this gene set was the transcriptional antiterminator LicT (Table 3.3 (A)). LicT has an important regulatory role in the utilization of β-glucan (alternate carbon source) in *B. subtilis* (Schnetz *et al.*, 1996). The expression of *licT* is regulated by CcpA, which is a global regulator of the carbon catabolite repression (CCR) and catabolite activation (Krüger, Gertz and Hecker, 1996; Lorca *et al.*, 2005). Whereas, IolD which is involved in the myo-inositol catabolism was also decreased in Δ*spoVG* in LB (Yoshida *et al.*, 1997, 1999, 2008).

Table 4.8: Significantly enriched Category 3 (SubtiWiki) gene sets identified in Δ*spoVG* from the proteomic data

The gene sets are arranged from most enriched gene set to the least for each growth condition.

Cat3 Gene Set	Total proteins identified in the gene set	Enrichment of the gene set (padj)	Proteins increased	Proteins decreased
LB				
Utilization of specific carbon sources	154	2.96E-02	49	69
Heat shock proteins	25	2.96E-02	21	4
PBSX prophage	7	2.96E-02	5	1
M9				
Miscellaneous metabolic pathways	58	5.55E-03	14	35
Biosynthesis of antibacterial compounds	34	5.55E-03	2	29
General stress proteins (controlled by SigB)	94	4.07E-02	20	60
SH2				
Translation	183	1.11E-02	100	77
SH5				
Sporulation proteins	260	1.11E-02	69	186

Carbon serves as the primary source of energy in all the organisms. For *B. subtilis*, glucose is the most preferred source of carbon whereas it can utilise secondary carbon sources e.g., malate, in glucose limiting conditions (Buescher *et al.*, 2012). While utilising glucose, *B. subtilis* represses the utilisation of secondary carbon sources, a mechanism which is commonly known as carbon catabolite repression (CCR). Since *B. subtilis* is a soil dwelling bacterium, it secretes a variety of enzymes that degrade complex sugars from plants which are further metabolised by the bacteria as alternate carbon sources.

Investigations in *L. monocytogenes* and *S. aureus* have also identified changes in the expression of genes and proteins of central carbon metabolism when *spoVG* was deleted (Bischoff *et al.*, 2016; Burke and Portnoy, 2016). Enrichment of the gene set utilization of specific carbon sources in the absence of SpoVG this study suggested a role of SpoVG in the carbon metabolism of *B. subtilis*. Targeted studies would be needed to identify such mechanisms in detail.

4.2.5.2 Biosynthesis of antibacterial compounds

From the GSEA in M9, the *Biosynthesis of antibacterial compounds* was one of the top significantly enriched gene sets (Figure 4.10). Most of the proteins identified in this gene set had a decreased production in Δ*spoVG*. Some of the most decreased proteins in this gene set were associated with the polyketide synthase (*pks*) operon, the sporulation killing factor (*skf*) operon and the plipastatin synthetase (*ppsABCDE*) operon (Table 4.9). Among the 29 decreased proteins in this gene set, the proteins PksE and SkfH decreased the most in Δ*spoVG*. PksE and SkfH were also identified as some of the most decreased proteins in Δ*spoVG* in M9 (Table 4.3 (B)).

The *pks, skf* and *ppsABCDE* operon encode for a non-ribosomal peptides (NRPs) that are translated independently of the ribosomal translational machinery (Tosato *et al.*, 1997; González-Pastor, Hobbs and Losick, 2003; Allenby *et al.*, 2006; Straight *et al.*, 2007; Tanovic *et al.*, 2008). Instead, they

are translated by synthases and produce peptides of smaller lengths. The resulting peptides have a range of properties such as antibiotics, pigments, siderophores or toxins (Bushley and Turgeon, 2010; Wang *et al.*, 2014; Martínez-Núñez and López, 2016). The decreased levels of the *pks, skf* and *ppsABCDE* operon proteins were in agreement with the previous observations from the transcriptomic data where the expressions of the *pks*, *skf* and *pps* genes were down-regulated in Δ*spoVG* in M9 (Chapter 3). The similar direction of change in these proteins indicated a role of SpoVG in the production of antimicrobial compounds in *B. subtilis*. In addition to the antimicrobial compounds, decrease in the levels of toxin proteins such as the SdpABC (logFC -3.0, -7.2, -5.2 respectively) and the YobL (logFC = -7.6) were also observed in the mutant in M9 (Holberger *et al.*, 2012).

The antimicrobial proteins are also known to be produced by the post exponential-phase cells. Also, in *B. subtilis* the biosynthesis of antimicrobial compounds has been linked to sporulation, competence and biofilm formation (Strauch *et al.*, 2007; Schultz *et al.*, 2009; Vlamakis *et al.*, 2013). When cells face extreme nutritional starvations and have exhausted all their strategies to escape these stresses, they face a difficult decision to whether sporulate or continue to grow. The cells that choose not to sporulate, increase their competence and production of antimicrobial compounds. Increased competence helps them to gain genes from their surroundings via horizontal gene transfer. This helps them to increase their fitness and production of antimicrobial compounds let them fight against the toxins from the sporulating cells and compete for resources from the surrounding amongst the other vegetative. Whereas, the sporulating cells decrease their competence and increase their toxin (skf) productions to kill the non-sporulating cells to gain energy out of the dying cells (Strauch *et al.*, 2007; Schultz *et al.*, 2009; Vlamakis *et al.*, 2013).

Table 4.9: Significantly enriched antimicrobial proteins in Δ*spoVG* in M9

Genes are arranged in ascending order of their genomic loci. Only the proteins with logFC < -1 are shown here.

Locus Tag	Gene	Product	Description	Function	logFC	padj
			***skf* operon**			
BSU01980	*skfH*	unknown	unknown	unknown	-9.307	7.60E-16
			***pks* operon**			
BSU17100	*pksC*	bacillaene synthase trans-acting acyltransferase	bacillaene synthase trans-acting acyltransferase	polyketide (bacillaene) synthesis	-1.634	6.45E-04
BSU17120	*pksE*	unknown	involved in polyketide synthesis	polyketide synthesis	-10.092	8.22E-17
BSU17170	*pksI*	unknown	involved in polyketide synthesis	polyketide synthesis	-8.262	5.94E-03
BSU17190	*pksL*	polyketide synthase of type I	polyketide synthase of type I	polyketide synthesis	-2.467	2.24E-05
			***ppsABCDE* operon**			
BSU18320	*ppsC*	plipastatin synthetase	plipastatin synthetase	production of the antibacterial compound plipastatin	-2.126	2.51E-05
BSU18340	*ppsA*	plipastatin synthetase	plipastatin synthetase	production of the antibacterial compound plipastatin	-2.044	1.07E-02

In addition to the changed levels of antimicrobial compounds, the proteomic analysis of Δ*spoVG* in M9 also identified proteins such as GerR, OppC and Spo0B as some of the DPPs (Table 4.3). The protein Spo0B (decreased) and GerR (increased) are sporulation proteins in *B. subtilis*. While, Spo0B is important for the initiation of sporulation (Burbulys, Trach and Hoch, 1991), GerR (a mother cell specific sporulation transcription factor) regulates the transcription of some spore coat genes and confers resistance to the spores against the hydrolytic enzyme lysozyme (Kuwana *et al.*, 2005; Cangiano *et al.*, 2010). OppC (increased) belongs to the oligopeptide transport (Opp) system encoded by the *oppABCDF* operon which is also responsible for the initiation of sporulation and competence in *B. subtilis* (Perego *et al.*, 1991; Solomon *et al.*, 2003). Although the cells were sampled in the M9 conditions and it was expected that some cells would sporulate in the nutritionally stressed cells. Based on the changes in the levels of sporulation proteins in M9 observed in this study and the results obtained in Chapter 3, it is speculated that SpoVG is involved in the complex regulation of antimicrobial activity, competence and establishing sporulation in *B. subtilis*. Although the mode of these regulation is still elusive.

4.2.5.3 General stress proteins (controlled by SigB)

The GSEA in M9 also identified *General stress proteins (controlled by SigB)* as one of the significantly enriched gene set in Δ*spoVG* (Figure 4.10). 94 proteins were detected in this gene set with most of the proteins decreasing in the mutant (Table 4.8). As *B. subtilis* is a soil dwelling bacterium, it is constantly challenged for food and by a variety of environmental stress such as heat, salt, ethanol, paraquat or changes in the membrane fluidity (Hecker and Völker, 2001; Price, 2002). To overcome these stresses, *B. subtilis* has developed a general stress response (GSR) mechanism. Most of the GSR genes are regulated by the sigma factor SigB (Kalman *et al.*, 1990; Boylan *et al.*, 1993). The SigB controlled GSR mechanism is complex and SigB regulates more than 200 genes (Helmann *et al.*, 2001; Petersohn *et al.*,

2001; Price *et al.*, 2001; Hecker, Pané-Farré and Völker, 2007; Nannapaneni *et al.*, 2012; Van Der Steen and Hellingwerf, 2015).

In the transcriptomic study, most of the significantly differentially expressed genes (DEGs) in the *General stress proteins (controlled by SigB)* gene set were up-regulated in Δ*spoVG* in M9 (Chapter 3). It was also identified that the expression of *sigB* significantly increased in Δ*spoVG* in M9 (Chapter 3). However, in this proteomic study of Δ*spoVG* in M9, most of the general stress proteins (GSPs) decreased in the mutant (Table 4.8). Also, the sigma factor SigB was not identified as a DPP in M9. Nonetheless, a range of SigB regulon proteins were affected in the mutant in M9, which included both the increased and decreased proteins in Δ*spoVG*.

The decreased levels of several GSPs in this proteomic study compared to the up-regulated GSR genes in the transcriptomic analysis of Δ*spoVG* in M9 indicated a possible post-transcriptional gene regulation of the GSR genes via SpoVG. Nonetheless, the mechanism of these changes remain elusive and further studies would be needed to understand these changes in detail.

4.2.5.4 Translation

Translation was the only significantly enriched gene set in Δ*spoVG* in SH2 (Figure 4.10). In this gene set, a total of 183 proteins were detected most of which were increased in Δ*spoVG* (Table 4.8). Some of the increased DPPs in this gene set comprised of ribosomal proteins and the non-ribosomal translational proteins such as tRNA synthases, transferases, and elongation factors (Table 4.4). Whereas, most of the decreased DPPs comprised of ribosomal proteins which belonged to both the large (50S) and the small (30S) subunit of ribosome (Table 4.4). Similar changes were also observed in the mutant in SH5.

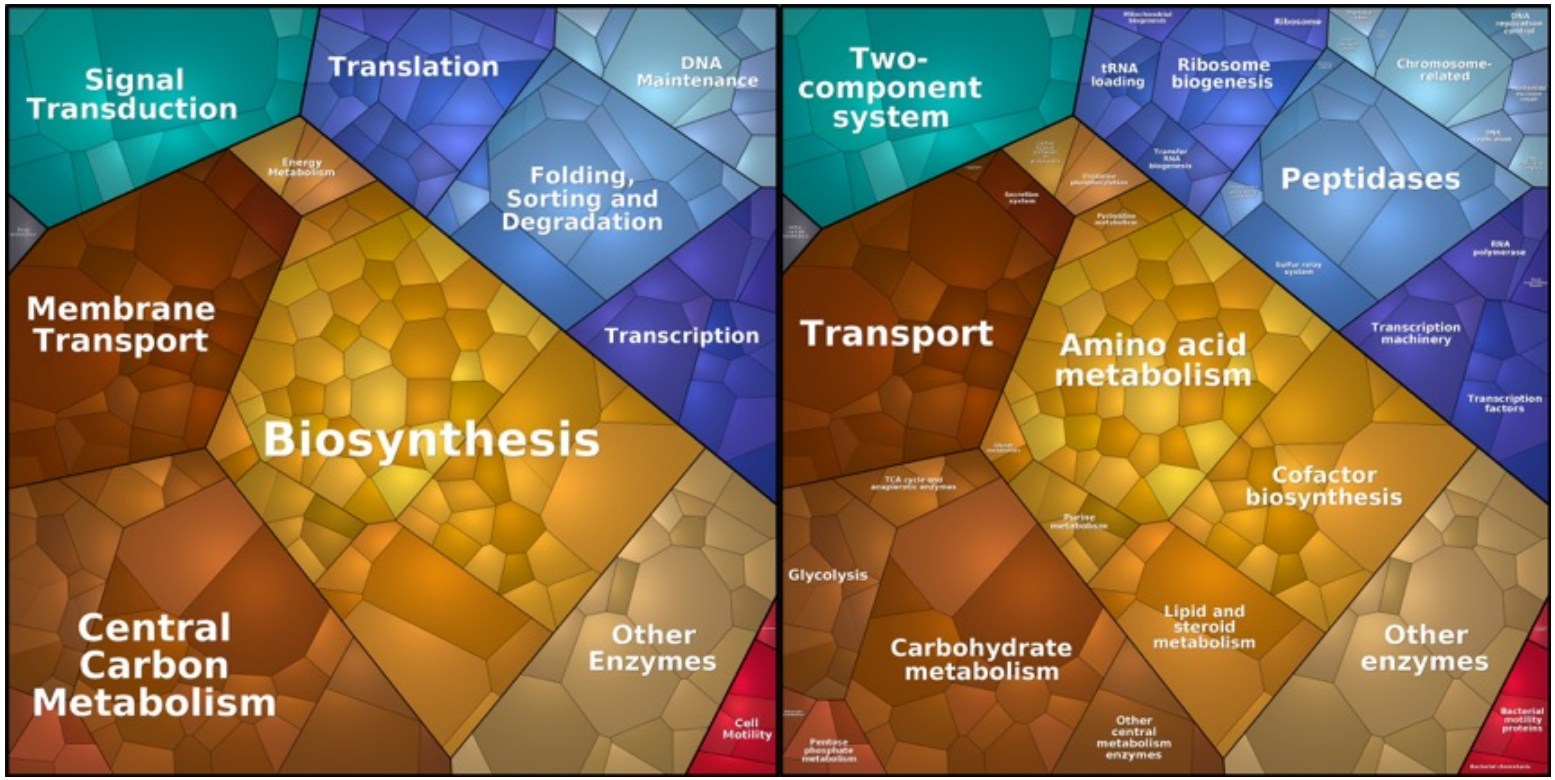
The ribosome is a macromolecular assembly of different ribosomal RNAs (rRNAs) and ribosomal proteins. The ribosomal proteins play an essential role in the process of translation and ribosome assembly (Cukras *et al.*, 2003; Diaconu *et al.*, 2005; Maguire *et al.*, 2005; Fei *et al.*, 2008; Zaher and

Green, 2010). In *B. subtilis*, in addition to these roles, ribosomes have also been studied in cellular differentiation and proliferation, different growth phases, antimicrobial resistance, RNA processing and cellular motility (Guha, Roth and Nierhaus, 1975; Ohashi *et al.*, 2003; Nanamiya and Kawamuray, 2010; Akanuma *et al.*, 2012; Baumgardt *et al.*, 2018; Crowe-McAuliffe *et al.*, 2018; Hummels and Kearns, 2019). The identification of the differentially produced ribosomal proteins in the absence of SpoVG in this study indicated a potential role of SpoVG in the regulation of ribosomal proteins in *B. subtilis*.

Hfq – the most well studied RNA-binding protein (RBP) in *E. coli* also associates with ribosomes and DNA (Kajitani *et al.*, 1994; Tsui, Leung and Winkler, 1994; Takada *et al.*, 1997). Interestingly, it was initially identified to be associated with ribosomes along with sRNAs while modulating the translational efficiency of mRNA transcripts (Kajitani *et al.*, 1994). Additionally, recent investigation have also identified Hfq as a crucial factor in ribosome biogenesis and translation (Andrade *et al.*, 2018). Hfq also participates in protein-protein interactions with ribosomal proteins (Caillet *et al.*, 2019); and its expression corelates with the growth of *E. coli*, highlighting the emerging new roles of RBPs in bacterial system (VO *et al.*, 2021).

Even though the roles and functions of SpoVG in *B. subtilis* largely remain unknown, from the significant differential production of the ribosomal proteins in Δ*spoVG* identified in this study along with the evidence from roles of Hfq in *E. coli* in ribosomal biogenesis and translation, it could be possible that SpoVG might have a function in the ribosome assembly and biogenesis in *B. subtilis* as well. Nevertheless, such generalisation would not be accurate without further investigation of SpoVG in such processes. Targeted experiments such as measuring the actual amounts of ribosomal gene products in Δ*spoVG* during sporulation conditions maybe helpful in highlighting its role in translation.

(A)


Signal Transduction
Translation
DNA Maintenance
Energy Metabolism
Folding, Sorting and Degradation
Membrane Transport
Transcription
Biosynthesis
Central Carbon Metabolism
Other Enzymes
Cell Motility
Two-component system
tRNA loading
Ribosome biogenesis
Chromosome-related
Peptidases
RNA polymerase
Transport
Amino acid metabolism
Transcription machinery
Transcription factors
Cofactor biosynthesis
Glycolysis
Lipid and steroid metabolism
Carbohydrate metabolism
Other enzymes
Other central metabolism enzymes

(B)

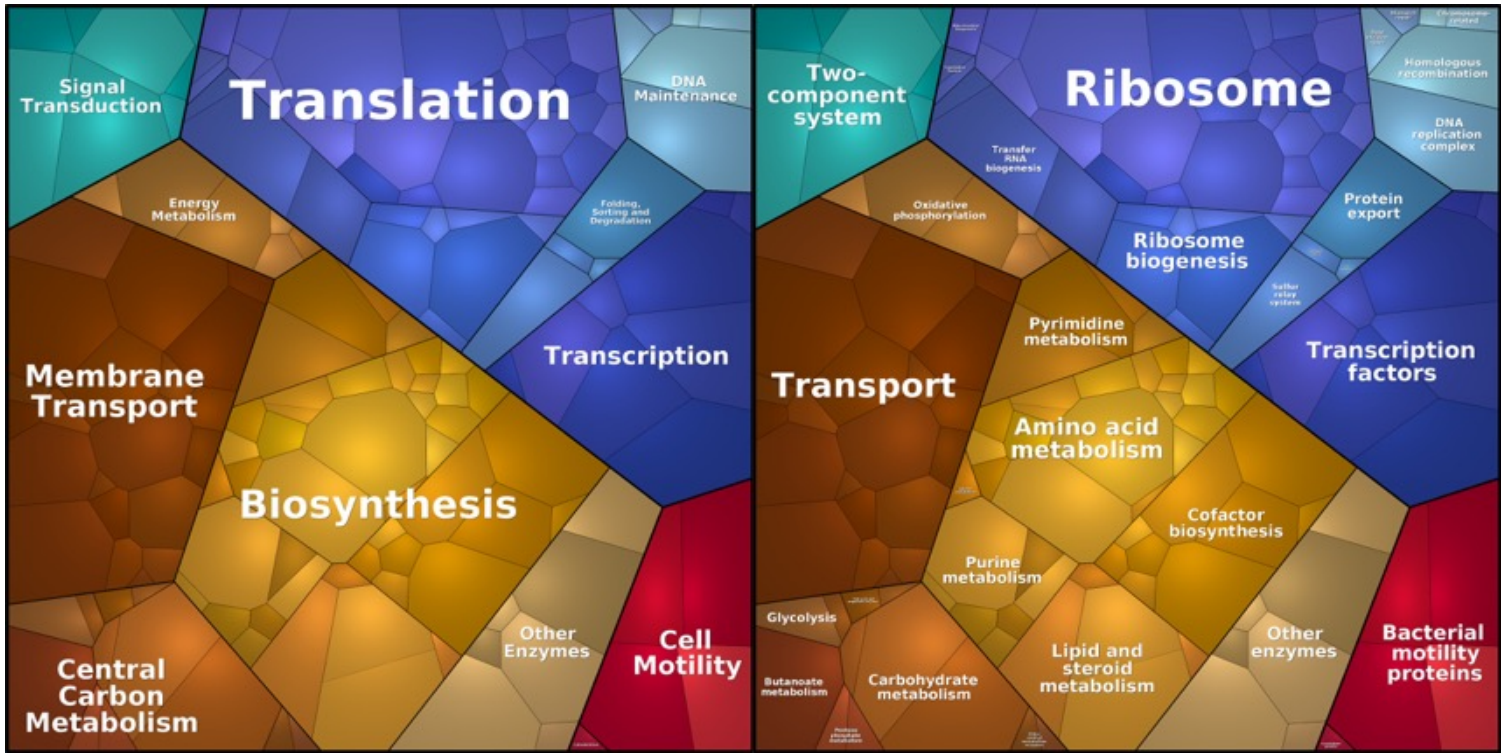


Figure 4.11: Biological processes affected in SH2 in Δ*spoVG*

Voronoi treemap of all the **(A)** increased **(B)** decreased proteins in Δ*spoVG*. The area of the polygon is proportional to the absolute logFC of the DPPs.

Table 4.10: Proteomic profile of decreased ribosomal proteins in Δ*spoVG* in SH2

Proteins are arranged in the ascending order of their logFC values. Only top 10 proteins are shown.

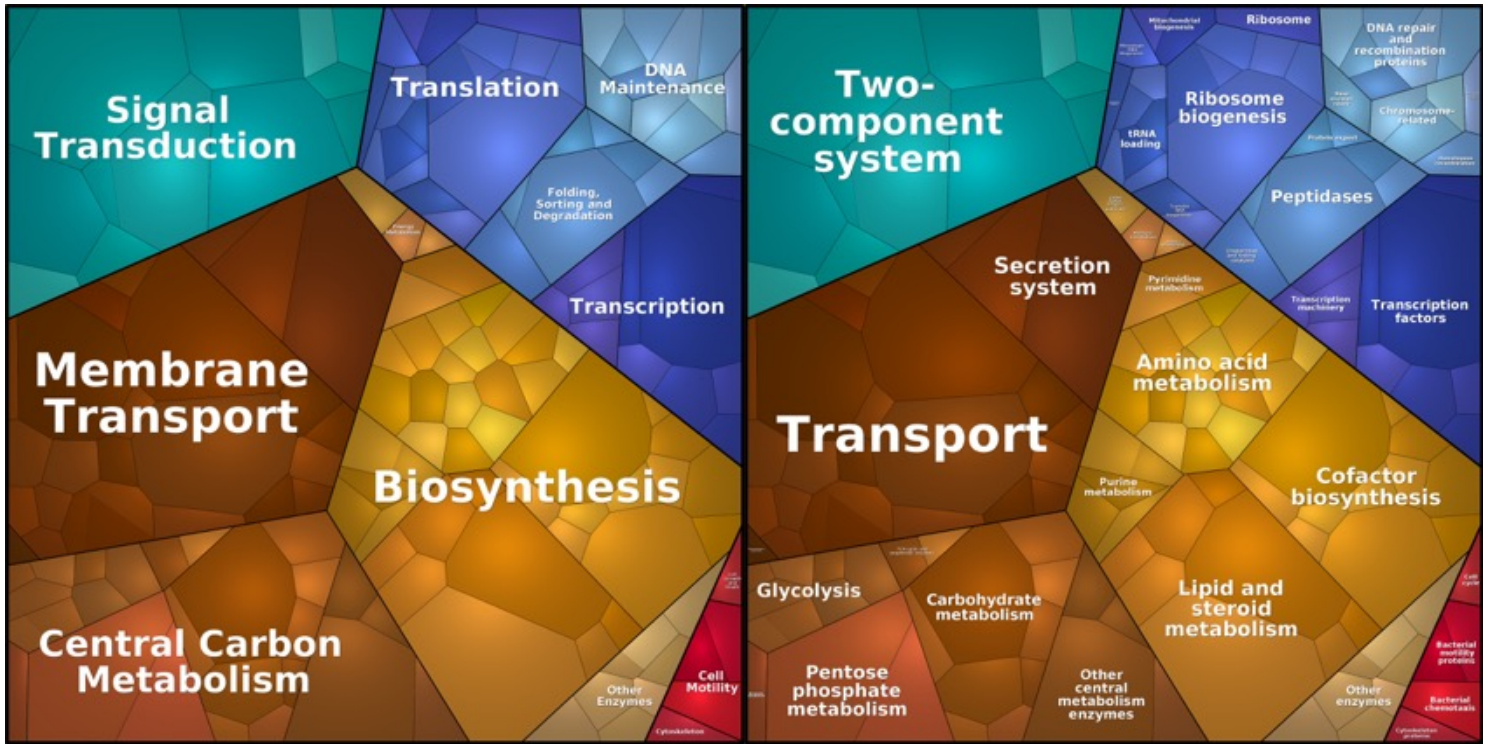
Locus Tag	Gene	Product	Description	Function	logFC	padj
BSU25410	*rpsU*	ribosomal protein S21	ribosomal protein	translation	-12.511	7.40E-17
BSU27940	*rpmA*	ribosomal protein L27 (BL24)	ribosomal protein	translation	-8.161	4.68E-03
BSU24900	*rpmGA*	ribosomal protein L33a	ribosomal protein L33a	translation	-7.970	5.25E-10
BSU01400	*rpmJ*	ribosomal protein L36 (ribosomal protein B)	ribosomal protein	translation	-7.144	3.25E-07
BSU40500	*rplI*	ribosomal protein L9	ribosomal protein	translation	-2.569	7.71E-09
BSU01240	*rpmC*	ribosomal protein L29	ribosomal protein	translation	-2.409	3.47E-07
BSU01200	*rpsS*	ribosomal protein S19 (BS19)	ribosomal protein	translation	-2.303	4.67E-05
BSU29660	*rpsD*	ribosomal protein S4 (BS4)	ribosomal protein	translation	-2.083	2.09E-05
BSU01190	*rplB*	ribosomal protein L2 (BL2)	ribosomal protein L2, required for assembly of RplP into the 50S subunit of the ribosome	translation	-2.044	1.04E-05
BSU01230	*rplP*	ribosomal protein L16	ribosomal protein L16, binding of L16 to the 50S subunit is required for the stimulation of RbgA GTPase activity, for release of RbgA from the ribosome, and for the conversion of the intermediate to the complete 50S subunit	translation	-1.998	2.16E-05

4.2.5.5 Sporulation proteins

The GSEA in SH5 identified Sporulation proteins as the only enriched gene set (Figure 4.10). There were 260 proteins detected in this gene set out of which majority of the proteins were decreased in the mutant (Table 4.8). Among the decreased DPPs in this gene set, proteins involved in processes such as activation of SigE (early mother cell-specific sporulation sigma factor), SigG (late forespore-specific sporulation sigma factor) and SigK (late mother cell-specific sporulation sigma factor); dissolution of septal wall (SpoIID); development, assembly and maturation of spore envelope (SpoVS, CotI), cortex (PdaB) and crust (CotY); spore morphogenesis (SpoVD); transcription of early mother cell (sporulation) genes; and spore germination (CwlJ, YnzB, GerT) were observed (Table 4.5 (B)). Additionally, proteins involved in biofilm formation and chemotaxis; initiation of sporulation and competence development (OppB); protection of spore (YwjD, YuzC), spore DNA (SspO and SspF) and spore coat proteins (SpoVC); killing of non-sporulating sister cells (SdpA, SdpC), maturation of toxin and antimicrobial resistance proteins were also decreased (Figure 4.12).

Sporulation is a complex developmental pathway in *B. subtilis* and is initiated in prolonged nutritional starvation conditions. The process of sporulation is divided into seven major stages and involves an array of proteins. Spo0A, a major transcriptional regulator, is central to the initiation of sporulation or biofilm matrix formation depending on its concentration of phosphorylated conformation in the cells (Molle *et al.*, 2003; Fujita, González-Pastor and Losick, 2005). The decreased amounts of many of the sporulation proteins at SH5 in ∆*spoVG* suggested a potential role of SpoVG in the later stages of sporulation.

(A)


Signal Transduction
Translation
DNA Maintenance
Folding, Sorting and Degradation
Transcription
Membrane Transport
Biosynthesis
Central Carbon Metabolism
Other Enzymes
Cell Motility
Two-component system
Ribosome
DNA repair and recombination proteins
Ribosome biogenesis
tRNA loading
Chromosome-related
Peptidases
Secretion system
Pyrimidine metabolism
Transcription machinery
Transcription factors
Amino acid metabolism
Transport
Purine metabolism
Cofactor biosynthesis
Glycolysis
Carbohydrate metabolism
Lipid and steroid metabolism
Pentose phosphate metabolism
Other central metabolism enzymes
Other enzymes
Bacterial motility proteins
Bacterial chemotaxis

(B)

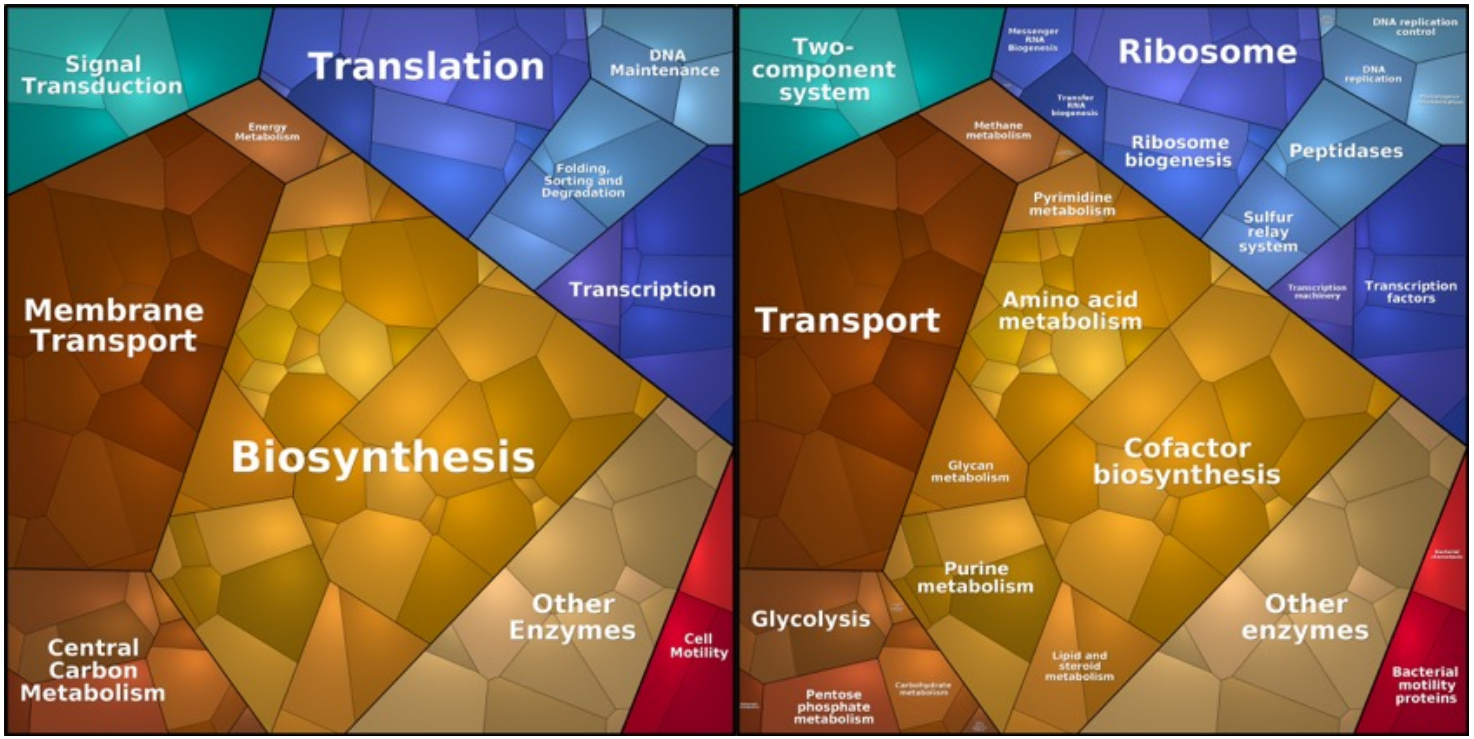


Figure 4.12: Biological pathways affected in SH5

Voronoi treemap of the **(A)** increased and **(B)** decreased DPPs in SH5. The size of each polygon is proportional to the absolute logFC of the proteins.

Further, in contrary to the decreased sporulation proteins, the production of Spo0B (a phosphotransferase that is involved in the phosphorelay in the initiation of sporulation (Trach and Hoch, 1989)) was increased (logFC ~5.3) in Δ*spoVG*. In the late stages of sporulation (SH5), a decreased level of Spo0B would have been expected. Therefore, the increased level of Spo0B in Δ*spoVG* in SH5 was interesting. Hence, it was hypothesised that perhaps SpoVG positively regulates many of the late-stage sporulation proteins while suppressing the sporulation initiation factors at SH2.

The results from SH2 showed a similar pattern of changes among the sporulation proteins in Δ*spoVG*. Especially, the stage V and stage VI proteins such as SpoVD (spore morphogenesis) (logFC = 0.52), SpoVID (spore coat assembly) (logFC = 1.6) and SpoVK (spore maturation) (logFC = 6.34) were increased in the mutant. The increased levels of late sporulation proteins at SH2 indicated that perhaps SpoVG represses the SH5 proteins at the early sporulation stages (SH2). This therefore enables the production of SH2 proteins to proceed to the later stages of sporulation.

Interestingly, in SH2 the levels of SpoVS (spore coat assembly, spore core dehydration) were decreased (logFC = -1.34) in Δ*spoVG*. The null mutation of *spoVS* when combined with the double null mutant Δ*spoVG*Δ*spoIIB* alleviates the sporulation blockage due to the Δ*spoVG*Δ*spoIIB* mutations (Resnekov, Driks and Losick, 1995; Perez, Abanes-De Mello and Pogliano, 2006). However, little is known about the relationship of *spoVG* and *spoVS.* Previously, *spoVS* has been shown to be transcribed by the SigH sigma factor (similar to *spoVG* and sporulation gene *spoIIB* (Britton *et al.*, 2002)) which is expressed early in the sporulation process and helps the cells to proceed to the stage V of sporulation (Resnekov, Driks and Losick, 1995). Therefore, from the findings of this study it is speculated that SpoVG may be positively interacting with SpoVS which promotes the cells to enter the stage V sporulation. The absence of SpoVG at SH2 restricts the complete functionality of SpoVS in facilitating the sporulation in SH2 hence the observed decrease in the sporulation proteins at SH5.

Some of the changed sporulation proteins in SH2 also included Spo0A (logFC = 1.65) and Spo0F (involved in the initiation of sporulation) (logFC = 0.56); Spo0M (sporulation-control gene) (logFC = -0.7), SpoIID (dissolution of the septal cell wall) (logFC = -9.34); SpoIIIAA (logFC = 10.01) and SpoIIIAG (activation of SigG) (logFC = -6.86), SpoIIIJ (membrane insertion of proteins and protein secretion) (logFC = -9.36), SpoIIP (required for dissolution of the septal cell wall) (logFC = 0.8) and SpoIIQ (forespore encasement by the spore coat) (logFC = 1.64). SpoIID is known to form a multi-protein complex with SpoIIM and SpoIIP (MPD complex) which is helps in the dissolution of the cell wall during sporulation (Chastanet and Losick, 2007; Khanna *et al.*, 2019). The localisation of the MPD complex driven by SpoIIM is dependent on SpoIIB (Chastanet and Losick, 2007).

SpoIIIAA and SpoIIIAG are both encoded from the *spoIIIA* operon (an operon with 7 to 8 genes) (Illing and Errington, 1991). SpoIIIAA is a protease and is required for the activation of sigma factor SigG (Camp and Losick, 2008). Whereas, SpoIIIAG is a component of the SpoIIIA-SpoIIQ type III secretion system that resides in the forespore membrane (Illing and Errington, 1991; Camp and Losick, 2008). SpoIIQ is a forespore protein and is a component of the SpoIIIAH-SpoIIQ type III secretion system that resides in forespore membrane which is required for anchoring of proteins on both sides of the sporulation septum; and for the localization and stability of SpoIIE (Camp and Losick, 2008; Campo, Marquis and Rudner, 2008; Levdikov *et al.*, 2012). SpoIIIJ is a vegetative protein and is encoded from *yidC1* which form an operon *rpmH-rnpA-yidC1-jag* or *yidC1-jag*. SpoIIIJ is a Sec-independent membrane protein translocase, which is essential for the activity of SigG at stage III sporulation and is also involved in the assembly of the SpoIIIAH-SpoIIQ complex (Camp and Losick, 2008). The changes in these proteins further indicated a role of SpoVG in the function of proteases, secretion proteins and the translocases in sporulation.

In summary, it was shown that absence of SpoVG impacts the sporulation process at both the SH2 and SH5 (Figure 4.10). It is hypothesized that SpoVG facilitates the sporulation proteins such as SpoVS at the early

stages of sporulation which in turn enable the cells to sporulate. The absence of SpoVG at SH2 decreases the levels of SpoVS thereby decreasing the sporulation proteins in SH5. More experimental investigations would be necessary to identify the exact mode of these regulation.

4.2.6 KEGG Pathway Enrichment Analysis of DPPs

GSEA helps to identify genes or proteins that are functionally associated with a biological process or molecular function based on the functional annotation. It does not consider the frequency or occurrence of the genes (or protein) compared to the entire set of the genes involved in a specific pathway. Therefore, to identify the significantly enriched biological pathways affected due to the DPPs alone in Δ*spoVG*, a Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis was carried out using hypergeometric statistical test. The pathways enriched with padj < 0.05 were considered statistically significant.

From this analysis, significantly enriched KEGG pathways were identified in each of the growth media used in the study (Figure 4.13). Two KEGG pathways - Histidine metabolism and Nonribosomal peptide structures were significantly enriched in LB. In M9, only one KEGG pathway - Nonribosomal peptide structures was significantly enriched (Figure 4.13). Among the sporulation conditions, two KEGG pathways – Ribosome and the RNA polymerase were significantly enriched in SH2 (Figure 4.13). Whereas, three KEGG pathways – Valine, leucine and isoleucine degradation, Arginine and proline metabolism and Taurine and hypotaurine metabolism were significantly enriched in SH5. Table 4.11 shows the details of the KEGG pathways enriched with DPPs in Δ*spoVG*.

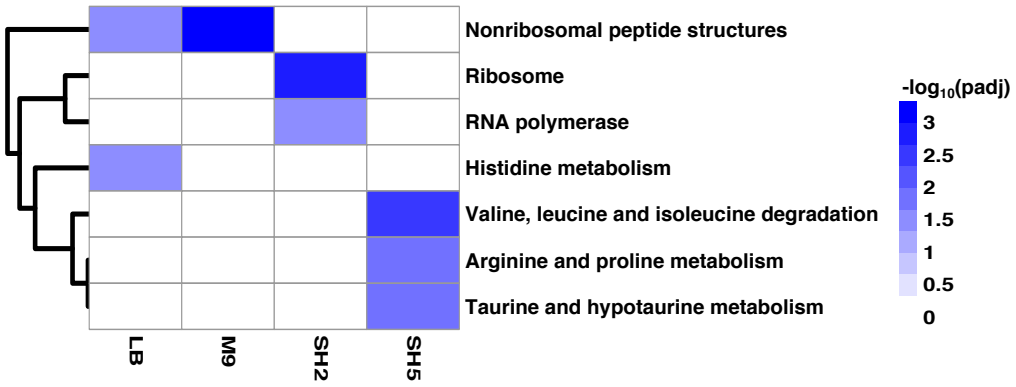


Figure 4.13: The KEGG enriched pathways due to the DPPs in Δ*spoVG*

Only the significantly enriched pathways are shown in shaded colours. The intensity is proportional to the significance (-log(padj < 0.05)) of the enrichment analysis. Deepest shade indicates the most significant pathway.

Among all the KEGG pathways identified across all the conditions, only Nonribosomal peptide structures was commonly enriched between LB and M9. In LB, plipastatin synthetase (increased in Δ*spoVG*) and the amino acid racemase (increased in Δ*spoVG*) were identified in the Nonribosomal peptide structures (Table 4.12). Whereas in M9, the plipastatin synthetases and surfactin synthetase proteins were identified and all the proteins decreased in the mutant. The decreased amount of the plipastatin proteins in M9 was expected due to the decreased levels of the antibacterial compounds observed in both the GSEA analysis of the transcriptomic (Chapter 3) and the proteomic profile of the mutant.

Table 4.11: Summary of the KEGG pathways enriched in due to the DPPs in Δ*spoVG* in all the growth conditions

KEGG pathways	padj	Total proteins in pathway	Proteins enriched
LB			
Histidine metabolism	3.05E-02	18	3
Nonribosomal peptide structures	4.52E-02	9	2
M9			
Nonribosomal peptide structures	4.62E-04	9	4
SH2			
Ribosome	1.25E-03	57	28
RNA polymerase	2.88E-02	5	4
SH5			
Valine, leucine and isoleucine degradation	2.20E-03	23	12
Arginine and proline metabolism	1.64E-02	25	11
Taurine and hypotaurine metabolism	1.94E-02	8	5

Table 4.12: Nonribosomal proteins affected in Δ*spoVG* in LB and M9

Proteins are arranged in ascending order of their genomic tag. Only the proteins with logFC > 1 or logFC < -1 are shown here.

Locus Tag	Gene	Product	Description	Function	logFC	padj
				LB		
BSU18340	*ppsA*	plipastatin synthetase	plipastatin synthetase	production of the antibacterial compound plipastatin	1.64	4.70E-02
BSU34430	*racX*	amino acid racemase	amino acid racemase, production of non-canonical D-amino acids	production of non-canonical D-amino acids	-5.57	2.62E-02
				M9		
BSU18320	*ppsC*	plipastatin synthetase	plipastatin synthetase	production of the antibacterial compound plipastatin	-2.13	2.51E-05
BSU18340	*ppsA*	plipastatin synthetase	plipastatin synthetase	production of the antibacterial compound plipastatin	-2.04	1.07E-02

Comparison of the LB and the SH5 KEGG pathways showed enrichment of the amino acid metabolism proteins. Additionally, most of the enriched proteins in these pathways were decreased in SH5 in Δ*spoVG* and all the identified proteins in the Histidine metabolism in LB were decreased. The amino acid metabolism is a subset of the nitrogen metabolism in *B. subtilis* which belongs to the core metabolic pathways of the cell (Fisher, 1991). In *B. subtilis* the carbon and nitrogen (glucose-glutamine) metabolic processes are intricately regulated. It has been shown that in glucose-glutamine nutrition limiting conditions e.g., sporulation, the expression of *spoVG* is repressed (Zuber and Losick, 1987) and the nitrogen metabolising genes is decreased (Boylan *et al.*, 1988; Hu *et al.*, 1999). Additionally, the amino acid metabolism was also affected in SH2. Especially, the production of GlnL was extremely increased in SH2 in the absence of SpoVG (logFC = 13.645) (Table 4.4). GlnL works with GlnK as a two-component system to regulate the genes involved in the biosynthesis and transport of glutamine (Satomura *et al.*, 2005). On the contrary, the levels of GltA and GltB (involved in the glutamate biosynthesis) were also reduced in SH2 (Bohannon and Sonenshein, 1989).

Amino acids are critical for several cell processes such as sporulation and competence development (Mäder *et al.*, 2002) and biofilm formation (Kimura and Kobayashi, 2020) in addition to the major metabolic processes. A significant decrease in the amino acid metabolic proteins in Δ*spoVG* suggested a role of SpoVG in the amino acid and nitrogen metabolism in *B. subtilis*. Although the mechanisms behind these changes are yet to be understood.

4.2.7 Affected Regulons in Δ*spoVG*

The results obtained so far strongly suggested that several regulons are perturbed in the absence of SpoVG in *B. subtilis*. In order to identify these regulons, analyses were carried out to identify the affected sigma factor regulons. The outputs showed that, in all the growth conditions, the proteins

of SigA regulon (comprising of household genes) and SigB regulons (consisting of GSR genes) were increased in the absence of SpoVG (Figure 4.14). This observation agreed with the previous observation that the GSPs were enriched in both the GSEA and KEGG pathway analysis. Furthermore, the proteins of SigD regulon (includes proteins participating in the chemotaxis and motility of the bacteria) were also increased in all the conditions except in LB (Figure 4.14). The change of SigD regulon proteins in LB was interesting since it suggested a different mode of regulation of the SigD regulon proteins in the absence of SpoVG than the rest of the growth conditions.

Analysis of the sporulation specific regulons i.e., SigE (early mother cell-specific sporulation), SigF (early forespore-specific sporulation), SigG (late forespore-specific sporulation) and SigK (late mother cell-specific sporulation) revealed that most of the changes occurring in the M9 conditions were similar to the SH2 compared to the SH5 (Figure 4.14). In M9 and SH2 all the sporulation sigma factor regulons were increased in the absence of SpoVG. Comparison of the sporulation sigma factor regulons in SH2 and SH5 showed that while most of the early sporulation regulon genes increased in SigE and SigF in SH2 compared to SH5, the late sporulation sigma factor regulon (SigG and SigK) genes decreased in SH5 vs SH2. Especially, the proteins of SigK regulons only increased in SH2 whereas they only decreased in SH5. From these observations it seemed that perhaps SpoVG represses the expression of the fore-spore specific proteins during SH2, thereby facilitating the cells to proceed to the later stages of sporulation and facilitates the expression of sporulation proteins in SH5.

Interestingly, contrary to M9 and SH2, all of the SigE, SigF, SigG and SigK regulon proteins were decreased in the SH5 in Δ*spoVG* (Figure 4.14). SpoVG is a known negative regulator of asymmetric septation in *B. subtilis* (Matsuno and Sonenshein, 1999). The decreased levels of sporulation protein in SH5 indicated a role of SpoVG in promoting stage V (late stage) sporulation in *B. subtilis*, an observation similar to the changed sporulation proteins identified in the GSEA.

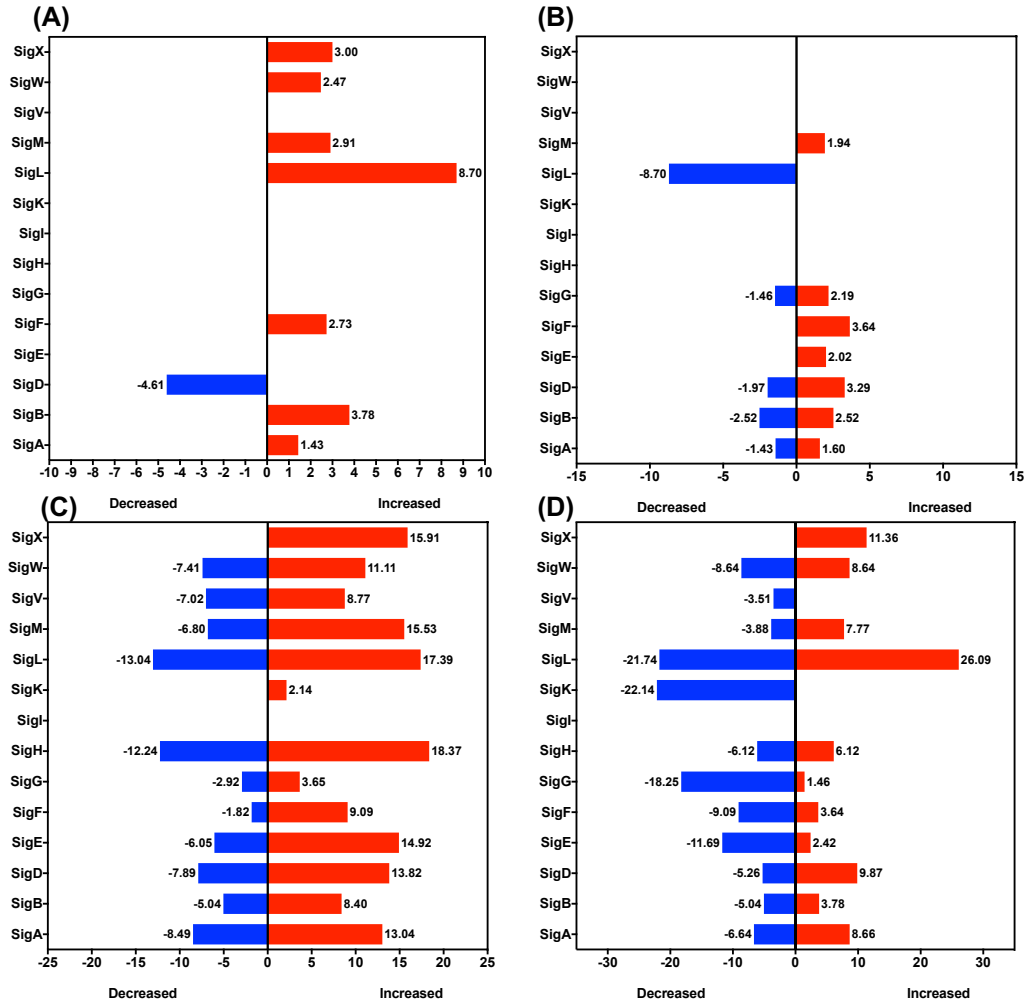


Figure 4.14: Sigma factors regulons affected in Δ*spoVG*

The bar graph shows the percentage of change (X-axis) of regulons (Y-axis) in each growth condition **(A)** LB, **(B)** M9, **(C)** SH2, **(D)** SH5. The negative bars on x-axis represent the percentage of change in decreased regulons.

SpoVG is a member of the SigH regulon (Britton *et al.*, 2002). The deletion of *spoVG* showed no change in the SigH sigma factor regulon in the LB and M9, however considerable changes were observed in SH2 and SH5 which included both increased and decreased DPPs (Figure 4.14). The regulation of SigH regulon genes is considered to be highly complex and not very well understood (Britton *et al.*, 2002). Major changes in the SigH regulon proteins in SH2 indicated a connection of SpoVG with the other SigH regulon

proteins. However, further investigations would be needed to understand the importance of these changes.

SigL is known to be an important link between the central carbon and nitrogen metabolism in *B. subtilis* (Choi and Saier, 2005). From this study, it was observed that SigL regulon proteins in LB only increased in Δ*spoVG* whereas the proteins significantly decreased in M9 (Figure 4.14 (A and B)). In both the sporulation conditions (SH2 and SH5) the proteins in the SigL regulon increased and decreased (Figure 4.14 (C and D)). The enrichment of SigL regulon proteins indicated a function of SpoVG in regulating both the carbon and nitrogen metabolism in *B. subtilis*.

Finally, the extra cytoplasmic function (ECF) sigma factor regulons SigM, SigV, SigW and SigX which regulate genes involved in cell wall metabolism and resistance to antimicrobial compounds were also affected (Helmann, 2016). Specifically, the SigV regulon was affected which provides resistance to lytic enzymes such as lysozyme (Guariglia-Oropeza and Helmann, 2011; Ho *et al.*, 2011). Both increased and decreased SigV regulon proteins were identified in SH2 (Figure 4.14 (C)). Whereas, in SH5 the SigV regulon proteins only decreased in Δ*spoVG* (Figure 4.14 (D)). This suggested a role of SpoVG in the resistance to lysozyme in SH5. Previous study by Burke and Portnoya, (2016) in *L. monocytogenes* had linked the lysozyme resistance to SpoVG where the loss of the protein rendered the cells hyper-lysozyme resistant. Therefore, the results from this study together with the previous evidence (Burke and Portnoy, 2016), indicated a role of SpoVG in lysozyme resistance in *B. subtilis*. Nonetheless, experimental evidence would be needed to validate these results *in vivo*.

4.3 Conclusion

This chapter presents the first global proteomic analysis of Δ*spoVG* in *B. subtilis* in LB, M9, SH2 and SH5 growth conditions. The results showed that 5studied. Maximal proteins were changed in SH2, where almost 29.4% of the total proteome was affected in Δ*spoVG*. This suggested an important

role of SpoVG in the early stages of the sporulation. Approximately 23% of the total proteins were also significantly differentially produced in SH5. These observations suggested an involvement of SpoVG in sporulation in *B. subtilis*. Significant changes in the proteomes of Δ*spoVG* were also observed in the nutritionally rich media (LB) and the minimal media M9 which comprised 3.8% and 4.6% of the total proteome respectively.

GSEA of the total identified proteins revealed a range of gene sets such as the Utilization of specific carbon sources, *General stress proteins (controlled by SigB)*, Translation, *Biosynthesis of antibacterial compounds* and Sporulation being affected in the absence of SpoVG. These observations were similar to the biological pathways enriched in the KEGG analysis of the DPPs. The KEGG pathway enrichment analysis in SH5 also identified the amino acid metabolic pathways, the central carbon metabolic pathways and changes in proteins that link these two pathways in Δ*spoVG*. Hence, a key role of SpoVG in both the central carbon and nitrogen metabolism during sporulation is speculated.

Proteins involved in altered carbon metabolic pathways were also altered in Δ*spoVG*. Investigations in other Gram-positive bacteria such as *L. monocytogenes* and *S. aureus* have also found SpoVG to have a role in the central carbon metabolism (Bischoff *et al.*, 2016; Burke and Portnoy, 2016). A similar observation from this study, suggested participation of SpoVG in the carbon metabolism of *B. subtilis* as well.

Major changes were also observed in several sigma factor regulons in Δ*spoVG*. For example, all the sporulation specific sigma factor regulons were affected in both SH2 and SH5, and a few in M9. Specifically, changes were also observed in the SigL sigma factor regulon across all the growth conditions in Δ*spoVG*. The SigL sigma factor regulon is known to be an important link between the central carbon and nitrogen metabolism (Choi and Saier, 2005). Hence, this suggested a potential role of SpoVG in the nitrogen and carbon regulation in *B. subtilis.*

From previous investigations of SpoVG in other bacteria, it is known that SpoVG participates in several metabolic processes such as the carbon

metabolism in *L. monocytogenes* (Burke and Portnoy, 2016), glycerol metabolism in *B. burgdorferi* (Savage *et al.*, 2018); cell wall metabolism and antibiotic resistance (Liu, Zhang and Sun, 2016), and central metabolism in *S. aureus* (Bischoff *et al.*, 2016). From the findings of the proteomic analysis of Δ*spoVG* presented in this chapter, it was shown that in addition to the previously known participations of the SpoVG, the protein also seems to be involved in processes such as general stress response, nitrogen metabolism, sporulation and translation. However, the exact role of SpoVG in *B. subtilis* is only beginning to be understood. Further phenotypic and biochemical experiments would be needed to validate the changes observed in this study, some of which will be discussed in the Chapter 7.

Chapter 5 RNA-Binding Profile of SpoVG in *B. subtilis*

5.1 Introduction

RNA-binding proteins (RBPs) form RNA-protein complexes called ribonucleoproteins (RNPs) that are involved in facilitating post-transcriptional gene regulation (Hentze *et al.*, 2018). Proteins such as Hfq and ProQ are some of the best-studied bacterial RBPs with roles in the small regulatory RNA (sRNA)-mRNA based regulation in the Gram-negative bacteria *Escherichia coli* and *Salmonella typhimurium* (Holmqvist and Vogel, 2018). Hfq was first identified as a host factor in *E. coli* for the RNA bacteriophage Qβ (Franze De Fernandez, Eoyang and August, 1968; Vogel and Luisi, 2011). Hfq is highly conserved across many bacteria which includes both the Gram-positive and negative bacteria. Based on the global regulatory role in the Enterobacteriaceae, Hfq has also been extensively investigated in Gram-positive bacteria such as *Listeria monocytogenes* (Nielsen *et al.*, 2009; Kovach *et al.*, 2014), *Clostridium difficile* (Salim *et al.*, 2012; Boudry *et al.*, 2014), *Bacillus anthracis* (Keefer *et al.*, 2017) and *B. subtilis* (Dambach, Irnov and Winkler, 2013). However, little to no effect on post-transcriptional gene regulation has been shown.

Although Hfq is able to bind RNA in *B. subtilis* (Dambach, Irnov and Winkler, 2013), it is not central in mediating sRNA-mRNA interactions in the post-transcriptional gene regulation (Hämmerle *et al.*, 2014; Rochat *et al.*, 2015). The absence of a global regulatory role of Hfq in *B. subtilis* indicates that either alternative RBPs exists which might participate in global regulation of sRNA-mRNA interactions, or the mechanism is independent of RBPs. *B. subtilis* may also have condition-specific RBPs which can regulate specific sRNA-mRNA interactions based on the growth condition, eliminating the need for a global gene regulator.

With recent investigations in a quest to identify RBPs with functions similar to Hfq, SpoVG has been identified as a novel RBP in the Gram-positive

bacterium *L. monocytogenes* (Burke and Portnoy, 2016) and the Gram-negative bacterium *Borrelia burgdorferi* (Savage *et al.*, 2018) with pleiotropic gene functions. In the *in vitro* analysis of SpoVG in *L. monocytogenes*, it was shown SpoVG binds to one of the abundantly expressed non-coding RNA (ncRNA) Rli31 in *L. monocytogenes* (Burke and Portnoy, 2016). Analysis of the Δ*spoVG* mutant in *L. monocytogenes* also identified pleiotropic phenotypic changes (i.e., motility, virulence and carbohydrate metabolism) in the deletion mutant, which was compared to the characteristic behaviour of the global regular CsrA in Gram-negative bacteria (Burke and Portnoy, 2016). Nevertheless, the mode of regulation of non-coding RNAs (ncRNAs) by SpoVG in *L. monocytogenes* remained unexplained. Also, no *in vivo* evidence of these findings was presented. *In vitro* experiments in *B. burgdorferi* also identified SpoVG as an RBP where it binds to the RNA transcripts of the *glpFKD* operon which encodes for proteins involved in the glycerol metabolism (Savage *et al.*, 2018).

SpoVG was first identified as a sporulation stage V protein (SpoVG) in *B. subtilis* and was shown to play a role as a negative effector of asymmetric septation (Segall and Losick, 1977; Matsuno and Sonenshein, 1999). Nonetheless, it is not known how it is able to influence this process. The RNA binding property of SpoVG in *B. subtilis* have yet to be investigated, which may provide the mechanism for this control. From the global study of *B. subtilis* in 104 growth conditions, *spoVG* was found to have a constitutive expression in all conditions (Nicolas *et al.*, 2012). Based on the evidence that SpoVG is an RBP in both Gram-positive and negative bacteria, and the constitutive expression of *spoVG*, it was hypothesized that SpoVG is an RNA-binding protein in *B. subtilis* and functions as a global post-transcriptional gene regulator.

Therefore, this chapter aimed to investigate the RNA-binding activity of SpoVG in *B. subtilis in vivo*. To identify the binding of RNAs to SpoVG, a Ultraviolet (UV) cross-linking and immunoprecipitation (CLIP) protocol (Ule *et al.*, 2003, 2005) was adopted. The RNA binding activity of SpoVG was also tested via UV cross-linking and analysis of cDNAs (CRAC)

(Granneman *et al.*, 2009), however the sequencing of cDNAs was not carried out. Finally, to identify the RNA-RNA interactions captured on SpoVG in *B. subtilis*, a Cross-linking, Ligation and Sequencing of Hybrids (CLASH) experiment was carried out (Kudla *et al.*, 2011). For CLIP and CRAC experiments, strain in which the SpoVG protein was tagged with epitope tags 3X-FLAG tag and His(6X)-TEV cleavage site-FLAG(3X) (HTF) tag were constructed (detailed in Chapter 2, section 2.2.3.3). In both the tagged strains, the *spoVG* gene was cloned under the control of its native promoter and integrated into the *amyE* locus. In CLIP, the RNAs are covalently cross-linked to the RBP via UV (λ = 254 nm). The cross-linking helps to retain only the bound RNAs to the protein upon stringent washing and cleaning steps. Although CLIP helps in the identification of the direct RNA-protein interactions, it poses some limitations such as semi-denaturing conditions and reliance on highly-specific antibodies for the purification of protein-RNA complexes in large RNPs (Granneman *et al.*, 2009). These limitations make CLIP less stringent and increases the likelihood of non-specific interactions also being identified. In the CRAC assay, the use of the bipartite tag His(6X)-TEV cleavage site-FLAG(3X) (HTF) provides better purification due to a two-step purification process even in harsh denaturing conditions (Granneman *et al.*, 2009).

Thus, the following sections of this chapter will discuss the results of the RNA-binding activity of SpoVG investigated via CLIP and CRAC experiments and the outcomes of the CLASH experiments, exploring the RNA-RNA interactions captured by SpoVG.

5.2 Results and Discussion

5.2.1 Construction of SpoVG-FLAG Strain

To construct a SpoVG-FLAG tagged strain the *spoVG* gene with its native promoter was tagged with a 3X-FLAG epitope tag (described in Chapter 2, section 2.2.3.2). The successful clone was chromosomally integrated into

the Δ*spoVG* strain to retain a single copy of the tagged protein in the cell. The cells expressing SpoVG-FLAG were screened via Western blotting against the monoclonal anti-FLAG antibody. A WT strain was used as a control for the un-tagged SpoVG. SpoVG is a small protein with a molecular weight of 10.75 kDa (Kim, Lee and Kwon, 2010; Michna *et al.*, 2014). The 3X FLAG tag used in this cloning produces a 23 amino acid peptide with a molecular weight of ~5.2 kDa.

The Western blot showed a single band for the tentative clone of SpoVG-FLAG at ~15 kDa (lanes 3, 4, 5, Figure 5.1). No bands were seen in the un-tagged control SpoVG from the WT strain (lanes 2, Figure 5.1). Therefore, the presence of a single band at ~15 kDa indicated the production of SpoVG-FLAG and were selected as positive clones for SpoVG-FLAG strains that were used for further experiments.

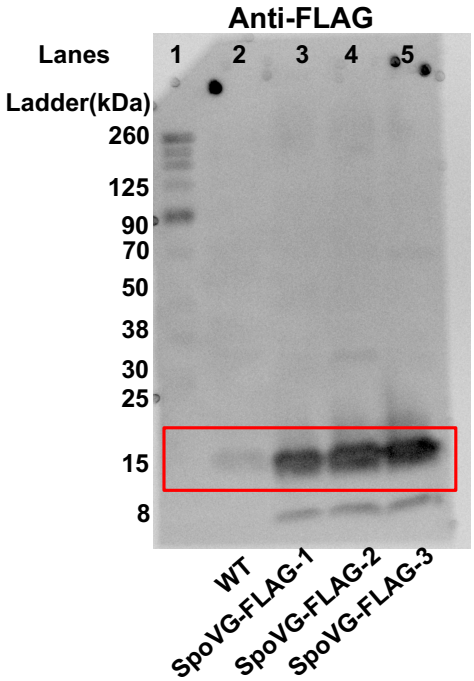


Figure 5.1: Western blot confirms the production of SpoVG-FLAG in *B. subtilis*

The Western blots were probed against monoclonal anti-FLAG secondary antibodies. The lane 2 shows the total cell protein lysates of the WT cells with native SpoVG (non-tagged) and lanes 3-5 show the total cell protein lysates of test biological triplicates of SpoVG-FLAG.

5.2.2 RNA-Binding Assay of SpoVG-FLAG with CLIP

NOTE: The following experiments were performed in the laboratory of Dr Sander Granneman at the University of Edinburgh.

5.2.2.1 Cell culturing and sample preparation

To identify the RNA-binding activity of SpoVG, the WT and SpoVG-FLAG strains were cultured in LB until exponential phase (OD_{600} = 0.5) and cross-linked *in vivo* with UV (λ = 254 nm) according to the CLIP protocol (Chapter 2, section 2.4.1). The tentative SpoVG-RNA complexes were purified for both the strains, and the RNAs of one set of the cross-linked samples were radiolabelled with ^{32}P-γATP as described in Chapter 2 (section 2.4.1). The other set was used as the un-labelled controls. Parallel autoradiogram and Western blot (WB) of the cross-linked labelled and un-labelled controls were produced.

5.2.2.2 CLIP identifies SpoVG-FLAG as an RNA-binding protein

The Western blot (WB) analysis of the immunoprecipitation against anti-FLAG monoclonal antibodies showed bands between 11-25 kDa (in a similar range to SpoVG-FLAG i.e., ~15 kDa) for the SpoVG-FLAG tagged strain only (lane 3 and 4, Figure 5.2 (B)). Whereas, no bands in this range were seen for the un-tagged control WT strains (lane 1 and 2, Figure 5.2 (B)). This, therefore, showed that IP was specific to the SpoVG-FLAG protein.

The parallel autoradiogram of WB from this CLIP experiment showed signals (smear) only in the radiolabelled SpoVG-FLAG sample (lane 4, Figure 5.2 (A)). Whereas, no signal was found for the radiolabelled WT or the un-labelled WT and SpoVG-HTF control strains (lane 1, 2, 3, Figure 5.2 (A)). The smeared signal of CLIP assays suggests the binding of non-uniform lengths of RNAs bound to the protein of interest, where the smallest

RNAs run to the bottom of the lane (Jensen and Darnell, 2008). Observation of smeared results in this investigation confirmed binding of RNAs of different lengths to SpoVG-FLAG, hence SpoVG as an RNA-binding protein in *B. subtilis in vivo*.

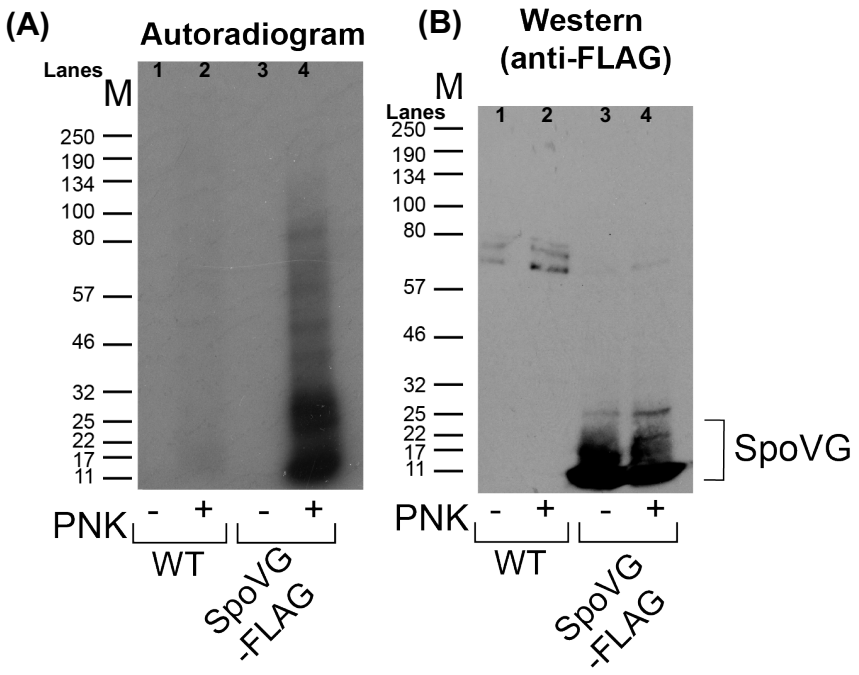


Figure 5.2: CLIP identifies SpoVG as an RBP in *B. subtilis in vivo*

(A) Autoradiogram and **(B)** Western blot of the UV cross-linked WT and SpoVG-FLAG cells investigated via CLIP analysis. +/- indicate radiolabelling of the cross-linked SpoVG-RNA complexes with/without ^{32}P- γATP respectively.

5.2.3 Construction of SpoVG-HTF Strain

To identify the RNAs binding to SpoVG and facilitate a better purification of the SpoVG-RNA complexes, a SpoVG-HTF expressing strain was constructed. The *spoVG*-HTF construct was expressed with the native promoter of *spoVG* (Figure 5.3 (A)), and a positive clone was transformed into the Δ*spoVG* strain to retain only a single copy of the tagged protein in the cells. To test the production of the SpoVG-HTF protein in the cells, total cell lysates of the successfully transformed cells were prepared and tested

with Western blotting. The WB was tested with the monoclonal anti-FLAG and polyclonal anti-His antibodies to confirm the presence of 3X-FLAG and His(6X) respectively. The lysates of WT cells were used as a negative control for the HTF tag. For testing the production of FLAG tag, a SpoVG-FLAG (~15 kDa) cell lysate was used as a positive control. Whereas, for testing the production of His(6X) tag in the HTF tag, a histidine specific protein marker and a SpoVG-His10 (~15 kDa) strain were used.

The results showed bands at ~20 kDa in all the SpoVG-HTF test strains for WB probed against both anti-FLAG and anti-His antibodies (lanes 2-5, Figure 5.3 (B and C)). No bands were observed in the control WT lysates for either of the anti-FLAG or anti-His blots (lanes 10, Figure 5.3 (B and C)). For anti-FLAG Western blot, no signals were detected in the histidine marker lane or the SpoVG-His10 tagged lanes. Additionally, no bands were observed for the negative control WT and SpoVG-FLAG in the anti-His WB. Hence, these results showed the successful construction, integration, and production of the SpoVG-HTF in the cells. For the CLASH experiments, one of these confirmed SpoVG-HTF strains was used.

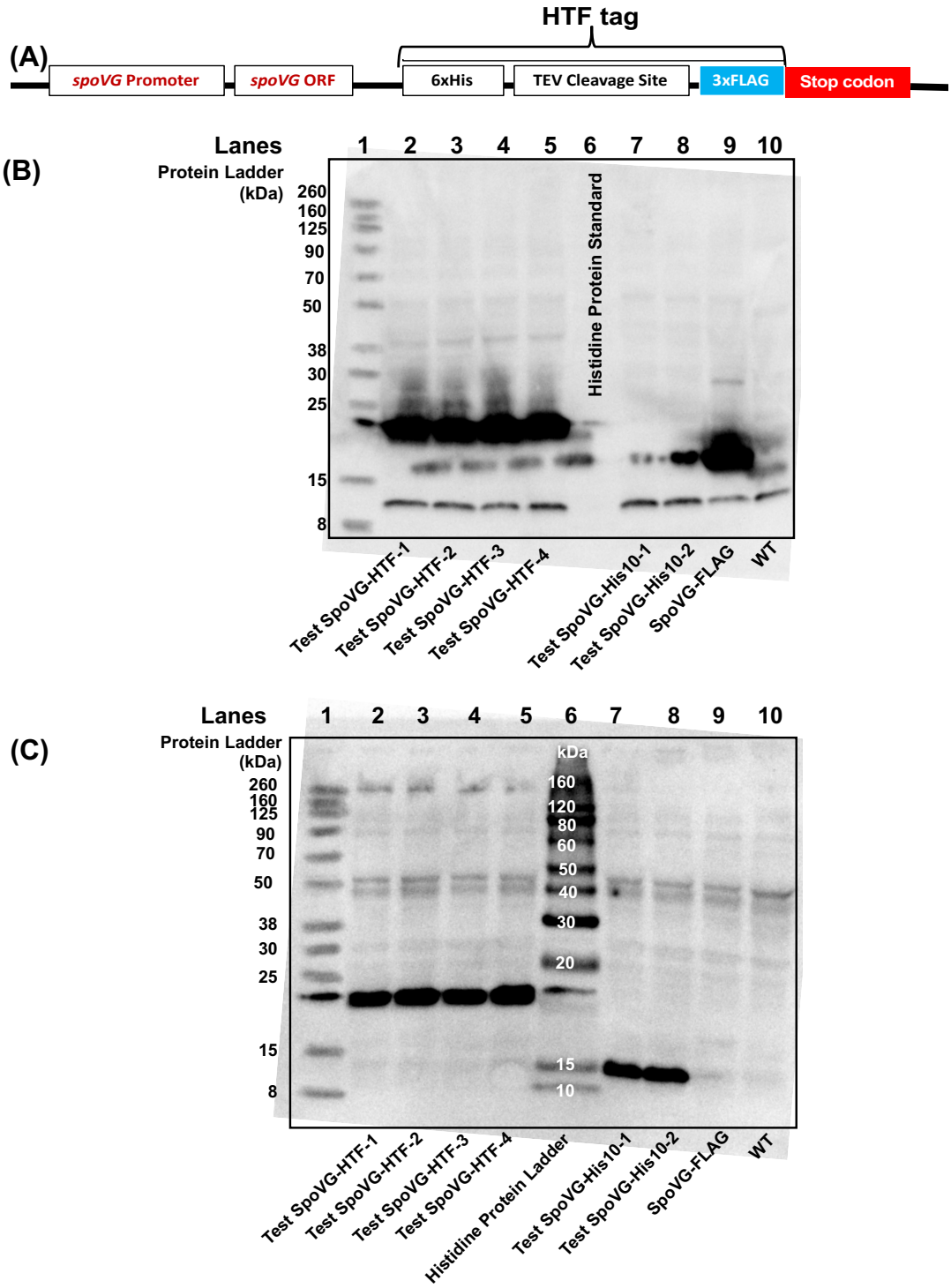


Figure 5.3: Western blot confirms the production of SpoVG-HTF in *B. subtilis* in LB

(A) The schematic of the SpoVG-HTF construct.; ORF = open reading frame codon **(B and C)** Western blots of total protein lysates probed with anti-FLAG and anti-His antibodies respectively. Lanes 2-5 show biological replicates of the test SpoVG-HTF strains, lane 9 = SpoVG-FLAG and lane 10 = WT.

5.2.4 RNA-Binding Assay of SpoVG-HTF

5.2.4.1 Cell culturing and sample handling

NOTE: The following experiments were performed in the laboratory of Dr Sander Granneman at the University of Edinburgh.

Before identifying the RNAs bound to SpoVG-HTF, the RNA-binding activity of SpoVG-HTF was tested. For crosslinking and visualization of the SpoVG-HTF-RNA complexes was carried out using the CRAC protocol except applying the sequencing steps (Chapter 2, section 2.4.2). In summary, a SpoVG-HTF strain was grown until stationary phase (OD_{600} = 3) in LB and UV cross-linked for 15s, 30s and 60s. The experiment also included a SpoVG-HTF strain that was not UV crosslinked and used as a negative control for crosslinking. The total protein lysates were immunoprecipitated via the HTF tag and the bound RNAs were ^{32}P-γATP radiolabelled (details in Chapter 2. Section 2.4.2). The purified and labelled SpoVG-RNA complexes were visualized via autoradiogram. A parallel Western blot of the same samples was also made using anti-TAP monoclonal antibodies to test specific IP for the HTF tag.

5.2.4.2 SpoVG-HTF binds RNA in vivo

The results of the Western blot analysis developed against monoclonal anti-TAP antibodies showed bands between 17-22 kDa (lanes 2-5, Figure 5.4 (B)). This was comparable to the Western blot of the SpoVG-FLAG against anti-FLAG in the CLIP analysis (lanes 3 and 4, Figure 5.2(B)), and the SpoVG-FLAG strain (lanes 3-5, Figure 5.1). Additional bands at ~32 kDa were also observed in all of the UV crosslinked samples (lanes 3-5, Figure 5.4 (B)), but not in the non-crosslinked control (lane 2, Figure 5.4 (B)). These bands were identified as potential SpoVG dimers which were kept together

by the bound RNA which remained intact after crosslinking. This suggested a potential hexameric structure of SpoVG as a biological unit.

Additionally, the comparison of the autoradiogram with the Western blot showed signals (smeared tracts) for all samples of SpoVG-HTF that were crosslinked for 15s, 30s and 60s (lanes 1-3, Figure 5.4 (A)). Whereas, no signal was observed in the non-crosslinked SpoVG-HTF control sample (lanes 4, Figure 5.4 (A)). All the crosslinked samples ran till ~22 kDa which was greater than the protein samples run on WB. This indicated an up-shift in the molecular weight due to the SpoVG-RNA complexes. Thus, this proved the binding of the RNAs were specific to the SpoVG-HTF.

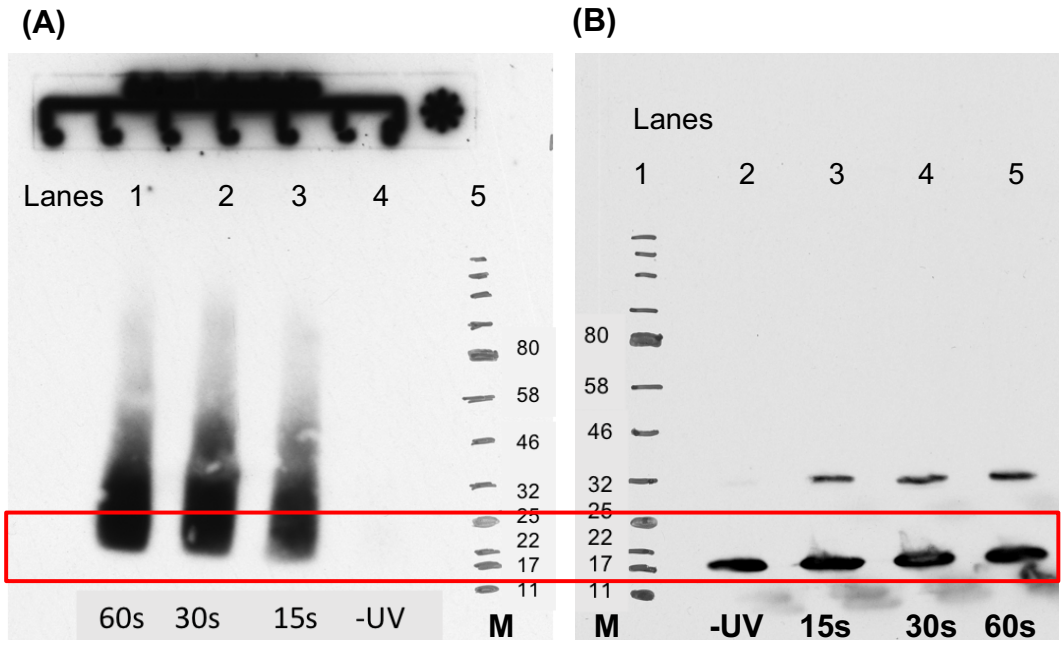


Figure 5.4: SpoVG-HTF binds to RNA transcripts (Courtesy: Dr. Sander Granneman – University of Edinburgh)

The R analysis of the SpoVG-HTF after being exposed to UV cross-linking after 15s, 30s and 60s. **(A)** Autoradiogram of the radio-labelled RNA molecules bound to SpoVG-HTF and exposed to UV **(B)** Western blot of the SpoVG-HTF samples assayed for autoradiogram and immunoprecipitated against anti-TAP monoclonal antibodies.

5.2.5 Analysis of RNAs Co-Immunoprecipitated with SpoVG

NOTE: The following experiments were performed and analysed in the laboratory of Dr. Sander Granneman at University of Edinburgh, UK.

5.2.5.1 Cell culturing and sample handling

To identify the RNA-RNA interactions mediated by SpoVG in *B. subtilis*, a CLASH protocol was adopted. Three biological replicates each of the WT and the SpoVG-HTF strains were UV-crosslinked, immunoprecipitated, sequenced and analysed as described in Chapter 2 (section 2.5).

5.2.5.2 Identification of RNAs bound to SpoVG

The sequencing analysis identified 2,265 RNA transcripts co-immunoprecipitated (co-IPed) with SpoVG-HTF. These RNAs included both protein coding and ncRNAs RNAs. Table 5.1 lists the top 20 co-IPed RNAs with SpoVG-HTF. Among the most abundant transcripts were the *srfAA*, *srfAB*, *srfAC* RNAs (Table 5.1). These transcripts belong to the *srf* operon of five genes that encode for the surfactant surfactin (Nakano and Zuber, 1990). Surfactin is pivotal for the development of competence and sporulation; and provides antimicrobial properties to *B. subtilis* (Zhao *et al.*, 2017). In addition to these three RNAs, *comS* and *srfAD* transcript (other members of *srf* operon) were also found in the CLASH assay.

Table 5.1: Top 20 co-immunoprecipitated (co-IPed) RNAs with SpoVG-HTF in this study

The transcripts are arranged in a descending order of the average read counts (abundance bound to SpoVG-HTF) of the SpoVG-HTF strain.

Locus tag	Gene	Average read counts (bp) (WT)	Average read counts (bp) (SpoVG-HTF)	Product	Description	Function	Gene length (bp)
BSU03490	*srfAB*	1,681	1,32,007	surfactin synthetase / competence	surfactin synthetase / competence	antibiotic synthesis	10,761
BSU03480	*srfAA*	1,205	1,09,515	surfactin synthetase / competence	surfactin synthetase / competence	antibiotic synthesis	10,764
BSU03510	*srfAC*	807	52,629	surfactin synthetase / competence	surfactin synthetase / competence	antibiotic synthesis	3,822
BSU39230	*wapA*	278	32,226	cell wall-associated protein precursor	cell wall-associated protein precursor, contact-dependent growth inhibition protein	intercellular competition	7,002
BSU11240	*carB*	703	18,339	carbamoyl-phosphate transferase-arginine (subunit B)	carbamoyl-phosphate transferase-arginine (subunit B)	biosynthesis of arginine	3,090
BSU01080	*rpoC*	136	17,155	RNA polymerase, βsubunit	RNA polymerase, βsubunit	transcription	3,597
BSU17210	*pksN*	58	15,545	polyketide synthase of type I	polyketide synthase of type I	polyketide synthesis	16,464

BSU12430	*rapA*	262	13,810	response regulator aspartate phosphatase	response regulator aspartate phosphatase, dephosphorylates Spo0F-P, control of the phosphorelay	control of sporulation initiation	1,134
BSU14860	*pycA*	107	13,547	pyruvate carboxylase	pyruvate carboxylase	replenishment of the oxaloacetate pool	3,444
BSU03210	*putC*	165	13,505	1-pyrroline-5-carboxylate dehydrogenase	1-pyrroline-5-carboxylate dehydrogenase	proline utilization	1,545
BSU35360	*hag*	134	13,069	flagellin protein	flagellin protein, about 20,000 subunits make up one flagellum	motility and chemotaxis	912
BSU18310	*ppsD*	50	12,977	plipastatin synthetase	plipastatin synthetase	production of the antibacterial compound plipastatin	10,809
BSU17180	*pksJ*	49	12,804	polyketide synthase	polyketide synthase	polyketide synthesis	15,135
BSU18000	*citB*	122	12,300	aconitase, trigger enzyme	aconitase, trigger enzyme	TCA cycle	2,727
BSU00500	*gcaD*	107	11,616	bifunctional N-acetylglucosamine-1-phosphate	bifunctional N-acetylglucosamine-1-phosphate uridyltransferase or glucosamine-1-phosphate acetyltransferase	cell wall metabolism	1,368
BSU17220	*pksR*	33	11,324	polyketide synthase	polyketide synthase	polyketide synthesis	7,629

BSU11430	*oppA*	123	11,310	oligopeptide ABC transporter (binding protein)	oligopeptide ABC transporter (binding protein)	initiation of sporulation, competence development	1,635
BSU37370	*albA*	90	11,054	radical S-adenosylmethionine enzyme	radical S-adenosylmethionine enzyme, antilisterial bacteriocin (subtilosin) production	antilisterial bacteriocin (subtilosin) production	1,344
BSU17190	*pksL*	35	10,708	polyketide synthase of type I	polyketide synthase of type I	polyketide synthesis	13,614
BSU37280	*narG*	44	10,488	nitrate reductase (alpha subunit)	nitrate reductase (alpha subunit)	nitrate respiration, nitrogen assimilation	3,684

In addition to the antimicrobial *srf* genes as some of the abundantly co-IPed RNAs with SpoVG-HTF, the CRAC data also found transcripts from other antimicrobial genes such as *pks* operon (involved in the biosynthesis of the polyketide synthases (Straight *et al.*, 2007)), plipastatin operons (*pps*) (operon encoding for the non-ribosomal peptide (NRP) plipastatin - an antifungal antimicrobial lipopeptide (Tsuge *et al.*, 1999)) in high abundance (Table 5.1). The transcripts of the antilisterial bacteriocin (subtilosin) operon (*sbo-alb)* were also amongst the most abundant transcripts co-IPed with SpoVG-HTF (Stein *et al.*, 2004). Many of the other abundantly co-IPed RNAs also had functions in motility and chemotaxis, transcription, amino acid metabolism, regulation of competence and sporulation initiation, nitrogen metabolism and central carbon metabolism (Table 5.1).

Also, interestingly amongst the top 20 abundantly co-IPed RNAs all the transcripts (except *hag*) had lengths greater than 1 kb (Table 5.1). This observation was intriguing and suggested perhaps a preferential binding of SpoVG with longer transcripts, thereby facilitating their stability. However, it remains unknown, whether these long RNAs were bound directly or indirectly with SpoVG. It is known that CLASH provides a highly stringent purification of sRNA-mRNA chimeras, ensuring only directly bound chimeras are immunoprecipitated with the protein of interest (Kudla *et al.*, 2011; Helwak *et al.*, 2013; Iosub *et al.*, 2020). Hence, to determine the specificity of binding of SpoVG to longer either independently or as a part of a chimera, further biophysical experiments would be needed.

5.2.5.3 Biological pathways affected by the transcripts bound to SpoVG

In order to visualize the biological pathways of the rest of the bound transcripts to SpoVG-HTF, a Voronoi tree map of unique transcripts with their average read counts (greater than 2,500) was constructed. The results identified metabolic processes such as, Central Carbon Metabolism, Nitrogen Metabolism, Membrane Transport, Energy Metabolism and many more among the other most bound transcripts (Figure 5.5). The data also

identified RNAs encoding for Genetic Information Processing such as Translation, DNA Maintenance, Transcription and Protein Folding, Sorting and Degradation. Whereas, few other transcripts belonged to Signal Transduction, Drug Resistance, Stress, Antimicrobial Compounds, Cell Division, Iron Metabolism, Secretion, Cell Wall Metabolism, Sporulation, Phages, Biofilm and Cell Motility in *B. subtilis*.

Many of these co-IPed transcripts, especially the antimicrobial genes, carbon metabolic genes, nitrogen metabolic genes, the amino acid metabolic genes and translation were also identified among the significantly differentially expressed genes and proteins in the transcriptomic and proteomic analysis of Δ*spoVG* (Chapter 3 and 4). These co-IPed transcripts were similar to the transcripts co-IPed with the global regulator Hfq and suggested a similar functional role of SpoVG to Hfq (Zhang *et al.*, 2003; Assis *et al.*, 2019).

To conclude, the sequencing data suggested that the RNAs bound to SpoVG participate in a range of biological processes. However, the exact role of SpoVG in binding to these transcripts is still to be elucidated. Furthermore, investigations such as identifying the interacting RNA-RNA partners would be helpful in identifying the kinds of interactions SpoVG participates. If SpoVG is a global post-transcriptional gene regulator with its role in mediating ncRNA and target RNA interactions, several ncRNAs shall be identified in the RNA-RNA complexes (Iosub *et al.*, 2020).

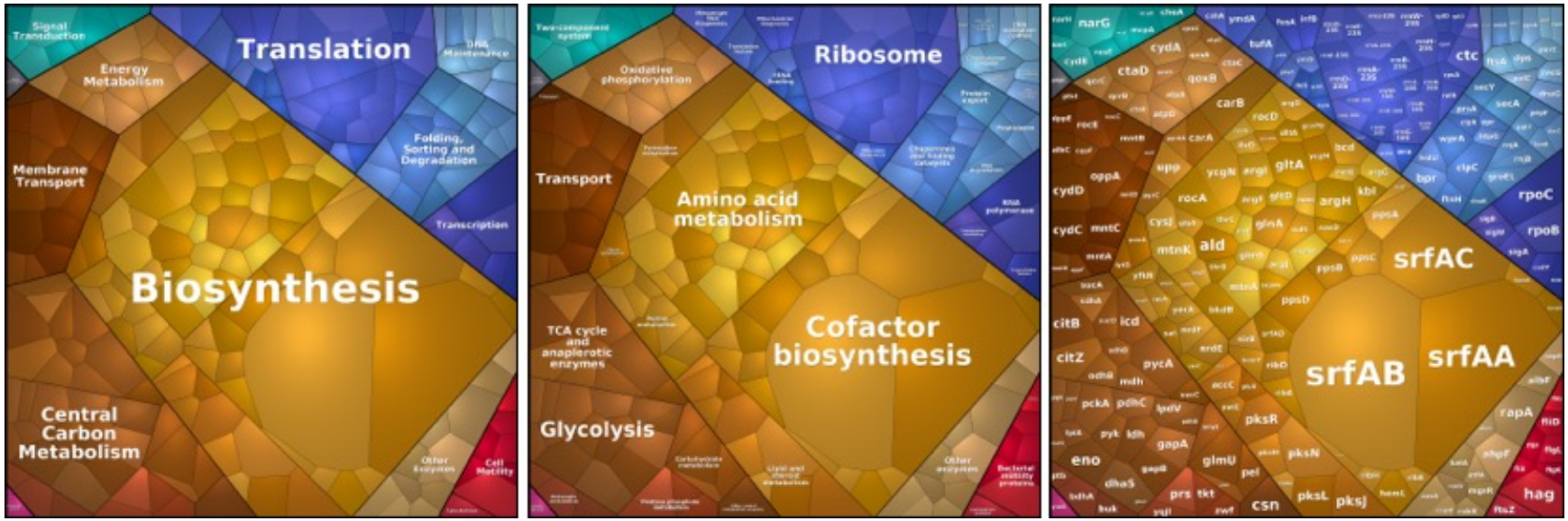


Figure 5.5: Functional characterisation of the transcripts bound to SpoVG

The Voronoi treemaps of top 336 RNA transcripts bound to SpoVG were generated via Bionic Visualizations' Proteomaps. The functional characterisation was based on the KEGG database. The area of each polygon represents the abundance of binding of the transcripts to SpoVG. From left to right → bigger to smaller functional classification of the transcripts bound to SpoVG and the corresponding transcripts identified in this study.

5.2.5.4 RNA-RNA interactions identified with SpoVG

The analysis of the CLASH sequencing data to identify RNA-RNA interactions identified 153 statistically significantly (with Benjamini-Hochberg adjusted *p*-value of 0.05 or less) enriched chimeras of ncRNA-mRNA co-IPed with SpoVG-HTF. This suggested that it is highly unlikely that these chimeras were generated due to random ligation of the RNA molecules. The most significantly enriched interaction was recorded between the transcripts of S276 and *sboA*. The *sboA* is an antimicrobial gene which is transcribed as a part of the *sbo-alb* operon, that encodes for the antilisterial bacteriocin subtilosin (Zheng *et al.*, 1999). Whereas, S276 is a 5' untranslated region (UTR) of the protein-coding gene *chaA,* that encodes for a calcium-specific antiporter that is involved in the Ca^{2+} signalling (Fujisawa *et al.*, 2009). In addition to this interaction, many other interactions between antimicrobial genes and other sRNAs were also observed. The identification of interactions between antimicrobial genes and the ncRNAs were interesting since the transcriptomic and proteomic analysis of the Δ*spoVG* strain identified a down-regulation of genes and proteins involved in the production of antimicrobial compounds (Chapter 3 and 4). Hence, these data taken together indicated a role of SpoVG in the antimicrobial compound synthesis in *B. subtilis*.

Interestingly, in addition to the interaction between antimicrobial genes and sRNAs, a known interactions between S415-*citZ* was also identified in this study (Durand *et al.*, 2015). Identification of the known interaction between S415-*citZ* was reassuring and validated the findings of this study. S415 (RoxS) is a known regulator of the TCA cycle gene *citZ* which is involved in the central carbon metabolism (Durand *et al.*, 2015). Furthermore, interactions were also identified between the small RNA S345 (RosA) and antimicrobial genes such as *ppsC*, *sboA*, *srfAB* and *srfAA* (Nakano *et al.*, 1991; Nakano, Xia and Zuber, 1991; Tsuge *et al.*, 1999; Zheng *et al.*, 1999). S345 (RosA) is a recently discovered novel sponge RNA in *B. subtilis* that interacts with sRNA RoxS/S415 and FsrA/S512 (Fur regulated small RNA)

(Durand *et al.*, 2019). Additionally, the expression of RosA is controlled by the carbon catabolite control protein A (CcpA), which links the activity of RoxS to the availability of carbon in *B. subtilis* (Durand *et al.*, 2019). In the CLASH analysis presented in this chapter, the previously known targets of S345 i.e., RoxS and FsrA were not identified. Nonetheless, the interaction between S345 and the antimicrobial genes were intriguing and suggested perhaps new interacting partners of S345 that are yet to be confirmed via direct biophysical experiments. Also, it remains unknown, how these interactions are mediated by SpoVG. To investigate these findings further, experiments would be needed to validate these findings and decipher the role of SpoVG in mediating such interactions.

In this chapter, even though several sRNA-mRNA interactions were co-IPed with SpoVG, the number of these interactions were much lower than the expected number of interactions if SpoVG was a general mediator of sRNA-mRNA interactions akin to Hfq. As global regulators RBPs such Hfq facilitate more than thousands of sRNA-mRNA interactions in *E. coli* (Iosub *et al.*, 2020; Quendera *et al.*, 2020). Thus, it was concluded that SpoVG in *B. subtilis* is not a global mediator of sRNA-mRNA interactions. However, it would be interesting to follow up some of the identified sRNA-mRNA interactions to validate the interactions and identify their roles in *B. subtilis*.

5.3 Conclusions

In this chapter, using the CLIP and CRAC technology it was shown that SpoVG is an RNA-binding protein in *B. subtilis in vivo.* The sequencing data of the co-immunoprecipitation experiment identified several RNAs co-IPed with SpoVG. The functional analysis of the co-IPed transcripts identified genes of diverse functions. Specifically, many antimicrobial RNAs such as the genes of the *srf, pks, pps* and *sbo-alb* operons were found to be among the most abundant RNAs bound to SpoVG-HTF. This observation was interesting since numerous of these antimicrobial genes and proteins were also found to be down-regulated in the transcriptomic and proteomic

analysis of the Δ*spoVG* strain in different growth conditions (Chapter 3 and 4).

In addition to the antimicrobial genes, the bound RNA molecules were also involved in a range of molecular processes such as central carbon metabolism, amino acid metabolism and nucleotide metabolism, cell-wall synthesis, and utilization of alternate carbons. These results were comparable to the previous -omics experiments (Chapter 3 and 4), where the differentially expressed genes and proteins in the Δ*spoVG* strain belonged to a range of biological processes and molecular functions. Hence, the observations in this study that SpoVG binds to an array of transcripts with functions ranging from metabolic genes to motility was both coherent and validated the data produced previously in this thesis.

Another interesting observation from the CLASH sequencing data set was that most of the abundantly co-IPed transcripts with SpoVG had a length of over 1 kb. Although intriguing, it suggested that perhaps SpoVG has a preference to bind to longer transcripts to protect and stabilize them, until their functions are completed. Nevertheless, targeted experiments would be needed to establish such roles of SpoVG *in vivo*. Binding of global RBPs to longer transcripts is under investigated. Most of the studies pertaining to the functions of global RBPs have focused on their interactions with sRNAs. However, data from a study on ProQ (an established bacterial RBP as a post-transcriptional regulator in *Salmonella* and *E. coli*) showed binding of ~44% of all the co-IPed mRNAs with ProQ to have length greater than 1 kb (Holmqvist *et al.*, 2018). It will be interesting to identify any factors that may contribute to the binding of global RBPs to longer transcripts (including SpoVG), if any.

The analysis of RNA-RNA interactions mediated by SpoVG identified 153 statistically significantly enriched sRNA-mRNA interactions. In this data set, the known interactions between S415-*citZ* and S345-S415 were identified. This finding validated the data set obtained from this study and suggested a role of SpoVG in the central carbon metabolism in the *B. subtilis*. Nonetheless, the number of sRNA-mRNA interactions identified with

SpoVG was rather less than the thousands of interactions mediated by the global regulator Hfq in Gram-negative bacteria (Quendera *et al.*, 2020). Hfq has ~4,000-10,000 hexamers per cell in *E. coli* (Kajitani *et al.*, 1994; Azam *et al.*, 1999; Moon and Gottesman, 2011). Whereas SpoVG is ~5,000 copies per cell (Zhu and Stülke, 2018). Thus, it is believed that even though SpoVG mediates sRNA-mRNA interactions in *B. subtilis*, it is not a global regulator of sRNA-mRNA interactions. In the future, it would be interesting to identify the binding sites of SpoVG among the bound transcripts and decipher any conserved binding domains.

Chapter 6 Protein-Protein Interaction with SpoVG

6.1 Introduction

Proteins rarely function in isolation. Proteins often interact with other proteins and are important in maintaining cellular and biological processes. Protein-protein interactions (PPIs) are generally defined as the physical interaction(s) between two or more proteins that occur *in vivo* (Mackay *et al.*, 2007; Chatr-aryamontri *et al.*, 2008; de Las Rivas and Fontanillo, 2010). Generally, proteins with PPIs are either involved in pathways where they participate with proteins of similar functions or they are produced from genes co-expressing in proximity.

RNA-binding proteins (RBPs) are known to form ribonucleoprotein (RNP) complexes with their non-coding RNAs (ncRNAs) and target mRNAs. The protein components of the RNPs are generally multimeric complexes that may include monomeric or heteromeric proteins (Bleichert and Baserga, 2010). The RBPs are known to participate in a range of PPIs (Caillet *et al.*, 2019). A well-documented example of the PPIs in RNPs is the interaction of ribonuclease E (RNase E) with the RBP Hfq in *E. coli* (Morita, Maki and Aiba, 2006). It was shown that Hfq stably associates with RNase E with its C-terminal scaffold region (Morita, Maki and Aiba, 2006). Morita, Maki and Aiba, (2006) also showed that the association of the sRNAs SgrS and RyhB to RNase E was facilitated by Hfq, highlighting the importance of PPI in the RNA-based post-transcriptional gene regulation. Hfq is also known to interact with other proteins, including Rho, poly(A)polymerase (PAP) and RNA polymerase (reviewed in Morita, Maki and Aiba, (2006)). Interaction of Hfq with ribosomal proteins has also been widely studied (Butland *et al.*, 2005; Arifuzzaman *et al.*, 2006). A recent study in *E. coli* also identified interaction of Hfq with the Elongation Factor Tu (EF-Tu) and several other non-ribosomal proteins including CsrA (carbon storage regulator), RpoB (DNA-directed RNA polymerase subunit beta) and GyrA (DNA gyrase

subunit A) (Caillet *et al.*, 2019). In another study, the well-characterised RBP Hfq was shown to form foci in prolonged nitrogen starvation in *E. coli*, which are speculated to be ribonucleoprotein complexes involved in the degradation of RNA molecules or ribosomes (McQuail *et al.*, 2020). Taken together these evidence show a vast protein-protein interactome of Hfq in *E. coli* and showcases the importance of PPI with RBPs in diverse cellular processes.

Several RBPs in both eukaryotes and prokaryotes are known to possess bi- or multi-functional properties. Many of these proteins are generally metabolically important enzymes which may 'moonlight' as RBPs, performing tasks that are physiologically distinct from their enzymatic activities. Such moonlighting activity of enzymatic proteins provides an additional layer of complexity in gene regulation. A well-characterized bi-functional metabolic enzyme that also acts as an RBP is aconitase. Aconitase is known to participate in the enzymatic interconversion of citrate and isocitrate, thus playing an important role in the central carbon metabolism (TCA cycle). Aconitase also participates in iron-metabolism and virulence of several bacteria (Wilson *et al.*, 1998; Somerville, Mikoryak and Reitzer, 1999). In *B. subtilis*, aconitase is essential for sporulation and binds to RNA sequences similar to the iron-responsive elements (IREs) found in the eukaryotic mRNAs (Alén and Sonenshein, 1999). Such pieces of evidence highlight the complexity of the functions of RBPs in the cells and their roles in mediating crosstalk between different biological processes. Therefore, it is possible that SpoVG acts in conjunction with other proteins in modulating both the transcriptional and protein interaction networks *in B. subtilis*.

The results from the *in vivo* RNA-binding assays with SpoVG in Chapter 5 showed that SpoVG as an RNA-binding protein that binds to a variety of RNAs in *B. subtilis*. The total transcriptomic and proteomic analysis of the Δ*spoVG* deletion strain also identified pleiotropic changes in the expression of several metabolic genes and proteins, which maybe a direct or indirect outcome of the deletion (Chapter 3 and Chapter 4). Therefore, to explore

how SpoVG may be bringing about the changes at the RNA and protein levels in either RNA stability or translation, we set out to determine any protein-protein interactions in which it may participate. To test this, an *in vivo* cross-linking of proteins and tandem mass-spectrometry (MS^2) were applied. For *in vivo* crosslinking, formaldehyde was used. Formaldehyde is a well-known protein cross-linker and forms covalent bonds between transiently interacting protein partners in close proximity (up to 2.3-2.7Å apart) (Sutherland, Toews and Kast, 2008). The cross-linked proteins were co-immunoprecipitated (co-IPed) using a His(6X)-TEV cleavage site-FLAG(3X) (HTF) epitope tag co-produced with SpoVG (SpoVG-HTF). The co-IPed proteins were quantified via a label free quantification (LFQ) method (explained in Chapter 4). Thus, the following sections will discuss the results from the first global PPI study of SpoVG in *B. subtilis*.

6.2 Results and Discussion

6.2.1 Cell Culturing and Sample Handling

Three biological replicates of the WT 168, *trp*+ and SpoVG-HTF were grown in LB to stationary phase and cross-linked as described in Chapter 2 (section 2.6.2). For the non-crosslinked control cells, no formaldehyde was added. The immunoprecipitation was carried out against the HTF tag using monoclonal anti-FLAG antibodies. The co-IPed proteins were eluted from the anti-FLAG antibodies using glycine. The proteins were quantified on an Orbitrap LC-MS/MS (Chapter 2, section 2.6.5), matched against the *B. subtilis* proteomic database (GenBank accession number = AL009126.3, UniProtKB proteome ID = UP000001570) and data filtered by Dr. Liangcui Chu (laboratory of Dr. Sander Granneman Lab, University of Edinburgh). The differential analysis of cross-linked vs non-crosslinked WT and SpoVG-HTF samples was carried out as described in Chapter 2 (section 2.6.6). Figure 6.1 illustrates a schematic of the major steps applied in this study.

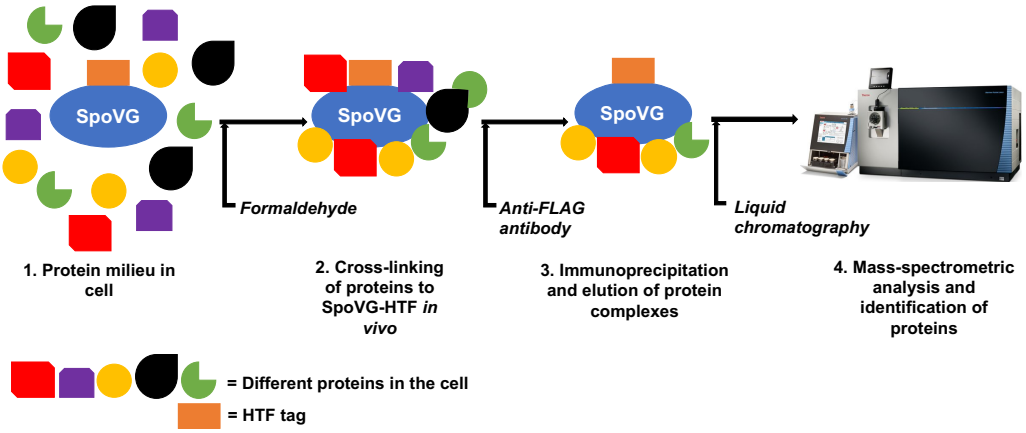


Figure 6.1: Schematic of the major steps in the PPI study with SpoVG-HTF and identification of the co-IPed proteins.

6.2.2 Confirmation of Protein Cross-linking and Immunoprecipitation

To test for successful protein cross-linking with formaldehyde, the total protein lysates were analysed by SDS-PAGE. The smearing of the cross-linked samples in the whole cell lysates indicated successful cross-linking of proteins *in vivo* (lane 3 and 5, Figure 6.2).

Further, to confirm specific immunoprecipitation (IP) against the HTF tag, the samples were separated by SDS-PAGE and a Western blot was carried out. The Western-blot analysis of the supernatant of the IP of the SpoVG-HTF samples showed a single band for the non-crosslinked samples at ~15 kDa (lane 4, Figure 6.3) and a smear for the cross-linked SpoVG-HTF samples (lane 5, Figure 6.3). Some smearing was also seen in WT samples (lane 2 and 3, Figure 6.3). However, since they were not as strong as the ones observed in the SpoVG-HTF samples (lanes 4 and 5, Figure 6.3), they were ignored as a biproduct of the high exposure of the Western-blot membrane. Thus, the single band at ~15 kDa indicated a specific IP of SpoVG-HTF with anti-FLAG antibodies (lane 4, Figure 6.3). Whereas presence of smear in the cross-linked SpoVG-HTF sample indicated cross-

linking of proteins of different molecular weights with SpoVG-HTF (lane 5, Figure 6.3).

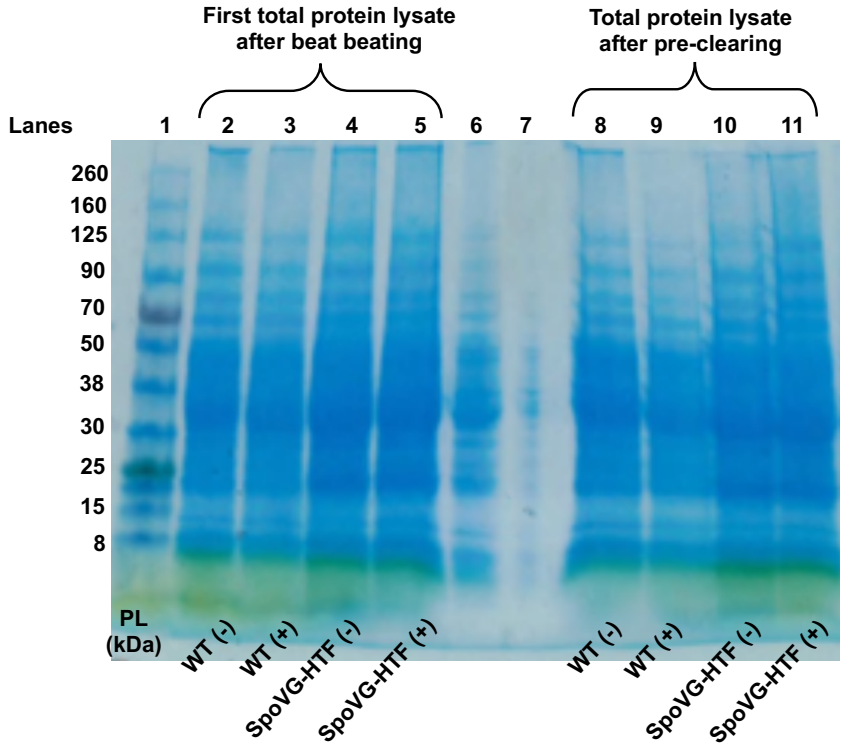


Figure 6.2: SDS-PAGE analysis shows more smearing in the total protein lysate of the cross-linked (+) samples compared to the non-crosslinked (-) controls.

The gel shows equal volume of the protein lysates for each sample run after beat beating and protein pre-clearing step. PL = protein ladder.

In the glycine eluted protein samples of both the cross-linked and non-crosslinked SpoVG-HTF showed single bands at ~15 kDa (lane 10 and 11, Figure 6.3). No bands were seen in either the cross-linked or the non-crosslinked WT samples of the glycine eluted samples (lane 8 and 9, Figure 6.3), which indicated the successful elution of the SpoVG-HTF-proteins complex with glycine.

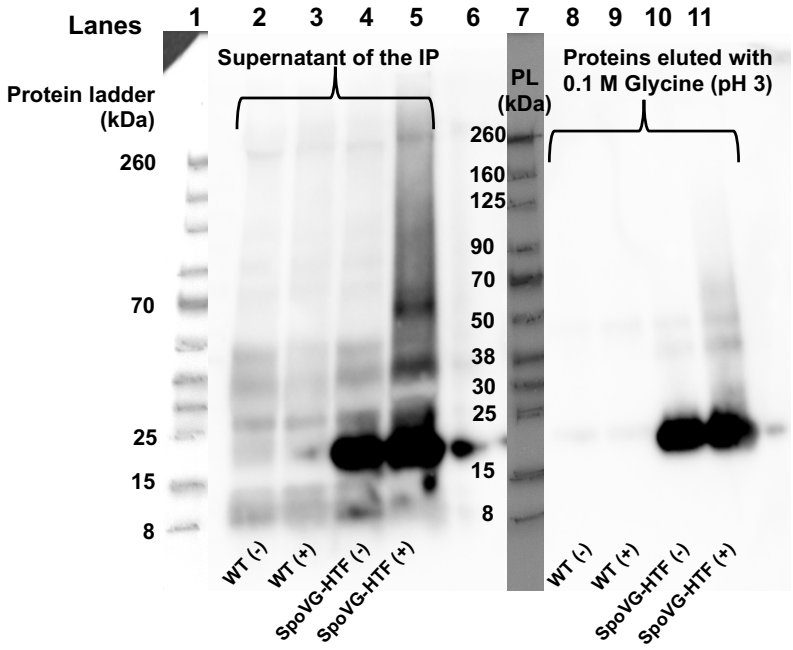


Figure 6.3: Western blot confirms successful cross-linking of proteins with SpoVG-HTF, specific immunoprecipitation of the protein complexes with anti-FLAG antibodies and successful elution of the protein complexes with glycine

Equal volumes of proteins were loaded on each lane. (+) = cross-linked, (-) = non-crosslinked, PL = protein ladder.

6.2.3 Identification of Proteins Cross-linked with SpoVG via MS/MS

The co-IPed and eluted samples were analysed by tandem mass-spectrometry (MS/MS) using the LFQ approach as described in Chapter 2 (section 2.6.5). The MS allowed identification of a cumulative number of 1,068 proteins for all the samples. To identify the significantly bound proteins with SpoVG-HTF, differential analysis of the cross-linked vs the non-crosslinked samples was carried out for both the WT and the SpoVG-HTF strains (described in Chapter 2, section 2.6.6). Comparison of the WT cross-linked vs non-crosslinked samples did not find any significantly changed proteins, which was expected. Nonetheless, comparison of the SpoVG-HTF cross-linked vs non-crosslinked samples identified 237 significantly co-IPed proteins with adjusted *p*-value (padj) < 0.05. Table 6.1

lists the top 10 co-IPed proteins with SpoVG with highest log_2 fold-change (logFC) values (as described in Chapter 2, section 2.6.6).

Analysis of these proteins showed that many of the co-IPed proteins (Eno, Icd, CggR, FbaA, RocA, Pgk, PckA) with high logFC values belong to the biological process carbon metabolism (Table 6.1). Proteins with roles in amino acid metabolism, nucleotide metabolism, other metabolic enzymes and biosynthesis pathways were also identified (Table 6.1). Many ribosomal proteins were also found to be co-IPed with SpoVG-HTF. The co-IP of the central carbon metabolism and amino acid metabolism proteins were also identified as significantly differentially produced in the absence of *spoVG* (Chapter 3 and Chapter 4). Therefore, these results together with observations from Chapter 3 and 4 indicated a potential role of SpoVG in the carbon and amino acid metabolism in *B. subtilis*.

Among the other co-IPed proteins was CcpA (Carbon catabolite control protein A) which had logFC = 7.67E+06 and padj = 3.61E-02. CcpA is a multifunctional global transcriptional regulator in *B. subtilis* which activates and represses a vast number of genes mediating carbon catabolite repression (CCR) (Henkin, 1996; Küster *et al.*, 1999; Warner and Lolkema, 2003). In a mass-spectrometric study by Wünsche *et al.*, (2012), SpoVG was recognised as one of the interacting proteins partners of CcpA.

Although this study identified an interaction between CcpA-SpoVG, no experimental evidence is present establishing a direct interaction between these two proteins (Wünsche *et al.*, 2012). Nonetheless, the identification of CcpA as one of the co-IPed proteins with SpoVG in this current study and the results from Wünsche *et al.*, (2012) suggested a potential interaction between CcpA-SpoVG in *B. subtilis*, which remains to be established with biophysical evidence of direct PPI.

Table 6.1: Selection of proteins co-IPed with SpoVG

Top 10 proteins arranged in the descending order of their log_2 fold change (logFC) values are shown here. Padj = adjust *p*-value.

Locus tag	Gene	Product	Description	Function	logFC	padj
BSU36890	*upp*	uracil phosphoribosyltransferase	uracil phosphoribosyltransferase	UMP synthesis	5.33E+09	1.69E-04
BSU33900	*eno*	enolase	enolase, glycolytic/ gluconeogenic enzyme, universally conserved protein	enzyme in glycolysis/ gluconeogenesis	3.22E+09	4.92E-03
BSU29130	*icd*	isocitrate dehydrogenase	isocitrate dehydrogenase	TCA cycle	1.36E+09	3.68E-03
BSU40320	*argI*	arginase	arginase	arginine utilization	1.05E+09	2.59E-04
BSU33950	*cggR*	central glycolytic genes regulator	repressor of the glycolytic *gapA* operon, DeoR family	transcriptional regulator	9.25E+08	3.21E-03
BSU37120	*fbaA*	fructose-1,6-bisphosphate aldolase	fructose 1,6-bisphosphate aldolase, glycolytic/ gluconeogenic enzyme	enzyme in glycolysis/ gluconeogenesis	8.81E+08	2.03E-02
BSU37780	*rocA*	3-hydroxy-1-pyrroline-5-carboxylate dehydrogenase	3-hydroxy-1-pyrroline-5-carboxylate dehydrogenase	arginine, ornithine and citrulline utilization	7.41E+08	2.48E-03
BSU31930	*ald*	L-alanine dehydrogenase	L-alanine dehydrogenase	alanine utilization	5.02E+08	3.99E-02
BSU33930	*pgk*	phosphoglycerate kinase	phosphoglycerate kinase, glycolytic/ gluconeogenic enzyme, universally conserved protein	enzyme in glycolysis/ gluconeogenesis	4.32E+08	4.42E-02
BSU30560	*pckA*	phosphoenolpyruvate carboxykinase	phosphoenolpyruvate carboxykinase	synthesis of phosphoenolpyruvate	3.79E+08	2.04E-02

6.2.4 Characterisation of Proteins Co-Immunoprecipitated with SpoVG

In order to understand the molecular functions and the biological pathways of the significantly co-IPed proteins with SpoVG, a KEGG pathways analysis was carried out and visualized on a Voronoi treemap. The results showed that a majority of the co-IPed proteins belong to the metabolic activity of the cell (Figure 6.4). These metabolic functions belong to central carbon metabolism (glycolysis, TCA cycle, pentose phosphate metabolism), nucleotide metabolism (purine and pyrimidine metabolism) and amino acid metabolism (Figure 6.4). Many of the co-IPed proteins also belong to the group genetic information processing such as transcription, translation, folding, sorting and degradation of proteins, DNA maintenance and RNA degradation. A few proteins also belonged to signal transduction and cellular motility (Figure 6.4).

To further understand the protein-protein association network of SpoVG, the 237 significantly co-IPed proteins analysed in the STRING database (11.0) using their logFC values (Szklarczyk *et al.*, 2019). The STRING database used Gene Set Enrichment Analysis (GSEA) to generate the interaction map with medium confidence (0.4). Only the interactions with known experimental evidence were retained and the unconnected proteins were ignored. The results of this analysis identified CcpA as the only protein with direct interaction with SpoVG (Figure 6.5). Several proteins with direct and indirect links to CcpA were also identified (Figure 6.5). Proteins with molecular functions such as catalytic activity, oxidoreductase activity, metal ion binding, ion binding and transferase activity were also enriched in this GSEA. Additionally, several proteins of the cytoplasm were also identified amongst the co-IPed proteins. Although at this point none of the interactions have been tested for their physical interaction with SpoVG, the diverse set of proteins co-IPed suggest a wide role of SpoVG in *B. subtilis*.

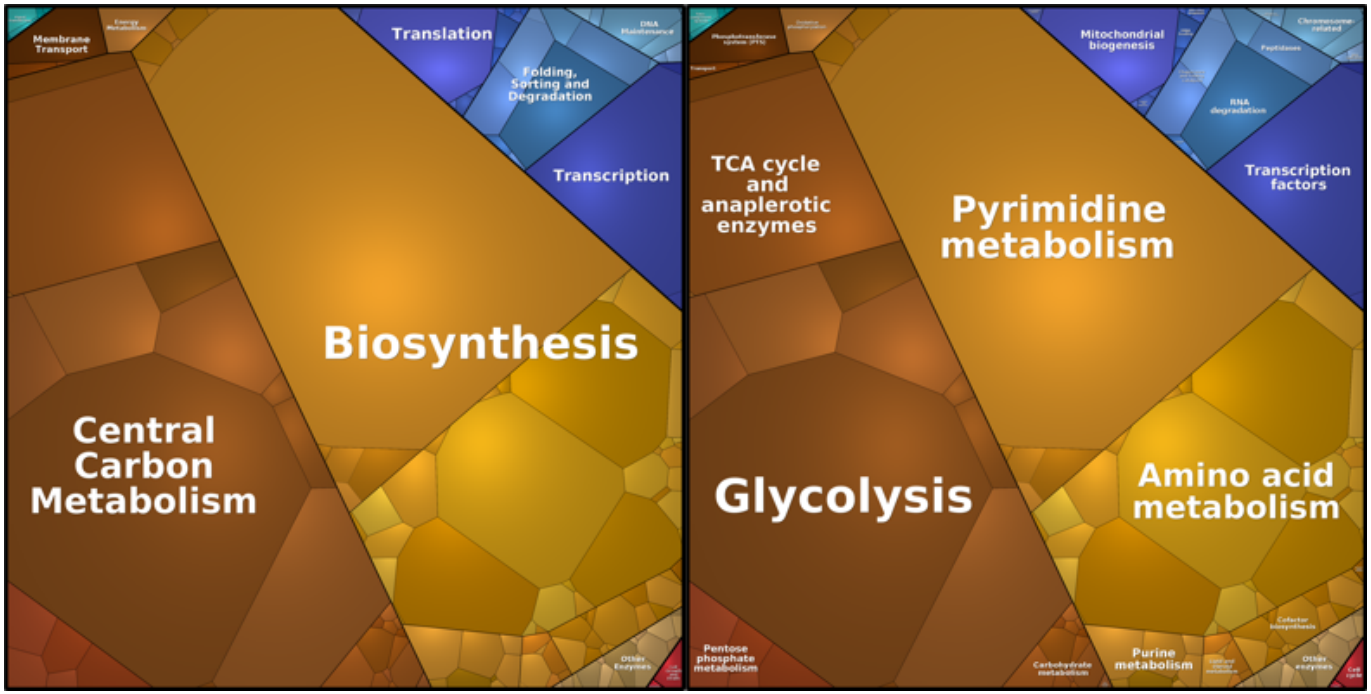


Figure 6.4: Significantly co-IPed protein with SpoVG belong to several KEGG pathways

Voronoi treemap of 237 significantly co-IPed proteins with SpoVG and their corresponding KEGG pathways. The area of each polygon in the treemap is proportional to the logFC values of the co-IPed proteins in the SpoVG-HTF cross-linked samples. Broader (left) and specific (right) of KEGG pathways

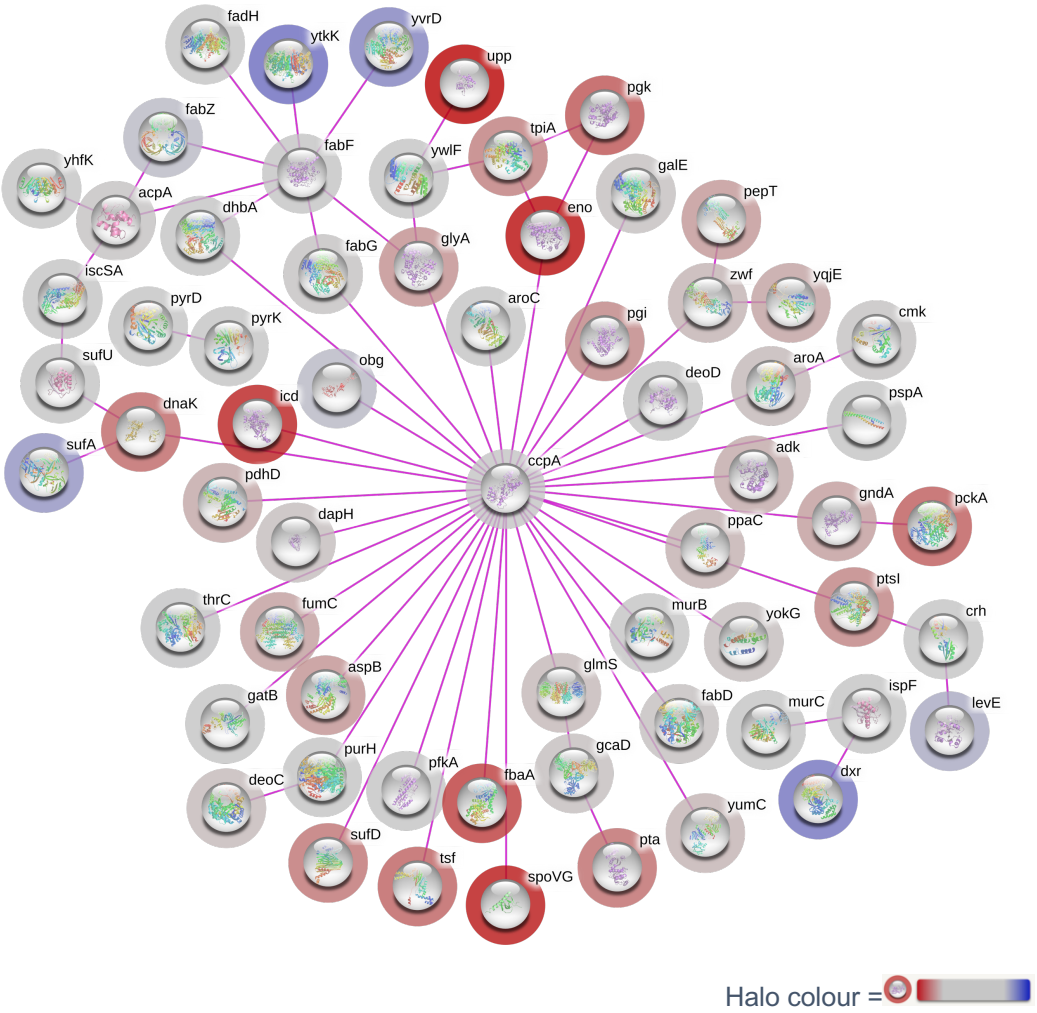


Figure 6.5: Experimentally identified protein-protein association network of SpoVG

The protein network shows experimentally established interactions that were identified in this protein-protein interaction study. The colour shades of halo indicate the logFC values (Red → Gray → Purple :: Max. logFC → median logFC → min. logFC) of the significantly co-IPed proteins with SpoVG.

6.3 Conclusion

This chapter presents the first *in vivo* global protein-protein interaction study of SpoVG in *B. subtilis*. 237 proteins were found to significantly co-immunoprecipitate with SpoVG. Some of the most abundantly proteins co-

IPed with SpoVG belong to biological processes such as carbon metabolism, amino acid metabolism and other biosynthesis pathways such as the biosynthesis of antibiotics, secondary metabolites, amino acids and nucleotides. The functional characterisation of all the significantly co-IPed proteins also identified proteins involved in transcription, translation, protein folding and degradation, DNA maintenance and RNA degradation. Whereas a few other co-IPed proteins also belong to signal transduction, uptake of additional carbon sources and cellular motility. These observations were comparable to the proteins binding to Hfq (Sukhodolets and Garges, 2003; Mohanty, Maples and Kushner, 2004; Butland *et al.*, 2005; Arifuzzaman *et al.*, 2006; Rabhi *et al.*, 2011; Caillet *et al.*, 2019). These observations were also in line with the diverse set of changes recorded in the pathway analysis of the transcriptomic and proteomic data of the ∆*spoVG* deletion strain (Chapter 3 and 4), and variety of RNAs co-IPed with SpoVG (Chapter 5).

SpoVG was initially identified in *B. subtilis* as a negative regulator of asymmetric division (Matsuno and Sonenshein, 1999) and a positive regulator of cortex formation (Rosenbluh *et al.*, 1981; Perez, Abanes-De Mello and Pogliano, 2000). SpoVG is also a regulatory protein in the human pathogen *Staphylococcus aureus*, where it executes its roles by binding to DNA (Liu *et al.*, 2016; Zhu *et al.*, 2019). In addition to *S. aureus*, SpoVG is also known as a regulatory protein in *Borrelia burgdorferi* and *Listeria monocytogenes* with DNA binding properties (Jutras *et al.*, 2013). Investigations in *B. burgdorferi*, *L. monocytogenes* and *Bacillus anthracis* have also identified SpoVG as an RNA-binding protein with pleiotropic regulatory functions (Burke and Portnoy, 2016; Savage *et al.*, 2018; Chen *et al.*, 2020). These growing body of evidence suggest a global regulatory role of SpoVG and its diverse interactions in bacteria. Nevertheless, the exact modes of these regulation are open for further investigations.

Some of the most well-studied RBPs such as Hfq, ProQ and CsrA are also known to possess multiple properties such as DNA-binding and protein-binding activities. These diverse properties of the RBPs in turn facilitate a range of functions which include the sRNA-mRNA interactions, sRNA

mediated regulatory functions and interactions with RNases and exonucleases. In chapter 5, it was shown that SpoVG is an RBP in *B. subtilis* which binds to RNAs of multiple functions and some sRNA-mRNA interactions. From this chapter, the identification of diverse proteins that co-IPed with SpoVG indicated a protein binding activity of SpoVG in *B. subtilis* as well. Taken together, these evidence pointed towards a multifaced functionality of SpoVG in *B. subtilis.* The mode of these interactions could be similar to the known global regulators or have novel interactions which remain to be investigated in the future.

In the future, it will be interesting to dissect the identified co-IPed proteins and explore their direct interaction with SpoVG. This will uncover more about the regulatory roles of SpoVG and *B. subtilis.* Comparing the available PPI data of Hfq with the output of this study may identify protein interactions that are conserved between different RBPs and SpoVG. The study of protein-protein interaction of SpoVG dependent or independent of its DNA and/or RNA binding activity will widen the understanding of its regulatory and RNA-binding activity in *B. subtilis*. The evidence presented support the hypothesis that SpoVG is a multifunction protein and is involved with several biological processes.

Chapter 7 Functional Characterisation of Δ*spoVG*

7.1 Introduction

Pleiotropy is a well-known characteristic of several bacterial RNA-binding proteins (RBPs). In organisms such as *Escherichia coli* , *Salmonella enterica*, *Pseudomonas aeruginosa* and *Legionella pneumophila*, RBPs such as Hfq, ProQ and CsrA are involved in a range of post-transcriptional gene regulation such as regulation of carbon metabolism, biofilm formation and virulence to name a few (Quendera *et al.*, 2020). These RBPs preferentially bind different variety of small non-coding RNAs (sRNAs) and mediate their interaction with the target mRNAs as RNA chaperones or help in the stability of the sRNAs structures (Melamed *et al.*, 2016, 2020). Additionally, RBPs also interact with ribonucleases (RNases) in promoting sRNA stability thereby playing a key role in the RNA based gene regulation in bacteria. The RNases interacting with RBPs include both endonucleases (e.g., RNase E and RNase III) and exonucleases such as polynucleotide phosphorylase (PNPase) (reviewed in Quendera *et al*., 2020). Though a number of studies have investigated the roles of these global RBPs in the bacterial gene regulation in the Gram-negative bacteria, not much is known about the role of RBPs in Gram-positive bacteria.

In an *in vitro* study in *L. monocytogenes*, Burke and Portnoya (2016) showed that SpoVG binds to RNA. They also showed that deletion of the *spoVG* gene leads to hypervirulence, absence of motility on soft-agar and hyper-lysozyme resistance phenotypes. A global transcriptomics approach also identified up-regulation of genes involved in carbon metabolism (Burke and Portnoy, 2016). Studies of *spoVG* in *Staphylococcus aureus* have also shown that deletion of *spoVG* renders the cell defective in secretion of extracellular enzymes (Schulthess *et al.*, 2011).

From the previous results presented in this thesis, it has been established that SpoVG is an RBP and binds to a range of RNA molecules in *B. subtilis*

(Chapter 5). In Chapter 6, it was shown that SpoVG co-immunoprecipitates with several proteins which are involved in processes such as central carbon metabolism, amino acid and nucleotide metabolism, cell-wall synthesis, and utilization of alternate carbon sources. The -omics data of Δ*spoVG* presented in Chapters 3 and 4 showed that the absence of *spoVG* perturbs hundreds of genes and proteins involved in a variety of metabolic and biological processes in the cell. Cumulatively, these observations indicated a pleiotropic function of SpoVG in *B. subtilis*.

The aim of this chapter was to compare the findings of the transcriptomic and the proteomic data obtained in the Chapters 3 and 4 respectively and to functionally validate some of these observations. For the functional validation of Δ*spoVG*, phenotypes such as motility, biofilm formation, secretion, competence, and growth in a range of carbon sources were tested. The phenotypes chosen for this investigation have all evolved as survival mechanisms to help *B. subtilis* thrive in a range of environments.

7.1.1 Motility in *B. subtilis*

B. subtilis can increase its motility when in nutrient-deficient conditions to both aid its search for additional food sources and to seek further locations for colonisation. *B. subtilis* exhibits different modes of motility, such as swimming, swarming and sliding, which are either dependent or independent of the flagellar movement (Kinsinger, Shirk and Fall, 2003; Cairns *et al.*, 2013; Mukherjee and Kearns, 2014). The flagella-dependent swimming and swarming motility are one of the well-studied phenotypes of *B. subtilis* (Mukherjee and Kearns, 2014). While swarming is achieved on solid media, the swimming motility is acquired only in the liquid growth conditions (Kearns and Losick, 2004; Senesi *et al.*, 2004). Additionally, the swarming motility is restricted to the undomesticated strain of *B. subtilis* NCIB 3610 (Kearns and Losick, 2004). Whereas, the lab domesticated strain 168 lacks the ability to swarm (Patrick and Kearns, 2009). A key characteristic of the undomesticated *B. subtilis* strain is its ability to secrete

a surfactant called 'surfactin' into its environment, which is encoded by the *srfAA* gene (Nakano *et al.*, 1991; Kearns and Losick, 2004; van Gestel, Vlamakis and Kolter, 2015). The availability of surfactin to bacteria reduces its surface tension which assists in swarming in addition to the flagellar movement on a solid surface (Kovács, Grau and Pollitt, 2017).

From the both the transcriptomics and proteomics analysis of Δ*spoVG* in M9 (Chapter 3 and 4), it was identified that the expression and production of several motility related genes and proteins including *srfAA* were reduced in the mutant. Therefore, to understand these changes in the context of the Δ*spoVG* strain, swimming and swarming motility tests were carried out in M9 using the 168, *trp*+ and NCIB 3610 strains respectively. Henceforth, the NCIB 3610 strain will be indicated as 3610 for simplicity.

7.1.2 Biofilm Formation in *B. subtilis*

Biofilms are ubiquitously found in the nature. They are made up of tightly surface-associated microorganisms that are encased in a self-produced extracellular matrix (Kolter and Greenberg, 2006). *B. subtilis* can either form biofilms on solid media or between air-liquid interface called as a pellicle. Formation of biofilms is a characteristic feature of *B. subtilis* which also serves as a model for microbial developmental.

The molecular mechanisms behind the formation of biofilms are complex and are often integrated with several regulatory pathways (Martínez and Vadyvaloo, 2014). For example, in *B. subtilis* the cell motility, biofilm formation and exopolysaccharide (EPS) synthesis are extremely complex and highly regulated phenotypes. An important operon in biofilm formation and regulation of motility in *B. subtilis* is the *tapA-sipW-tasA* operon (Chu *et al.*, 2006). In a recent study, it was shown that secreted TasA is essential for the flagellar motility in *B. subtilis* in biofilm formation (Steinberg *et al.*, 2020). Expression of the gene *swrA* is responsible for both swimming and swarming motility in *B. subtilis* (Calvio *et al.*, 2005). In addition to these, some studies in *B. subtilis* have also linked the activation of the cell motility

and biofilm formation to factors such as ribosomal proteins (Takada *et al.*, 2014), general stress response (Nagórska *et al.*, 2008), capacity of cells to produce surfactin (Connelly, Young and Sloma, 2004) and antibiotic resistance (Lai, Tremblay and Déziel, 2009).

In the transcriptomics data, it was observed that the expression of the *tapA-sipW-tasA* operon was decreased in the Δ*spoVG* mutant in minimal media supplemented with glucose (M9) (Chapter 3). The protein BslA contributes to both the architecture and the hydrophobicity of the biofilm matrix (Hobley *et al.*, 2013; Bromley *et al.*, 2015; Arnaouteli *et al.*, 2017). Expression of the *bslA* gene was decreased in the M9 transcriptomic data. Therefore, biofilm formation capacity of the mutant was tested in on both solid and liquid media with minimal media.

7.1.3 Secretion in *B. subtilis*

B. subtilis is a soil dwelling bacterium and it naturally secretes a number of proteins and enzymes into its immediate environment to degrade and metabolise complex polymers in its environment. The ability of *B. subtilis* to secrete enzymes in large amounts is exploited in the biotechnology industry and thus *B. subtilis* is known as a molecular biology and industrial 'workhorse'. The secretion system in *B. subtilis* is well-studied and comprises a variety of systems including the Sec-SRP pathway, twin-arginine (Tat) pathway and the ATP-binding cassette (ABC) transporters (reviewed in (Su *et al.*, 2020)). Of these pathways, the largest number of proteins are known to be exported by the Sec-SRP pathway. Whereas the other secretion systems are activated only in special conditions such as competence, and secrete only a limited number of proteins.

HtrA and HtrB are membrane-anchored proteases involved in protein quality control (Noone *et al.*, 2001). The expression of *htrA* and *htrB* is induced by the two-component system CssRS, which is sensitive to secretion stress (reviewed in (Yan and Wu, 2019)). The *cssRS* pathway can also be triggered by various sigma factors and growth conditions such as

the SigB, general stress response factor or heterologous expression of proteins (reviewed in Yan and Wu, 2019). The transcriptomics data showed an altered expression of secretion genes *htrA* and *htrB* in Δ*spoVG* in M9 (Chapter 3). The production of HtrA was also decreased in the mutant in M9 (Chapter 4) Also, the *cssRS* genes were down-regulated in Δ*spoVG* in M9 (Chapter 3). However, no other major secretion pathway was differentially expressed in either of the omics data. In contrast to the secretion genes, many general stress response genes were up-regulated in Δ*spoVG* in M9 (Chapter 3). However, their relationship with the down-regulated secretion genes was not clear. Therefore, experiments were set up to verify the secretion response of Δ*spoVG* which will be discussed in the sections below.

7.1.4 Growth of *B. subtilis* in Different Carbon Sources

For all the living organisms, carbon serves as the primary source of energy. However, in their natural habitat bacteria encounter a mixture of carbon sources. Therefore, bacteria have developed selective modes of uptake of carbon and its metabolism to be able to survive and compete in nutritionally limiting environments against other bacteria and microorganisms. For most bacteria including *B. subtilis*, glucose is the preferred source of carbon. *B. subtilis* can also utilize xylose, malate and arabinose as alternate sources of carbon (Fisher, 1991; Stülke and Hillen, 2000; Kleijn *et al.*, 2010; Buescher *et al.*, 2012). The alternate carbons are either converted into simpler sugars such as glucose that undergo glycolysis or are converted into intermediary compounds that feed into the glycolytic pathway, tri-carboxylic acid (TCA) cycle or other modes of carbon metabolism.

The proteomics data of Δ*spoVG* identified ‘Utilization of specific carbon sources’ as one of the significantly enriched gene sets in LB growth condition (Chapter 4). The same get set was also enriched in the transcriptomic analysis of the mutant in sporulation hour 5 (SH5) (Chapter 3). This suggested a role of *spoVG* in the uptake of specific carbon sources

for the survival of the cells in differential growth conditions. Therefore, to understand the role of *spoVG* in the uptake of alternate energy sources, the growth of Δ*spoVG* in response to different carbon sources involved in central carbon metabolism was tested. Growth curve experiments were set up in defined minimal media with different carbon sources in microtiter plates and the growth of Δ*spoVG* was compared with the WT. Sucrose, fructose, succinate, galactose, glycerol, arabinose, xylose, and myo-inositol (MI) were chosen as the alternate carbon sources.

7.2 Results and Discussion

7.2.1 Motility and Biofilm in Δ*spoVG*

7.2.1.1 The ΔspoVG cells swarm less on M9 agar

Swarming motility of the Δ*spoVG* 3610 strain was monitored on M9 plates with 0.7% agar, which showed that the mutant had a reduced swarming motility compared to the WT (Figure 7.1). The WT strains covered the plate in a uniform featureless manner, whereas the Δ*spoVG* mutant produced a channel like pattern on the surface of the agar. The complemented strain (Δ*spoVG*::*spoVG*) was able to restore the swarming deficiency.

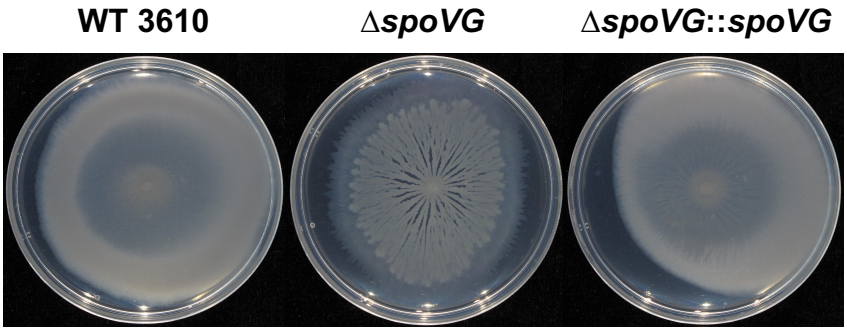


Figure 7.1: The Δ*spoVG* (3610) cells have reduced swarming motility and form distinctive "channel" like morphology on M9 agar

The M9 swarming plates of the 3610 strain of *B. subtilis* WT, Δ*spoVG* and complemented Δ*spoVG*::*spoVG* after overnight incubation.

*7.2.1.2 Δ*spoVG* has restricted growth on M9 swarming plates and lacks pigment formation*

The Δ*spoVG* mutant covered less of the swarming plate after 16 hours compared to the WT and the complemented strain (Figure 7.2 (A), upper panel). Therefore, it was speculated whether, after a longer incubation period, the Δ*spoVG* strain would also be able to reach the edge of the plate. The plates were incubated for 24 hours, by which time the WT and the complemented strain had completely covered the plate. The Δ*spoVG* strain was still restricted to where it had grown at 16 hours, but that growth was thicker (Figure 7.2 (A), lower panel). Also, after the prolonged incubation of the swarming cells for up to 24 hours, it was observed that the Δ*spoVG* mutant strain lacked a pink pigment that was present for both the WT and the complemented strain (Figure 7.2 (B)).

B. subtilis is known to produce a variety of pigments as primary and secondary metabolites (Reva *et al.*, 2004). The coloured pigments include black melanin-like pigment (Hinojosa-Rebollar *et al.*, 1994, 1995), the characteristic brown coloured pigments of the spores (Barnett *et al.*, 1983), and one or more red pigments (Uffen and Canale-Parola, 1972; Yan *et al.*, 2003; Reva *et al.*, 2004). The red pigment has been identified as pulcherrimin that is encoded from the *yvmC-cypX* operon and functions as a siderophore i.e., it chelates iron (Uffen and Canale-Parola, 1972; Tang *et al.*, 2006).

Therefore, it was speculated that the pink pigment found on the swarming plates was pulcherrimin and the Δ*spoVG* mutant is deficient in the production of pulcherrimin. Re-analysis of the transcriptomics data also identified the decreased gene expression of the *yvmC-cypX* operon in Δ*spoVG* in M9 (Chapter 3). However, the proteins products were not differentially produced in the proteomics data. The gene set enrichment analysis (GSEA) in Δ*spoVG* in M9 transcriptomic data also identified the significant enrichment of the gene set *Acquisition of iron* and the significant enrichment of the KEGG pathway *Biosynthesis of siderophore group*

nonribosomal peptides (Chapter 3). Hence, it was speculated that *spoVG* has a role in pulcherrimin production in M9.

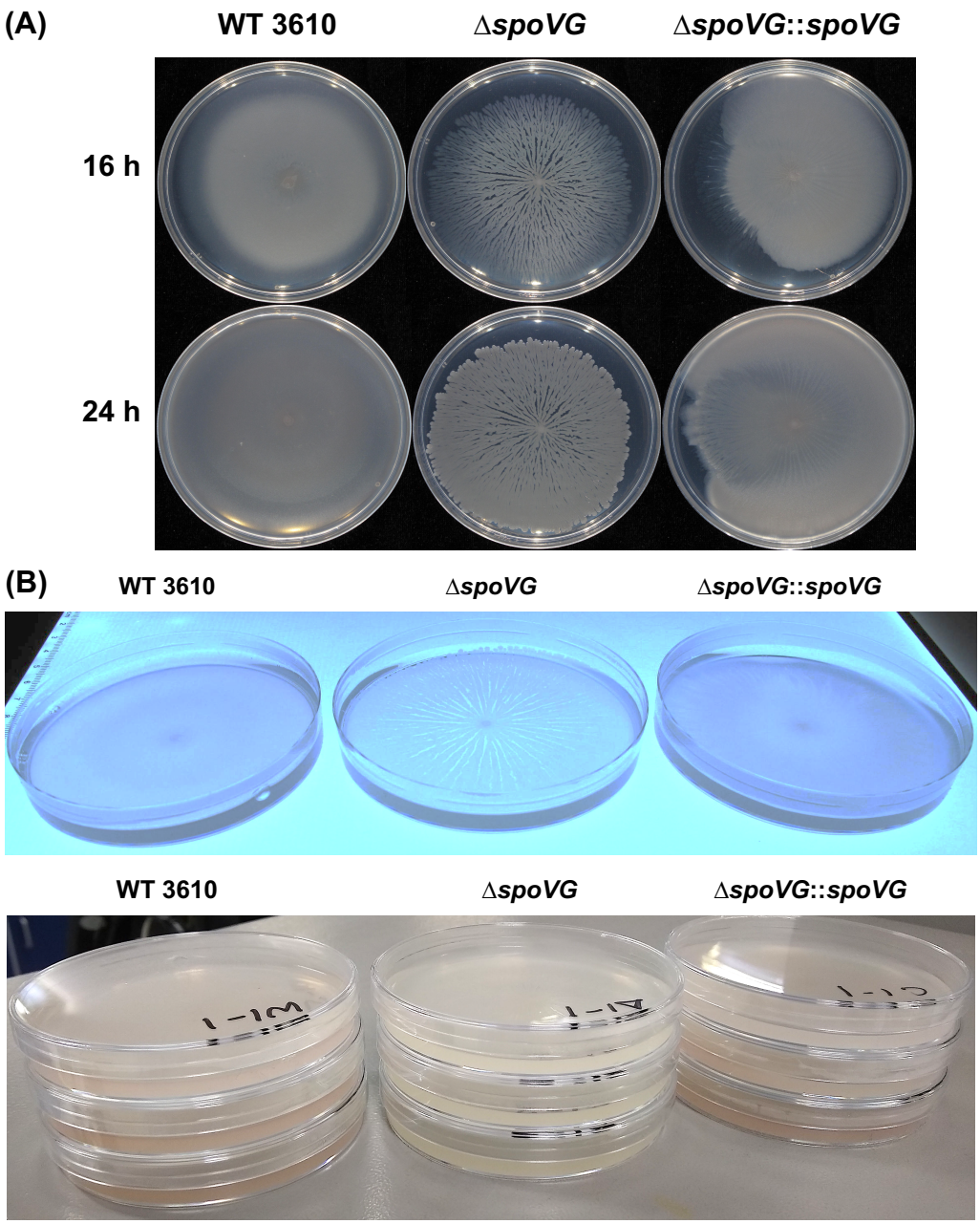


Figure 7.2: The Δ*spoVG* (3610) swarming cells have restricted growth and do not produce pink pigment on M9 agar plates after 24 hours

(A) The swarming plates of the 3610 strain of WT, Δ*spoVG* and Δ*spoVG*::*spoVG* incubated up to 24 hours. **(B)** The plates in the upper panel show the top view of the swarming plates incubated for 24 hours. The lower panel shows the side view of the stacked technical triplicates of the same plates with no pink pigment in the Δ*spoVG* plates.

*7.2.1.3 Biofilms of Δ*spoVG* are smaller and have more "wrinkles" on MSgg agar*

Since the Δ*spoVG* 3610 cells showed a restricted swarming ability and formed distinct channel like features on M9 agar, it was speculated whether the phenotype was reproducible in other minimal media. To test this, the 3610 strains of the WT, Δ*spoVG* and the complemented strain were tested in the lab of Prof. Nicola Stanley-Wall for their capacity to form biofilms on MSgg agar as an alternative minimal media.

The results from the Stanley-Wall lab showed a restricted growth of Δ*spoVG* 3610 on MSgg agar after 5 days of incubation (Figure 7.3). The restricted biofilms of Δ*spoVG* also had more "wrinkles" and "channel-like" architectures compared to the WT 3610. The complemented strain restored the biofilm phenotype defect. These observations were comparable with the previously observed phenotype of Δ*spoVG* on M9 agar (Figure 7.1 and Figure 7.2). Thus, it was concluded that *spoVG* has a role in the swarming motility and biofilm formation in *B. subtilis* on solid minimal media.

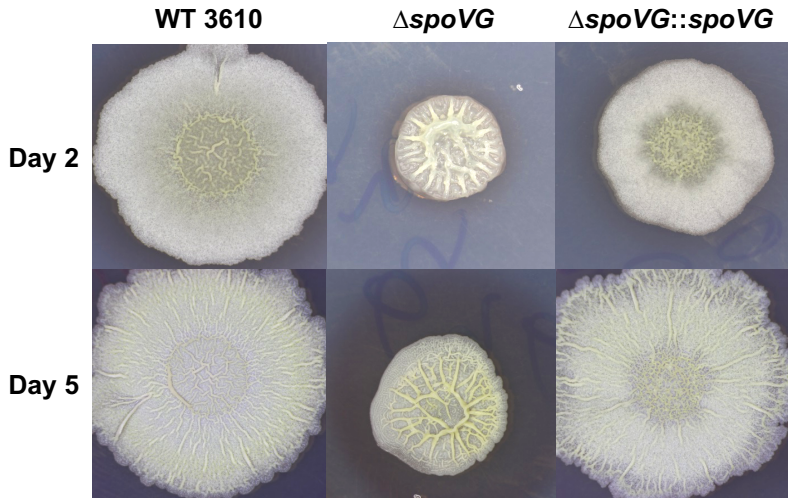


Figure 7.3: (*Picture courtesy: Nicola Stanley-Wall Lab*) The biofilms of the Δ*spoVG* (3610) are smaller and have more "wrinkles" on MSgg agar

The 3610 strain of the WT, Δ*spoVG* and *spoVG*::*spoVG* were grown on MSgg and inoculated on MSgg agar (1.5%) (Arnaouteli *et al.*, 2019). The cells were incubated for up to 5 days and imaged on day 2 (upper panel) and day 5 (lower panel) post inoculation.

7.2.1.4 Swimming of ΔspoVG is reduced on M9 agar

From the swarming experiments it has been established that the Δ*spoVG* cells are deficient in motility in M9. However, the reason behind this reduced swarming phenotype was not clear from these studies. To understand if this deficiency of motility in Δ*spoVG* was solely due to flagellar impairment or the inability of the cells to secrete surfactin, swimming motility tests in M9 agar were set up. If the cells lacked only in the surfactin production, the Δ*spoVG* cells should be motile on swimming plates equivalent to the WT but if the flagella were compromised, the mutant should also show impairment in its capacity to swim. Therefore, swimming motility tests for the 168 strains of the WT, Δ*spoVG* and the complement were set up on M9 agar and monitored for up to 16 hours.

The results revealed that the Δ*spoVG* 168 had a reduced swimming phenotype compared to the WT 168 (Figure 7.4). This observation was comparable to the reduced swarming ability of the Δ*spoVG* 3610 in M9 (Figure 7.1). Also, the complemented strain was able to restore the WT phenotype.

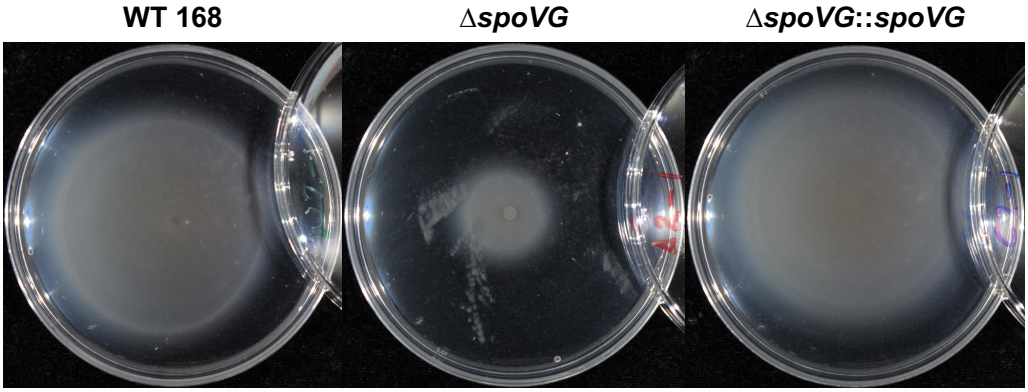


Figure 7.4: The swimming phenotype of Δ*spoVG* (168) is reduced on M9 agar

The swimming phenotype of the 168 strains of the WT, Δ*spoVG* and Δ*spoVG*::*spoVG* were tested on M9 agar (0.26%). Cells were incubated at 37˚C for ~ 16 hours before imaging.

Interestingly, from the proteomics data it was observed that the production of the FliK and FlhG proteins of the *fla*/*che* operon were increased in Δ*spoVG* in M9 (Chapter 4). However, no other flagellar proteins were identified in the proteomic data. Also, no flagellar genes were found to be differentially expressed in the transcriptomic analysis of Δ*spoVG* in M9 (Chapter 3).

FliK and FlhG proteins are encoded from the *fla*/*che* operon which produces flagellar proteins and facilitates in the swimming and swarming motility in *B. subtilis* (Guttenplan, Shaw and Kearns, 2013; Mordini *et al.*, 2013). A major activator of the *fla*/*che* operon is the master regulator SwrA (Patrick and Kearns, 2012; Guttenplan, Shaw and Kearns, 2013; Mordini *et al.*, 2013). SwrA also antagonizes the SlrA/SinR/SlrR system that reduces the *fla*/*che* operon transcripts levels (Kobayashi, 2007; Mordini *et al.*, 2013). Also, FlhG is an ATPase which activates the GTPase activity of FlhF which is necessary for the activity of SigD (sigma factor involved in the chemotaxis and motility of the bacteria) (Carpenter, Hanlon and Ordal, 1992; Guttenplan, Shaw and Kearns, 2013; Schuhmacher, Thormann and Bange, 2015). FliK is a regulatory proteins which regulates the length of the flagella in *B. subtilis* (Courtney, Cozy and Kearns, 2012; Mukherjee and Kearns, 2014).

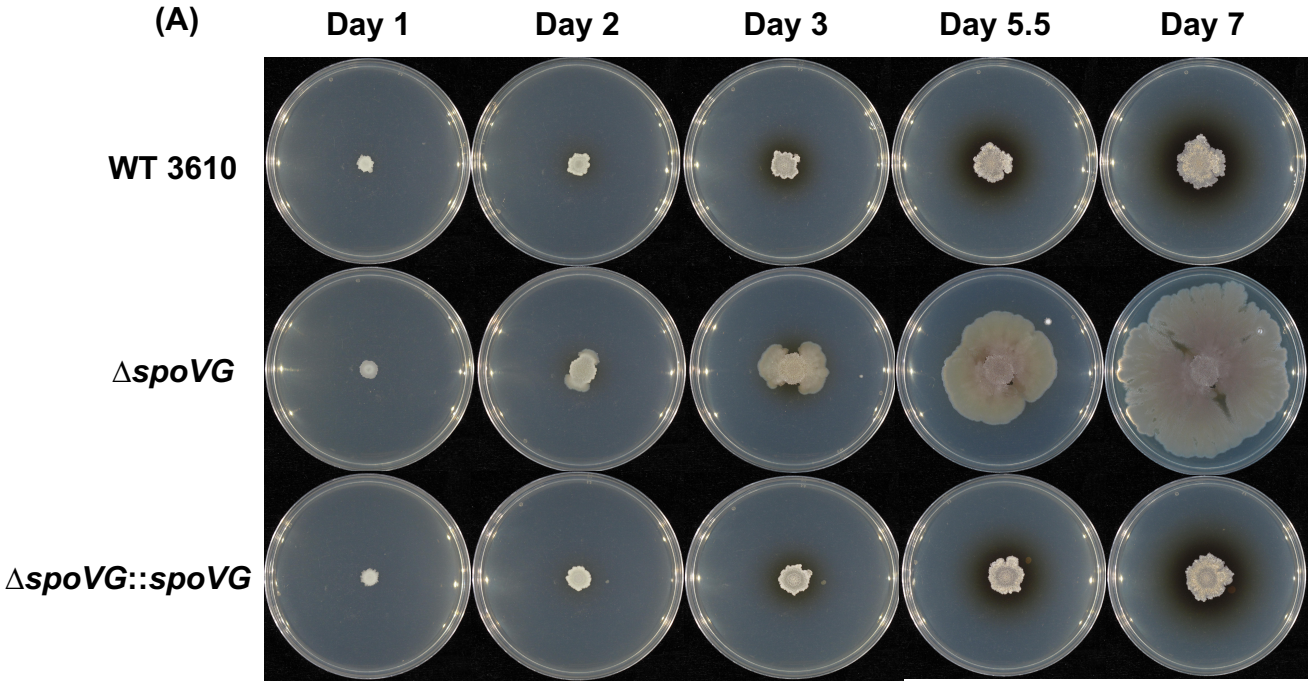
Therefore, a lack of direct link between the increased motility proteins from the proteomic experiments (Chapter 4) and the decreased swarming and swimming motility in Δ*spoVG* in minimal media was puzzling. Perhaps, *spoVG* plays a complex role in regulating the flagellar motility in *B. subtilis*, the mechanism of which remains elusive. Further molecular and phenotypic experiments would be needed to uncover the participation of *spoVG* in the motility of *B. subtilis.*

*7.2.1.5 Δ*spoVG *forms expanded biofilms and produces less pigment on LB agar*

Since Δ*spoVG* showed reduced motility and biofilm formation on minimal media, the phenotypes were also tested in the rich media LB. The swimming and swarming motility of the Δ*spoVG* mutant was tested in the 168 and the 3610 strains respectively. The results from these experiments did not show any difference in the motility of the mutant compared to the WT (data not shown).

Further, the biofilm formation capacity of Δ*spoVG* was also tested on LB agar using the 3610 strain. For biofilm experiments, it has been reported that addition of glycerol and manganese (GM) in the growth media enhances the biofilm formation in *B. subtilis* in both the solid and liquid media (Shemesh and Chai, 2013). Therefore, for this study LB agar supplemented with glycerol and manganese (LBGM) were used to monitor the formation of biofilms in Δ*spoVG*.

The results from this experiment revealed that the Δ*spoVG* (3610) mutant formed an expanded biofilm compared to the WT on LBGM agar (Figure 7.5 (A)). The complemented strain also restored the WT phenotype. In addition to these observations, it was also noticed that Δ*spoVG* (3610) produced less brown pigment compared to the WT and the complemented strains (Figure 7.5 (B)). The enhanced biofilm formation of Δ*spoVG* on LBGM agar was opposite to that observed in the minimal media (MSgg) agar where the mutant had a decreased biofilm formation compared to the WT. However, the pigment formation of Δ*spoVG* was decreased in both the LB and MSgg. These opposite results of the biofilm formation in LB vs M9 in Δ*spoVG* were intriguing and suggested an indirect involvement of SpoVG in a complex regulatory network. In the future, it will be interesting to understand these differences in detail to study the role of SpoVG in media-specific biofilm formation.


(A)
Day 1
Day 2
Day 3
Day 5.5
Day 7
WT 3610
ΔspoVG
ΔspoVG::spoVG

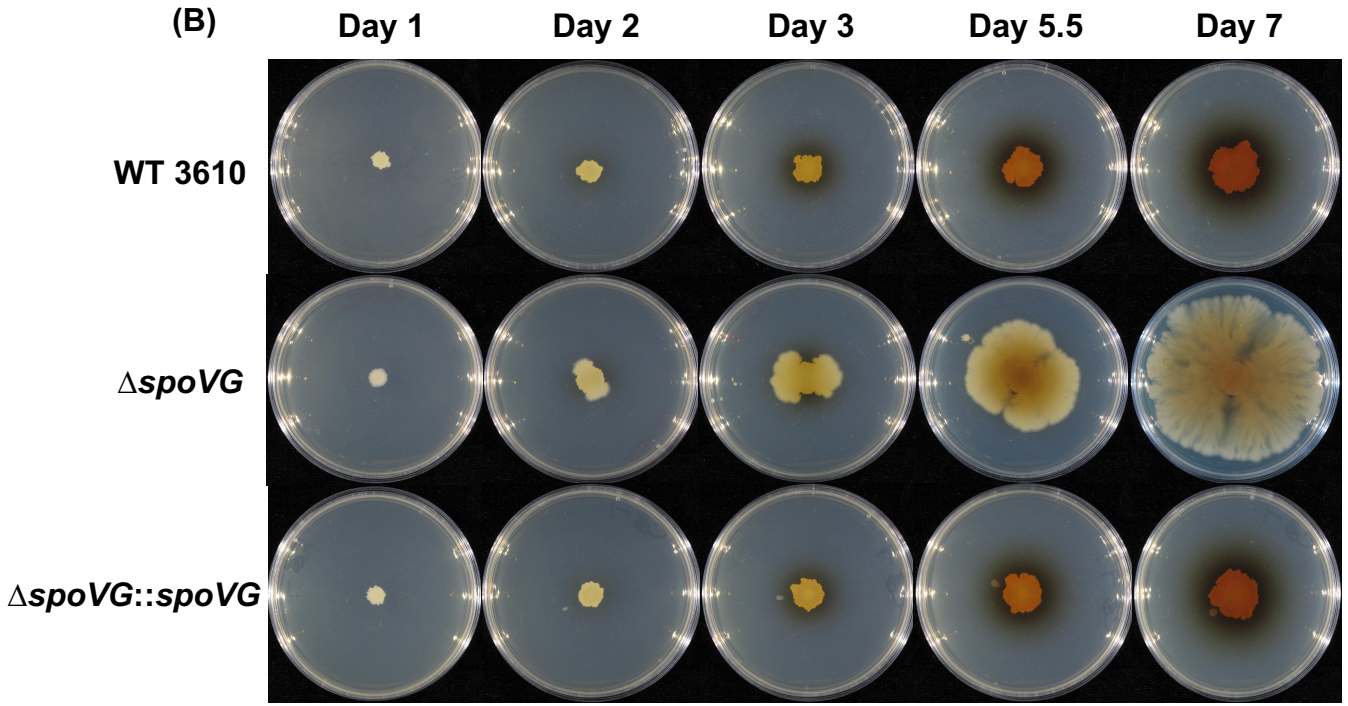


Figure 7.5: The biofilms of Δ*spoVG* (3610) have an expanded morphology and produce less brown pigment on LBGM agar

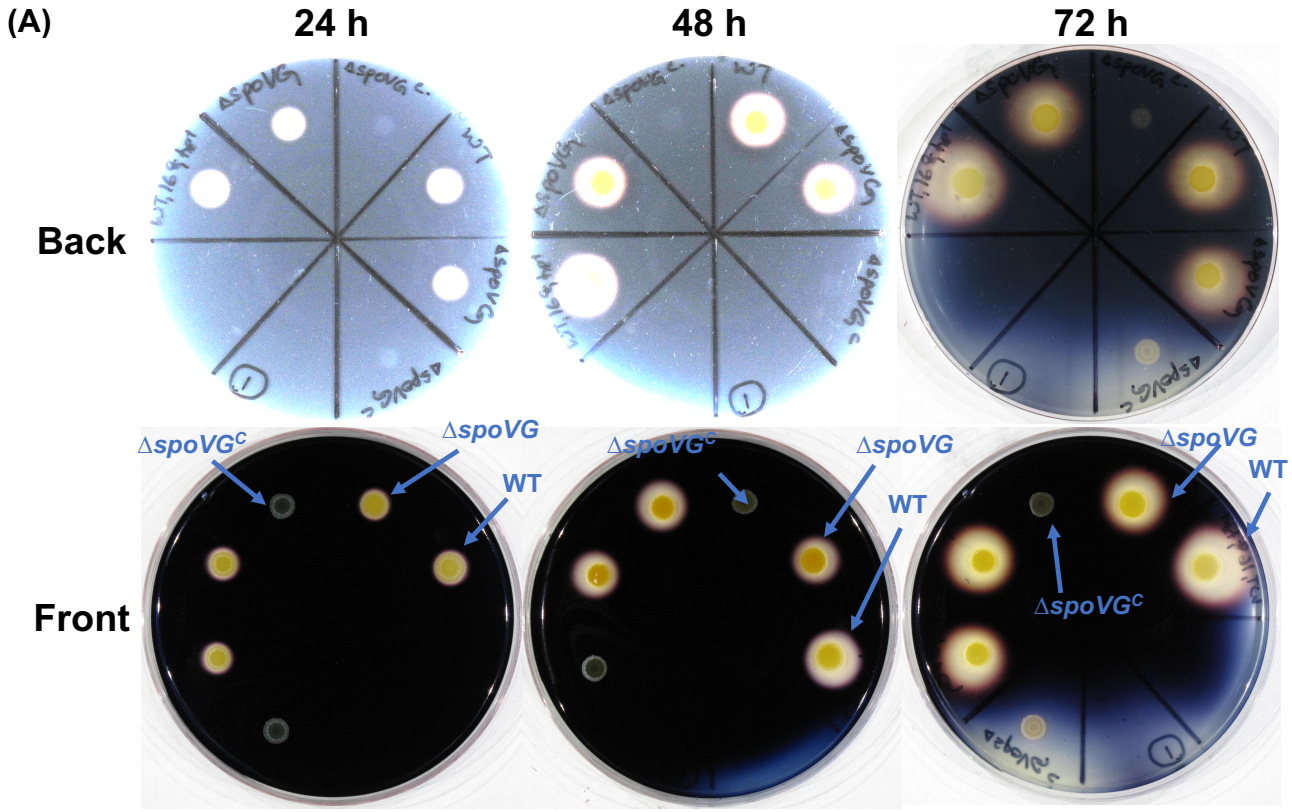
Imaging of the biofilms at different times over 7 days of the 3610 strain of WT, Δ*spoVG* and complement Δ*spoVG*::*spoVG* on LBGM agar (1.5%) plates. **(A)** The top view of the plates showing the expanded biofilms of the Δ*spoVG*. The colonies of Δ*spoVG* start to expand from day 2 of inoculation. **(B)** The bottom view of the same as plates as (A) showing the difference in the pigment formation in the Δ*spoVG*. A noticeable change in colour in Δ*spoVG* can be observed after day 5.

7.2.2 Secretion in Δ*spoVG*

*7.2.2.1 Δspo*VG *secretes less α-amylase in LB*

α-amylase is a starch degrading enzyme that is secreted by *B. subtilis* into its environment and is encoded by the *amyE* gene. To investigate if the secretion system of Δ*spoVG* is affected in LB, the ability of Δ*spoVG* to secrete amylase was tested by the iodine/starch assay using both the 3610 and the 168, strains of WT, Δ*spoVG* and the complement (Δ*spoVG*::*spoVG*).

The results from this experiment showed that Δ*spoVG* (168) produced reduced zones of clearing on starch plates compared to the WT, which directly correlates with reduced production of amylase (Figure 7.6 (A)). The complemented strain showed no zone of starch clearance due to the interrupted *amyE* locus. Similar results were observed for the 3610 strain of the Δ*spoVG* mutant and the complemented strain (Figure 7.6 (B)). These evidence suggested an impairment in the secretion of α-amylase in the absence of *spoVG* suggesting a role of SpoVG in the secretion system of *B. subtilis*. A proteomic analysis of the secretome of Δ*spoVG* would further enable to quantify these changes in the secretome quantitatively.


(A)
24 h
48 h
72 h
Back
Front
ΔspoVGC
ΔspoVG
WT
ΔspoVGC
ΔspoVG
WT
ΔspoVG
WT
ΔspoVGC

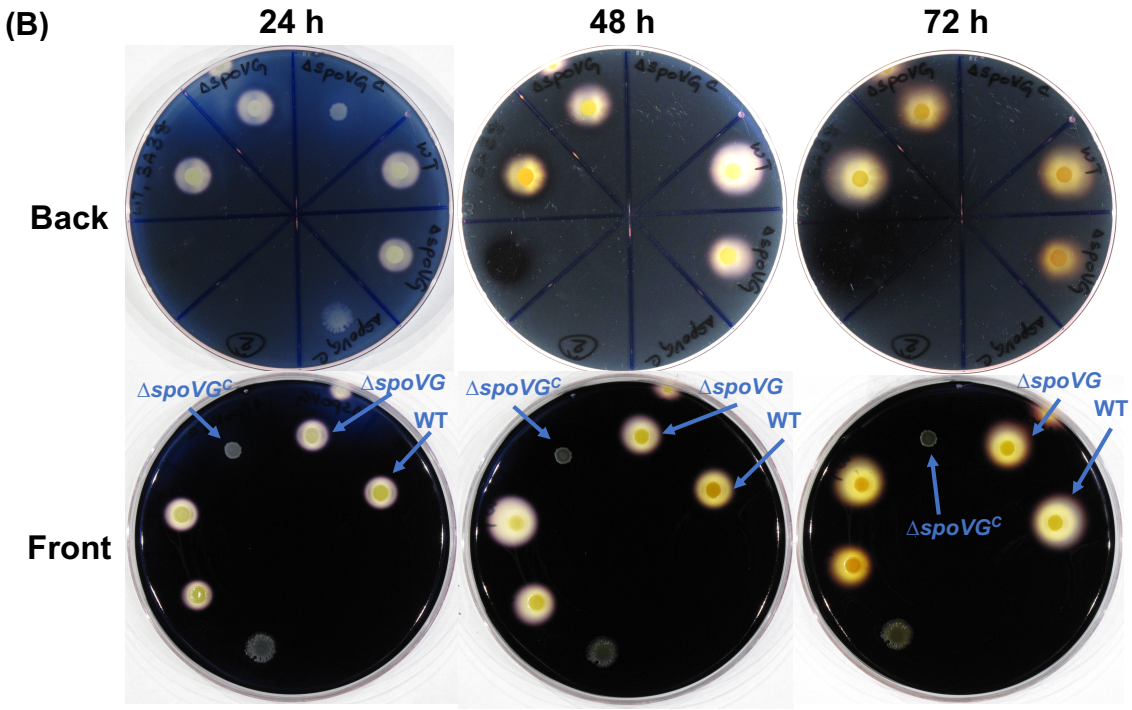


Figure 7.6: Secretion of α-amylase in Δ*spoVG* (168 and 3610) is decreased in LBA-starch

The starch-iodine test showing the secretion of α-amylase from the **(A)** 168 **(B)** 3610 (3A38) strains of *B. subtilis* on LBA-starch (1%). The halo generated around the colonies show the digested starch by the α-amylase. The complement strain Δ*spoVG*::*spoVG* (Δ*spoVG*^{C}) does not secret α-amylase due to the disrupted *amyE* locus by the complementation of *spoVG* at the locus. Plates were incubated up to 72 hours and exposed to iodine in every 24 hours interval before imaging.

*7.2.2.2 Secretion of total proteins is reduced in Δ*spoVG* in LB*

In addition to the α-amylase starch assay, the secreted proteins were precipitated and were analysed by SDS-PAGE. The Δ*spoVG* strain appeared to secrete fewer proteins into its extra-cytoplasmic environment when grown in LB (Figure 7.7). This observation was in line with the observation from the M9 omics data where secretion genes were down-regulated in the Δ*spoVG* and also the previous secretion experiments. The complemented strain restored the phenotype of the WT.

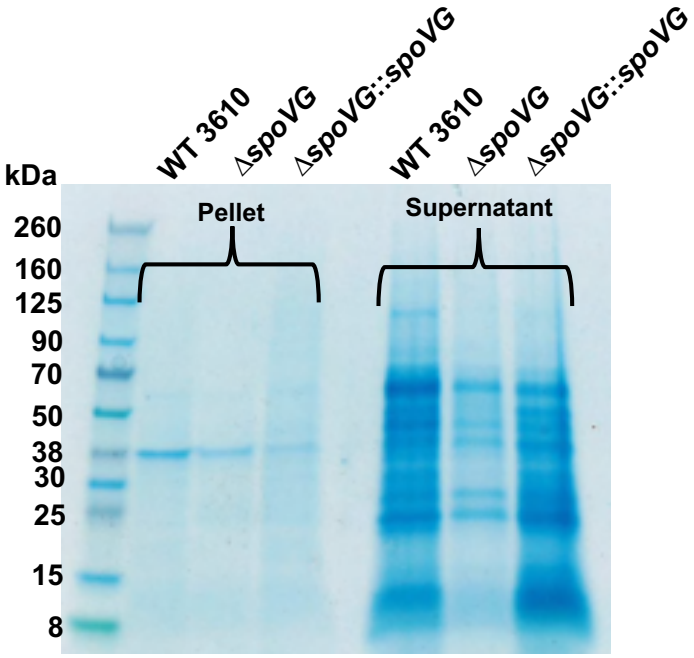


Figure 7.7: Secretion of total proteins in Δ*spoVG* (3610) is reduced in LB

The 3610 strains of the WT, Δ*spoVG* and Δ*spoVG*::*spoVG* were grown in LB till OD_{600} of 1.3. The total supernatant (with secreted proteins) was precipitated with trichloroacetic acid (TCA) overnight. The total precipitated supernatant proteins and 18% of the cell pellets were loaded on the gel.

These observations together with the α-amylase assay indicated a potential role of *spoVG* in the secretion system of *B. subtilis*. In the absence of the *spoVG*, the cells seem to be responding in a stress like situation which in turn makes the cell incapable of maintaining the normal cellular processes such as motility, secretion, and biofilm formation.

7.2.3 Growth of Δ*spoVG* in Alternate Carbon Sources

To test the ability of Δ*spoVG* to grow with different carbon sources, growth experiments were set up in PM supplemented with the carbon sources glucose (0.3%), sucrose (0.3%), fructose (0.3%), succinate (0.3%), galactose (0.3%), glycerol (0.3%), arabinose (0.3%), xylose (0.3%), and myo-inositol (MI) (10 mM) as the sources of carbon. Pre-culture was carried out in glucose, before the cells were transferred to media containing the different carbon sources.

The growth of Δ*spoVG* was similar to the WT, though the mutant had a continued growth in sucrose (0.3%), succinate (0.3%), galactose (0.3%), glycerol (0.3%), and arabinose (0.3%) when the cells entered stationary phase (Figure 7.8). The complemented strain in each of these conditions showed comparable growth to the WT, except for succinate where the complemented strain had a growth similar to Δ*spoVG*. Milder changes in growth were observed with fructose (0.3%), xylose (0.3%) and MI (10 mM). In xylose (0.3%), the growth of Δ*spoVG* was decreased between 10-14 hours time point compared to the WT and complemented strain. Also, all of the growth differences in Δ*spoVG* were observed post the 6 hours time point. Of all the conditions, a maximal change in the growth of Δ*spoVG* was observed in galactose (Figure 7.8). To better understand these behaviours of Δ*spoVG*, further experiments would be needed.

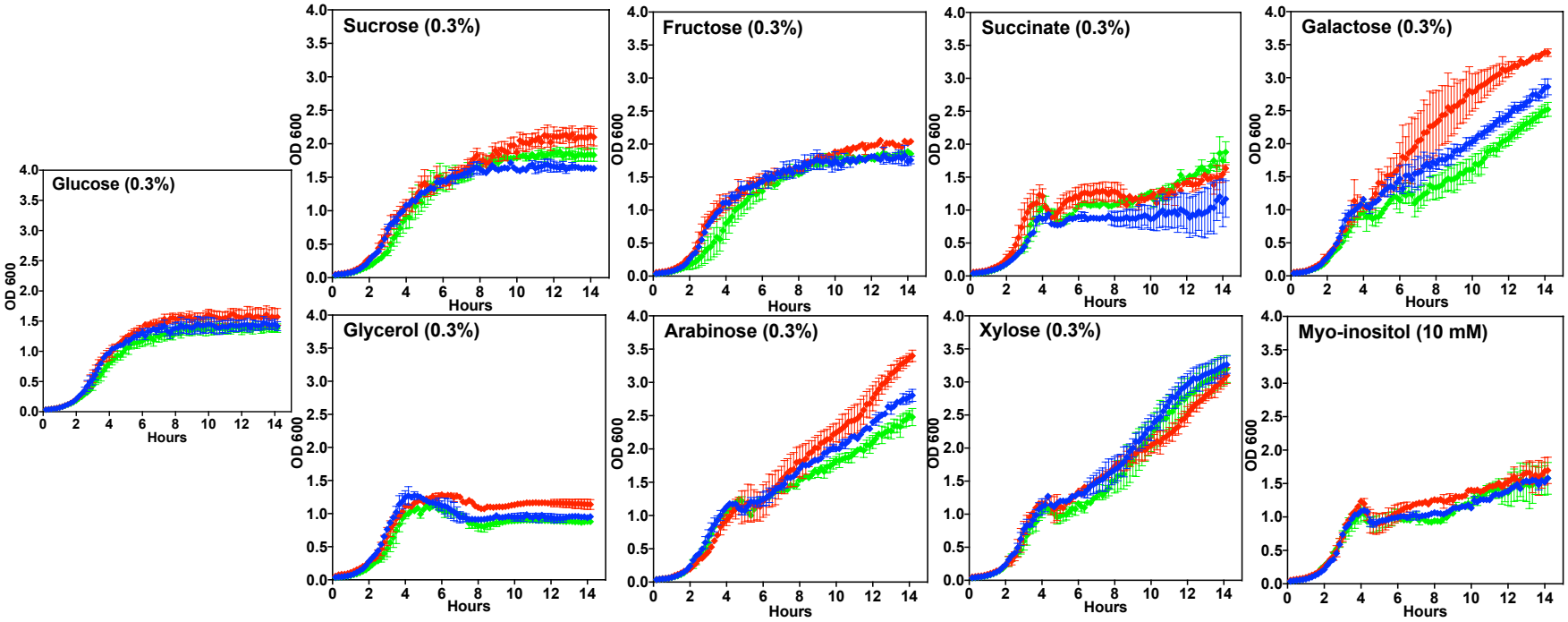


Figure 7.8: Growth of Δ*spoVG* (168) is affected in Paris media in response to various carbon sources

Each growth curve shows the growth of Δ*spoVG* in Paris media supplemented with different sources of carbon (indicated on each curve) and grown in a microtiter well for 14 hours. The WT, Δ*spoVG* and the complement strain Δ*spoVG*::*spoVG* are shown in blue, red and green lines respectively. The curves are cumulative of at least 2 biological replicates (4 biological replicates for glucose and sucrose) ± SEM.

7.3 Conclusion

In this chapter, it was shown that deletion of *spoVG* in *B. subtilis* impacts the cell in a pleiotropic fashion. Specifically, it was shown that in the absence of *spoVG* in minimal media (M9), the cells have a retarded motility that includes both swarming and swimming motility. This indicated that *spoVG* might have a role in the regulation of flagellar motility and surfactin production in *B. subtilis*.

Further, it was identified that in the absence of *spoVG*, the swarming cells in M9 lacked the production of a pink pigment. This pink pigment was speculated to be pulcherrimin. A similar reduction in the pigment formation was observed in Δ*spoVG* when it was tested for its capacity to form biofilm on LBGM agar. When tested for the capacity to form biofilms, the Δ*spoVG* formed smaller biofilms on MSgg minimal media agar and had an expanded morphology on LBGM agar. Several genes and proteins were identified from the -omics experiments that are speculated to have an effect in the biofilm formation and motility of *B. subtilis* in connection to *spoVG*. Additionally, it is hypothesised that perhaps *spoVG* may have a different mode of participation in the biofilm formation based on the growth media however, its role in the formation of pigment is independent of its capacity to form biofilms. Both molecular and phenotypic experiments would be necessary to understand these changes in detail.

The investigations in this chapter also identified impaired secretion in the Δ*spoVG* mutant. It was shown that, there was an overall decrease in the ability of Δ*spoVG* to secrete α-amylase. When tested for the total secretome, decrease in the secreted proteins in the mutant was also identified. These observations take together suggested a role of *spoVG* in the facilitating the secretion of *B. subtilis*.

When tested for the ability to grow in alternate carbon sources, the Δ*spoVG* strain showed a better growth in carbon sources such as sucrose, succinate, galactose, glycerol and arabinose when the cells entered stationary phase of growth. However, no changes in growth were observed

in Δ*spoVG* when grown in LB and M9 with glucose (Chapter 3 and 4). These observations suggested that deletion of *spoVG* likely dysregulates the regulation of the uptake of glucose as the only carbon source.

Studies of prominent RBPs such as Hfq, ProQ and CsrA have shown that deletion of the RBPs result in a pleiotropic behaviour in bacteria. Some of these behaviours include changes in biofilm formation (Jackson *et al.*, 2002; Kulesus *et al.*, 2008; Sheidy and Zielke, 2013), development of virulence (Potts *et al.*, 2019; Westermann *et al.*, 2019), bacterial motility (Kulesus *et al.*, 2008), bacterial fitness (Christiansen *et al.*, 2004), expression of sigma factors (Ding, Davis and Waldor, 2004), cell-envelope stress (Kulesus *et al.*, 2008), localisation with ribosome (Sheidy and Zielke, 2013) and uptake of sugars (Ren *et al.*, 2012). These pleiotropic characteristics are attributed to the ability of RBPs to bind to a variety of sRNAs that assist the post-transcriptional gene regulation by either binding to mRNAs or increasing the stability of sRNAs. Some RBPs also interact with a range of DNAs and protein molecules such as ribonucleases and ribosomes which expand the roles of RBPs in bacteria. SpoVG was identified as an RBPs in *L. monocytogenes* by Burke and Portnoya (2016), and they showed that deletion of *spoVG* makes the cell non-motile on soft-agar, hyper-lysozyme resistant, hypervirulent and results in up-regulated carbon metabolism genes. SpoVG is also an RBP and DNA-binding protein in *Borrelia burgdorferi* with regulatory functions (Savage *et al.*, 2018). Previously, SpoVG was shown to have regulatory functions and DNA-binding properties in *Staphylococcus aureus* (Liu, Zhang and Sun, 2016). These growing evidence point towards an important role of SpoVG in the bacteria gene regulation as pleiotropic.

From the previous experiments presented in this thesis, it was shown that SpoVG is an RBP in *B. subtilis* which interacts with a range of RNAs (Chapter 5). It was shown that SpoVG potentially binds with proteins with diverse functions, which were similar to the functions identified with RNAs bound to SpoVG (Chapter 6). The transcriptomics and proteomics analysis of Δ*spoVG* identified hundreds of genes and proteins perturbed in the

mutant and significant enrichment of several biological pathways (Chapters 3 and 4). From these observations, it is speculated that SpoVG in *B. subtilis* might have a role similar to many of the known RBPs. Although, its role as a global regulator of gene expression in *B. subtilis* remains elusive.

In summary, this chapter identified changes in the motility, biofilm formation, secretion, and growth in alternate carbon source in the absence of *spoVG* in *B. subtilis*. These changes are speculated to be due to the capacity of SpoVG to bind to a wide range of RNAs (including both protein coding and non-coding RNAs) and proteins. Nevertheless, further investigations would be needed to understand the precise mode of these changes that are mediated by SpoVG.

Chapter 8 Discussion

To sense and survive in a constantly changing environment, bacteria need to rapidly modulate their gene expression. The post-transcriptional gene regulation including non-coding RNAs (ncRNAs), small non-coding RNAs (sRNAs) and RNA-binding proteins (RBPs) provides one such platform for bacteria to fine-tune their gene expression. Several studies have investigated the roles of ncRNAs and sRNAs as versatile regulators in post-transcriptional gene regulation in prokaryotes. However, the roles of RBPs in bacteria are still far from complete.

RBPs are ubiquitously found in each domain of life. They have several crucial and highly specific roles that make them important for the survival and the fitness of many bacteria. RBPs also have a wide range of interacting RNA partners that comprises of RNAs such as transfer RNAs (tRNAs), ribosomal RNAs (rRNAs) and the regulatory RNAs that include sRNAs and ncRNAs (Quendera *et al.*, 2020; Ng Kwan Lim *et al.*, 2021). The role of RBPs dependent on the regulatory RNAs include the facilitator of sRNA-mRNA interaction in the post-transcriptional gene regulation. Additionally, RBPs also help in the recruitment of ribonucleases (RNases) in sRNA-based gene regulation. Furthermore, RBPs influence the activity of many sRNAs that affect the translational efficiency of the target mRNAs which in turn influence the properties of cells such as virulence, competence, motility, biofilm formation, consumption of alternate carbon sources or secretion. RBPs may also provide stability to the interacting sRNAs or an mRNAs, thereby providing stability against transcriptional degradation. Moreover, various RBPs also have regulatory functions that are independent of their RNA regulators, such as in transcription termination and metabolic enzyme activities (reviewed in Ng Kwan Lim *et al.,* 2020).

In the model bacterium *E. coli,* the RBP Hfq has been studied as the global RNA chaperone. It was mainly known to stabilise the interactions between sRNA and their target mRNAs interactions during post-transcriptional gene

regulation. However, new studies have challenged these ideas. For instance, recently it was shown that Hfq is not essential for the base-pairing of sRNA DsrA with its target mRNA *rpoS*, nor it is required for the regulatory activity of DsrA (Kim *et al.*, 2019). Nevertheless, Hfq still significantly enhances the regulatory effect of DsrA by facilitating its base-pairing with *rpoS.* Another recent finding has also shown Hfq independent recruitment of RNA degradasome complex to its cognate target mRNA in the absence of Hfq (Baek *et al.*, 2019), which challenged the general understanding of the role of Hfq in sRNA-mRNA based gene regulation. Furthermore, deletion studies of Hfq have identified increased errors in the protein synthesis and defect in ribosome assembly (sRNA-independent process) in the absence of Hfq (Andrade *et al.*, 2018). Further investigation by Andrade *et al.*, 2018 showed the role of Hfq in the maturation of 16S rRNA by binding to 17S rRNA in *E. coli*. Taken together, these evidence have uncovered new roles of Hfq and prompted to investigate the functionality of RBPs from a different angle.

With recent genome-wide and high-throughput studies, RBPs such as CsrA and ProQ have also emerged as major RBPs with regulatory functions in the bacterial RNA-based gene regulation and beyond (Quendera *et al.*, 2020; Ng Kwan Lim *et al.*, 2021). However, the majority of studies pertaining to the RBPs in bacteria, have been limited to the *Enterobacteriaceae E. coli* and *Salmonella* (Quendera *et al.*, 2020; Ng Kwan Lim *et al.*, 2021). Little is known about the roles of these RBPs in the Gram-positives. Additionally, many of the Gram-positive bacteria either do not code for the homologues of these RBPs (Jousselin, Metzinger and Felden, 2009), or these RBPs do not participate in the same way (including in the model Gram-positive bacteria *B. subtilis*) as in their homologues in the *Enterobacteriaceae* (Hämmerle *et al.*, 2014; Mars *et al.*, 2015; Rochat *et al.*, 2015). Cumulatively, these evidence indicate the existence of alternative post-transcriptional gene regulation strategies in the Gram-positive bacteria. These alternate mechanisms may either include undiscovered RBPs with

global functions, or RBPs with condition-dependent roles, or existence of sRNA-mRNA base-pairing without the need of RBP.

Recent investigations to explore the roles of RBPs in post-transcriptional gene regulation in the Gram-positive bacteria such as *S. aureus*, *L. monocytogenes* and *B. burgdorferi* have identified a protein SpoVG with *in vitro* RNA-binding activity (Burke and Portnoy, 2016; Liu *et al.*, 2016; Savage *et al.*, 2018). From these studies, SpoVG has been shown to participate in RNA based gene regulation which affects cells in a pleiotropic manner, similar to what is observed in many of the Hfq deletion strains. The absence of SpoVG in these bacteria results in increased virulence, decreased motility, increased lysozyme activity, changes in the carbon metabolism and ability of the bacteria to colonise their host species (Burke and Portnoy, 2016; Liu *et al.*, 2016; Savage *et al.*, 2018).

SpoVG was initially identified in *B. subtilis* as a stage V sporulation protein with its role as a negative regulator of asymmetric division in sporulation. *spoVG* is a highly conserved gene across bacterial species and has a constitutive expression in 104 growth conditions in *B. subtilis* (Nicolas *et al.*, 2012). Nevertheless, the role of SpoVG as an RBP in *B. subtilis* was not known. Furthermore, the homologue of Hfq in *B. subtilis* does not participate in the regulation of sRNA-mRNA interaction in the bacteria and *B. subtilis* lacks a global regulator of post-transcriptional gene expression. Based on the studies in *S. aureus*, *L. monocytogenes* and *B. burgdorferi* it was hypothesised that SpoVG is an RBP in *B. subtilis* and is a global post-transcriptional gene regulator.

Therefore, the aims of this thesis were to explore the RNA-binding activity of SpoVG in *B. subtilis*, identify the effect of the deletion of *spoVG* on the gene expression and protein profile of the cell and functionally characterise SpoVG in *B. subtilis*. To address these aims, a multi-omics approach was adopted. Finding a global regulator of post-transcriptional gene regulation in the model organism *B. subtilis* would help us to broaden our understanding of the gene regulation in many other Gram-positive bacteria and help to decipher the fine-tuning mechanisms involved in the

pathogenicity, motility, sporulation, biofilm formation, carbon metabolism, virulence, stress response, susceptibility to antibiotics and many more.

8.1 Investigating the Global Transcriptome of Δ*spoVG*

Chapter 3 presents the first global transcriptomic analysis of the Δ*spoVG* in *B. subtilis* 168, using the RNAtag-Seq. The gene expression profile of the mutant in three growth conditions; minimal media supplemented with glucose (M9), sporulation hour 2 (SH2) and sporulation hour 5 (SH5) were investigated. The differential gene expression analysis identified 444, 161, and 227 statistically significantly differentially expressed genes (DEGs) in M9, SH2 and SH5. While 303 out of 444 DEGs in M9 were down-regulated in the mutant, 101 and 131 DEGs were up-regulated in SH2 and SH5 respectively. Additionally, 32, 7 and 19 significantly enriched putative ncRNAs were also identified in M9, SH2 and SH5 respectively. These results suggested a functional role of *spoVG* in each of the media investigated in this study.

To characterise the genes identified in Δ*spoVG*, functional analysis of the identified genes via gene set enrichment analysis (GSEA) was carried out. Gene sets represent biological processes. Results from GSEA identified 12 significantly enriched gene sets in M9, followed by in 3 in SH2 and 8 in SH5. The most significantly enriched biological processes in M9 had functions in *Biosynthesis of antibacterial compounds*, *General stress proteins (controlled by SigB)* and Miscellaneous metabolic pathways. No gene set was commonly significantly enriched across all the three growth conditions. Nonetheless, Biosynthesis/acquisition of nucleotides was enriched in both M9 and SH2, *Genetic competence* was common in SH2 and SH5 and two gene sets, *General stress proteins (controlled by SigB)* and Quorum sensing were common between M9 and SH5. Identification of the commonly significantly enriched gene sets indicated a common role of *spoVG* in these growth conditions. Interestingly, 49 out of the 50 genes in *Biosynthesis of antibacterial compounds* in M9 were down-regulated and 127 out of 151

genes involved in *General stress proteins (controlled by SigB)* were up-regulated in Δ*spoVG* in M9. This indicated a role of *spoVG* in the biosynthesis of antimicrobial compounds and stress response. Additionally, biological processes such as Toxin, antitoxins and immunity against toxins, Cell envelope stress proteins (controlled by SigM, V, W, X, Y), Quorum sensing, *Acquisition of iron* and *Biosynthesis/acquisition of nucleotides* were also significantly enriched in M9. The significant enrichment of genes in the process of Motility and chemotaxis, Biosynthesis/acquisition of amino acids, Sporulation proteins, Utilization of specific carbon sources, Utilization of lipids and *Genetic competence* in the sporulation conditions suggested a role of *spoVG* in the tight regulation between sporulation and competence.

Furthermore, characterisation of DEGs using KEGG pathways enrichment analysis revealed three significantly enriched KEGG pathways in M9,10 significantly enriched KEGG pathways in SH2 and one KEGG pathway (*Inositol phosphate metabolism*) in SH5. No KEGG pathway was common between in any of the growth conditions. Biological processes such as *Biosynthesis of antimicrobial compounds*, nonribosomal peptide structures, and secondary metabolites in M9 were significantly enriched in both the GSEA and KEGG pathways analysis of Δ*spoVG*. These observations indicated a role of *spoVG* in the synthesis of antimicrobial peptides in *B. subtilis*. Other commonly significantly enriched biological processes in sporulation conditions via GSEA and KEGG pathway analysis comprised of amino acid metabolism, translation, biosynthesis of siderophore groups, competence, quorum sensing, acquisition of nucleotides and uptake of alternate carbon sources. Taken together, the similar outcomes of differential analysis and the functional characterisation of the genes in Δ*spoVG* validated the outcomes of these analyses and suggested new roles of *spoVG* not only in sporulation conditions but also in minimal media.

The analysis of the sigma factor regulons in response to the significantly enriched DEGs also showed changed across all the growth conditions. In M9 all the sigma factor regulons except SigV (extracytoplasmic function (ECF)) and SigM (ECF, high salt concentrations) were changed in response

to Δ*spoVG*. The SigI regulon in M9 was the most affected regulon where ~44% genes (all down-regulated) were affected due to Δ*spoVG*. The genes of SigI sigma factors are involved in iron metabolism (Ramaniuk *et al.*, 2018). Interestingly, in SH2, all the affected sigma factor regulons comprised of up-regulated DEGs except the SigA (household) regulon. Furthermore, in SH2 maximal changes were recorded in the SigF (early forespore-specific sporulation) regulon where ~7% genes were affected due to Δ*spoVG*. In sporulation hour 5, the maximal changes (~13%) were recorded in SigL where all the DEGs were down-regulated.

Therefore, it is concluded that the deletion of *spoVG* affects the transcriptome of *B. subtilis* in M9, SH2 and SH5 and affects the regulation of various genes associated with different biological processes. This further suggested a divergent regulatory role of *spoVG* in modulating the gene expression under these growth conditions.

8.2 Exploring the Global Proteome of Δ*spoVG*

Further, in order to investigate the changes in the proteome levels due to deletion, the first global proteomic quantification of the Δ*spoVG* strain of *B. subtilis* was performed in nutrient rich lysogeny broth (LB), M9, SH2 and SH5 media (Chapter 4). The differential analysis of the proteins in Δ*spoVG* identified 92, 110, 710, 556 statistically significantly differentially produced proteins (DPPs) in LB, M9, SH2 and SH5 respectively. These changes accounted for approximately 3.8%, 4.6%, 29.4% and 23% of the total proteins identified in the proteomic quantification of LB, M9, SH2 and SH5 respectively. These observations indicated that SpoVG has some functional roles in each of the growth conditions tested, with most activity in SH2.

The GSEA analysis of the identified proteins revealed three significantly enriched gene sets in each of LB and M9. Whereas, only one significantly enriched gene set was identified for both SH2 (i.e., Translation) and SH5 (i.e., Sporulation proteins). No gene sets were commonly affected in all the four conditions. The significantly enriched processes in LB comprised of

Utilization of specific carbon sources and *Heat shock proteins* and processes such as the *General stress proteins (controlled by SigB)*, *Biosynthesis of antibacterial compounds* and Miscellaneous metabolic pathways were significantly enriched in M9. Comparison of the GSEA analysis of the M9 transcriptomic data compared to its corresponding proteomic data of Δ*spoVG* unveiled three gene sets, i.e., *Biosynthesis of antibacterial compounds*, *General stress proteins (controlled by SigB)*, and *Miscellaneous metabolic pathways* that were commonly significantly enriched. These observations further indicated a role of SpoVG in the *Biosynthesis of antibacterial compounds* in *B. subtilis*. Among the sporulation conditions, no gene set was common between the -omics dat in SH2. Nonetheless, the gene set Sporulation proteins was significantly enriched in both the -omics data of SH5. This observation suggested a role of SpoVG in sporulation.

The KEGG pathway enrichment analysis of the DPPs discovered two, one, two and three significantly enriched KEGG pathways in LB, M9, SH2 and SH5 respectively. The KEGG pathway *Nonribosomal peptide structures* was the only commonly significantly enriched KEGG pathway across all the growth conditions. Enrichment of *Ribosome and RNA polymerase* in SH2 was comparable to the identification of Translation in GSEA in the proteomic analysis of Δ*spoVG* in SH2. Additionally, *Ribosome* was also a significantly enriched KEGG pathway in the transcriptomic analysis of Δ*spoVG*. These observations strongly suggested a potential involvement of SpoVG in ribosome assembly and translation in *B. subtilis* during sporulation.

The impact of the absence of SpoVG was also analysed in the regulation of sigma factor regulons. The absence of SpoVG affected sigma factor regulons across all the growth conditions i.e., LB, M9, SH2 and SH5. Out of the 14 sigma factor regulons analyses in this study, 13 regulons were affected in both the SH2 and SH5 except SigI. Changes were also recorded in the sporulation sigma factor regulons viz., SigE (early mother cell-specific sporulation), SigF (early forespore-specific sporulation), SigG (late forespore-specific sporulation) and SigK (late mother cell-specific

sporulation) in both the SH2 and SH5. Interestingly, specific pattern of changes was observed in the SigK regulon in SH2 and SH5. That is, in SH2, all the affected genes in SigK were up-regulated, which was contrary to the down-regulation of all the genes in SH5 in SigK sigma factor regulon. As SigK regulon genes comprise of gene involved in the late mother cell-specific sporulation, these observations suggested a potential role of SpoVG in promoting sporulation in the later stages of sporulation, while suppressing the sporulation proteins in the beginning of the sporulation process. This result corroborated with the previous understanding of *spoVG* in sporulation, where its transcription is known to initiate in the initial stage of sporulation whose protein product is active in the later stages of sporulation (Rosenbluh *et al.*, 1981; Hudspeth and Vary, 1992; Matsuno and Sonenshein, 1999). However, these studies were conducted in *B. subtilis* with genetic backgrounds in other than 168 such as PY79, SMY and JH642. As no evidence is available for the role of *spoVG* in the 168 genetic background, validating these results from the previous literature would be difficult.

In summary, it was shown that absence of SpoVG in *B. subtilis* perturbs the global proteome in each of LB, M9, SH2 and SH5. The perturbations identified in this study were akin to the changes identified in transcriptome of Δ*spoVG* (Chapter 3) and the deletion strain of *spoVG* in *L. monocytogenes*, *B. burgdorferi* and *S. aureus* (Bischoff *et al.*, 2016; Burke and Portnoy, 2016; Liu, Zhang and Sun, 2016; Savage *et al.*, 2018). These observations taken together pointed to a pleiotropic role of SpoVG, with similar functions across the bacterial species studied thus far. However, further phenotypic, biochemical and molecular investigations would be needed to reveal the true nature and importance of these changes.

8.3 RNA-Binding Activity of SpoVG

Chapter 5 explored the RNA-binding property of SpoVG *in vivo* by utilising Ultraviolet (UV) crosslinking and co-immunoprecipitation (co-IP) techniques

such as UV cross-linking and immunoprecipitation (CLIP) and UV cross-linking and analysis of cDNAs (CRAC) (Chapter 5). Both the CLIP and CRAC experiments were carried out in LB stationary condition and investigated the RNA-binding activity via radiolabelling and corresponding Western blot analysis. For the preliminary analysis of the RNA-binding activity and co-IP, the CLIP analysis was employed where SpoVG with a 3X-FLAG tag was used. The autoradiogram and Western blot analysis of the CLIP experiment identified SpoVG as an RNA-binding protein *in vivo*. Furthermore, for a stringent purification of the SpoVG-RNA complexes, a CRAC analysis was carried out using SpoVG with His(6X)-TEV cleavage site-FLAG(3X) (HTF) tag without sequencing analysis. Instead, the CRAC experiment investigated the RNA-binding activity of SpoVG-HTF via the autoradiogram and Western blot analysis. The results from CRAC experiment also identified SpoVG as an RBP *in vivo*. Thus, the result from CLIP analysis together with the CRAC assay established SpoVG as an RNA-binding protein in *B. subtilis*.

Further, to identify the RNA transcripts binding to SpoVG and the RNA-RNA interactions mediated by SpoVG, a Cross-linking, Ligation and Sequencing of Hybrids (CLASH) protocol with SpoVG-HTF was carried out in LB stationary condition. The sequencing analysis of the RNA transcripts co-immunoprecipitated (co-IPed) with SpoVG-HTF compared to the WT in this study identified 2,265 RNA transcripts. The characterisation of top 20 abundantly co-IPed transcripts identified most of the abundant transcripts with antimicrobial properties. Whereas, many other abundantly co-IPed transcripts also had functions in metabolic processes such central carbon metabolism, amino acid and nucleotide metabolism, cell-wall synthesis and utilization of alternate carbons.

These observations were in line with the results of transcriptomic and proteomic analysis of the Δ*spoVG* (Chapter 3 and 4). For example, in both the transcriptome and the proteome of Δ*spoVG*, the antimicrobial genes were significantly decreased in the mutant (Chapter 3 and 4). The abundance of antimicrobial transcripts binding to SpoVG and the decreased

levels of these genes and proteins indicated a preference of SpoVG in the regulation of antimicrobial compounds in *B. subtilis*. Similar observations were also made for the central carbon and nitrogen metabolism RNA transcripts, which were found to be decreased in both the transcriptomic and proteomic analysis of Δ*spoVG* but were abundantly co-IPed with SpoVG.

Amongst the top 338 abundant transcripts analysed with average read count >2,500 bound to SpoVG-HTH identified transcripts encoding for RNases such as RNase J1, RNase J2 and RNase Y. This observation implied a role of SpoVG in the post-transcriptional gene regulation via sRNAs and RNases. Moreover, transcripts encoding 16S rRNA and 23S rRNA were also identified which indicated a role of SpoVG in the ribosome assembly and translation mechanism in *B. subtilis*.

Another important observation from this co-IP data was the identification of the *spoVG* mRNA co-IPed with SpoVG. This observation was similar to the observation from the experiments in *B. burgdorferi* where SpoVG was shown to bind to its own mRNA transcript, possibly as result of structural affinity rather than specific nucleotide sequence (Savage *et al.*, 2018). Hence, from our data we speculate that SpoVG might also have an autoregulatory or RNA sequestering role in modulating the expression of *spoVG* in late stationary growth in *B. subtilis*.

Another interesting observation from the CLASH experiment was the enrichment of longer RNA transcripts with lengths greater than 1 kb among the most abundant co-IPed transcripts. This observation was intriguing and suggested a role of SpoVG in stabilizing longer transcripts. However, specific RNA stability experiments would be needed to prove this hypothesis. Furthermore, the RNA-binding activity of SpoVG in other growth conditions such as M9, SH2 and SH5 remains unknown. To generate a comprehensive interactome of SpoVG and identify its functions, the RNA-binding data hence generated could be analysed with the transcriptomic and proteomic analysis of Δ*spoVG*.

The analysis to identify the RNA-RNA interactions mediated by SpoVG revealed 153 significantly enriched interactions. A majority of the protein coding transcripts encode for antimicrobial compounds such as surfactin, sublancin, subtilosin A, antilisterial bacteriocin (subtilosin) and plipastatin. Whereas some of the other transcripts also participate in DNA replication, acquisition of iron, cell division, metabolic processes, stress, protein degradation, RNA metabolism, TCA cycle, competence and uptake of different compound.

This study also identified an interaction between the sRNA S415 and the TCA cycle gene *citZ*. S415 (RoxS) is a known regulator of *citZ* (Durand *et al.*, 2015, 2019). Hence, the identification of this interaction validated the output of the CLASH experiment. However, the total number of sRNA-mRNA interactions with SpoVG were very low as compared to the thousands of interactions mediated by the global regulators such as Hfq in *E. coli* (Iosub *et al.*, 2020). From these results, it could be assumed that SpoVG is not a global regulator in mediating the sRNA-mRNA interactions in *B. subtilis*. Another reason for the lower number of sRNA-mRNA interactions captured on SpoVG could be due to the missed stages of growth conditions with active sRNA-mRNA interactions. Furthermore, perhaps the post-transcriptional gene regulation in the Gram-positive bacteria are different to our current understandings from the Gram-negatives. Or alternate RBPs with functions in sRNA-mRNA based gene regulation are still to be identified.

8.4 Protein-Protein Interaction with SpoVG

Studying protein-protein interactions in bacteria provide a powerful tool to identify how proteins function in given growth conditions and cellular pathways. In this thesis, chapter 6 investigated the protein-protein interaction (PPI) with SpoVG in *B. subtilis* in LB stationary growth condition *in vivo* for the first time. For this, a formaldehyde cross-linking of bacteria in solution was adopted and the immunoprecipitated proteins with SpoVG

(with an HTF epitope tag) were identified using quantitative mass-spectrometry.

The results from this study identified 237 significantly co-IPed proteins with SpoVG. A majority of these co-IPed proteins had functions in central carbon metabolism, amino acid metabolism and biosynthesis pathways (antibiotics, secondary metabolites, lipids, amino acids and nucleotides). Whereas some of the other co-IPed proteins were involved in transcription, translation, protein folding and degradation, DNA maintenance, RNA degradation, signal transduction, uptake of additional carbon sources and cellular motility. A diverse set of proteins identified in this study indicated a multitude function of SpoVG in *B. subtilis*. This result corroborated with the previous observations of the Chapter 3, 4 and 5 where *spoVG* showed a wider participation in the cell.

Interestingly, RNases such as RNase M16 (endoribonuclease, required for the maturation of the 3' end of 16S rRNA), RNase PH (3' → 5' exoribonuclease), RNase HIII (endoribonuclease) and a putative RNase (involved in the processing of the 5' end of pre-16S rRNA) were also among the co-IPed proteins with SpoVG. This indicated a potential participation of SpoVG with RNases in mediating sRNA-mRNA interactions, a process which is similar to the major RBPs such as Hfq in Gram-negative bacteria. The data also identified several glycolytic enzymes and DNA nuclease, replication and repair enzymes that indicated a regulatory role of SpoVG in addition to its RNA-based regulation.

Among the top abundant co-IPed proteins was the TCA cycle protein Icd (isocitrate dehydrogenase). Previously, Icd has been shown to have roles in the in the regulatory role of SpoVG in the asymmetric septation during sporulation (Matsuno and Sonenshein, 1999). However, the mode by which SpoVG regulates the asymmetric septation remains unknown. Although the identification of Icd with SpoVG in our study in the LB stationary condition validated the outputs of the co-IP experiment, it however did not explain the importance of this interaction in the non-sporulating conditions. Nonetheless, we speculate that perhaps SpoVG is actively associated with

TCA cycle and glycolytic enzyme which derives the post-transcriptional gene regulation in the late stationary phase. Also, the RNA-binding activity of SpoVG at this stage remains unsolved. In the future, it would be interesting to explore if the regulatory interactions of SpoVG with Icd or other TCA and glycolytic enzymes is dependent or independent of its RNA-binding activity.

Further, we speculate that SpoVG in addition to being an interacting protein partner with TCA cycle and glycolytic enzymes, also participates in linking these two processes. This argument was backed by our discovery of the protein PdhD (dihydrolipoamide dehydrogenase E3 subunit of both pyruvate dehydrogenase and 2-oxoglutarate dehydrogenase complexes) in the protein co-IP data set. PdhD is a TCA cycle enzyme that functions in linking glycolysis and TCA cycle and are involved in sporulation process (Gao *et al.*, 2002).

Additionally, proteins involved in cell wall metabolism were also identified in the co-IP data. As SpoVG is a negative regulator of asymmetric septation (Matsuno and Sonenshein, 1999), therefore. identification of the cell wall metabolic proteins was reassuring. This observation was also corroborated with the significant enrichment of cell wall genes and proteins in both the -omics data set of Δ*spoVG*. Although, the roles of SpoVG in connection to these proteins in LB condition remain to be explored.

The interaction data identified the global transcriptional regulator CcpA as one of the significantly co-IPed proteins with SpoVG. CcpA is a multi-functional transcriptional regulator and is best known for its participation in the carbon catabolite repression (CCR) in *B. subtilis* (Henkin, 1996; Küster *et al.*, 1999; Warner and Lolkema, 2003). A study by Wünsche *et al.*, (2012) identified SpoVG as an interacting protein of CcpA (Wünsche *et al.*, 2012). Thus, this validated the data from this study and indicated a role of SpoVG with CcpA. However, further investigations would be needed to determine the nature (direct or indirect) of the SpoVG-CcpA interaction.

To summarise, identification of a number a significantly co-IPed proteins suggested a vast interactome of SpoVG in the cell. In future, it would be

interesting to characterise the identified proteins and unravel the importance of these interactions in *B. subtilis*. It will be also useful to investigate the protein partners of SpoVG in M9, SH2 and SH5 to identify if SpoVG specific protein partners (a single protein or a complex of proteins) for its activity in all the growth conditions or has a diverse set of proteins in different growth conditions.

8.5 Phenotypic Analysis of Δ*spoVG*

Finally, the transcriptome (Chapter 3) and proteome (Chapter 4) profiles obtained from *spoVG* deletion were compared to investigate the changes due to the gene deletion at the different levels of cellular processes. I further present some phenotypic experiments based on the -omics experiments to characterise the Δ*spoVG* mutant (Chapter 7). It was shown that only a handful of genes were commonly differentially expressed in both the transcriptome and proteome of the mutant (Figure 8.1). However, the functional characterisation of most of the proteomics data were coherent with the transcriptomic data recorded in M9.

Based on the findings of the -omics analysis of Δ*spoVG*, the phenotypes of Δ*spoVG* were investigated in LB and minimal media M9 and MSgg. This thesis explored the phenotypes such as motility, biofilm formation, production of pigment and secretion which are essential adaption behaviour of *B. subtilis*. The investigations revealed that the Δ*spoVG* has a reduced swimming and swarming motility in M9. Also, it has a retarded capacity to form biofilms on MSgg agar. The Δ*spoVG* swarming cells on M9 agar also lacked the formation of pink pigment (pulcherrimin) compared to its WT and complemented strain. Additionally, the Δ*spoVG* swarming cells has a restricted growth even after 24 hours of incubation. Similar experiments in nutrient rich media LBGM identified that Δ*spoVG* forms expanded biofilms on solid media, however, has reduced capacity to form biofilms on liquid media (pellicles). The biofilms formed on LBGM agar also had a decreased

production of pigment which was comparable to the observations made in minimal media.

When tested for secretion, Δ*spoVG* showed decreased α-amylase production suggesting a deficiency in its secretion system. Comparison of the total secreted proteins of Δ*spoVG* via gel electrophoresis and staining showed decreased proteins levels in the mutant indicating a decrease in the secretion of *B. subtilis* in the absence of *spoVG*. Taken together, it is speculated that *spoVG* has a role in facilitating secretion in *B. subtilis*.

Lastly, the efficiency of Δ*spoVG* to growth in the presence of alternate carbon sources in its culture media was tested in PM. From this study, it was identified that Δ*spoVG* has a better growth in the presence of sucrose, succinate, galactose, glycerol and arabinose in comparison to the WT growth in the same media. These observations suggested that perhaps *spoVG* supresses the growth of the *B. subtilis* in the presence of alternate carbon hence allowing the preferential uptake of glucose. However, more phenotypic and biochemical experiments would be needed to establish exact reasons behind these changes.

In summary, my results showed a pleiotropic characteristic of the Δ*spoVG* mutant in the standard laboratory conditions. Pleiotropy is known to be a common characteristic feature of the major bacterial RBPs (Holmqvist and Vogel, 2018). Prominent phenotypes affected due to the deletion of the major RBPs include virulence, biofilm formation, cell wall metabolism, motility and chemotaxis, carbon metabolism and cellular stress. From the previous investigations, it has also been shown that deletion of *spoVG* in *L. monocytogenes, S. aureus* and *B. burgdorferi* produced pleiotropic changes in the cell (Schulthess *et al.*, 2011; Burke and Portnoy, 2016; Liu, Zhang and Sun, 2016; Savage *et al.*, 2018).

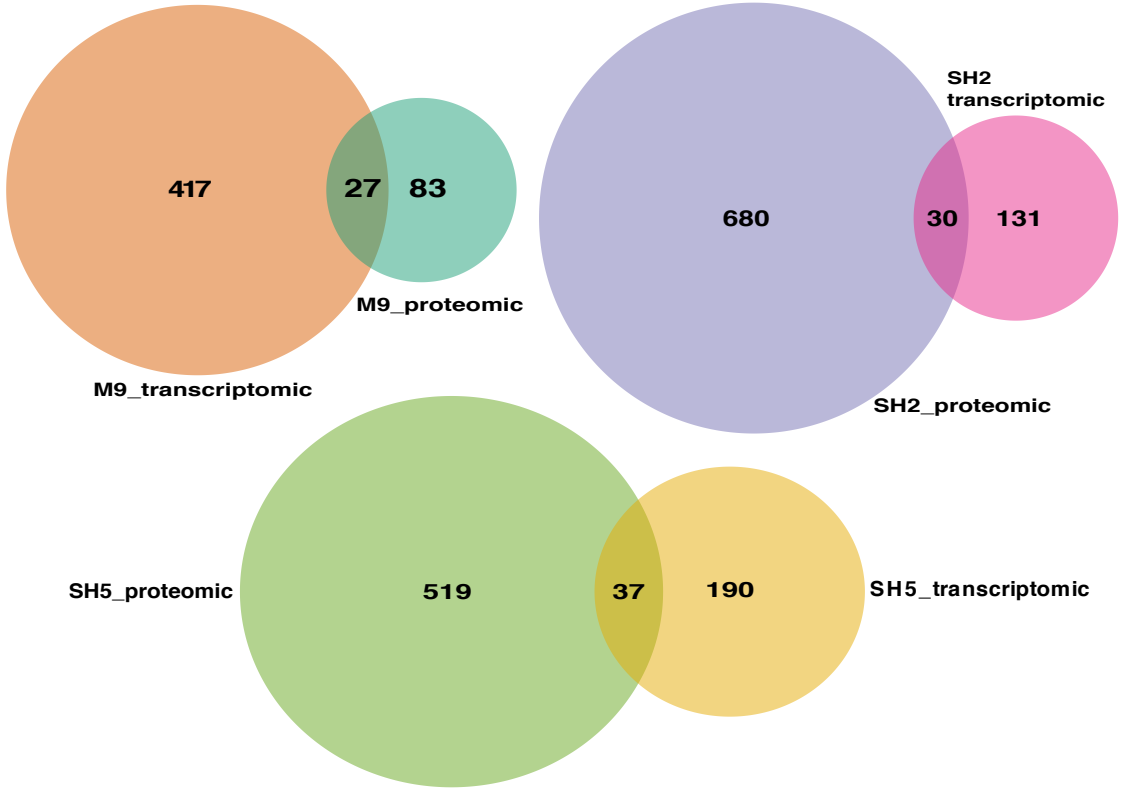


Figure 8.1: Comparison of transcriptomic and proteomic changes in the Δ*spoVG* mutant in M9, SH2 and SH5

Even the deletion of *spoVG* in *L. monocytogenes* and *B. burgdorferi* produced pleiotropic changes in the cell. For the first time, through this study, I showed that the pleiotropy in *B. subtilis* is present in the absence of *spoVG* (Figure 8.2). Many of these changes are similar and comparable to the previously known RBPs and the deletion effect of *spoVG* in other bacterial species. Hence, we summarize that these changes are a classical signature of a global post-transcriptional gene regulator. Specific molecular experiments would be needed to understand the exact roles of *spoVG* in the physiology of *B. subtilis*.

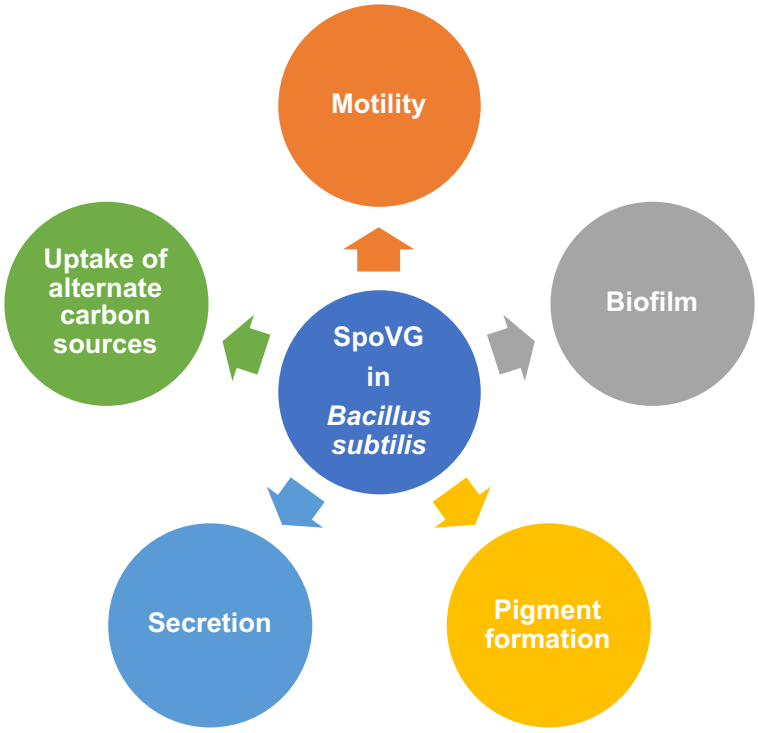


Figure 8.2: Summary of phenotypes identified with Δ*spoVG* in this thesis

8.6 Limitations and Future Work

The thesis explored the total transcriptome and proteome of the Δ*spoVG* mutant in LB and M9 in the mid-exponential growth phase (Chapter 3 and 4). However, the RNA-binding assays and the PPI experiments were carried out in late stationary phase (Chapter 4 and 5). Also, the phenotypic investigation in the Δ*spoVG* mutant identified changes (such as biofilms, motility and growth on alternate carbon sources) in late stationary phase cells (Chapter 7). Therefore, it would be interesting to investigate the transcriptome and proteome of Δ*spoVG* in late stationary phases in LB and M9 to decipher the changes in these cells. Additionally, easily identifiable phenotypic changes, such as the changes observed in the motility (swarming and swimming) and biofilm of Δ*spoVG*, could be explored as amenable systems to investigate the mechanism of gene regulation in the absence of *spoVG*. Experiments involving the use of reporters for global measurements, microscopy and population differentiation could also be designed to further investigate the findings of thesis.

Further, for the sporulation conditions only the hour 2 and hour 5 cells were sampled. Better comparisons in the -omics data could be made with cells sampled at hour 0 (SH0) as a refence point for the Δ*spoVG*. From the -omics data analysis, a comparison of the Δ*spoVG* strain in SH5 vs SH2 shall also be made to identify the differentially expressed genes and differentially produced proteins in Δ*spoVG* in SH5 compared to Δ*spoVG* in SH2. Also, it will be exciting to identify the changes in the phenotypes of a strain over-expressing *spoVG*. Many follow up experiments could also be carried out from the -omics data to validate some of the changes identified in the Δ*spoVG* strain.

8.7 Summary

To conclude, this thesis presents a global multi-omics approach in characterising the highly conserved bacterial gene *spoVG* in *B. subtilis*. From the work carried out in this study, it was shown that SpoVG in *B. subtilis* is an RBP *in vivo*. SpoVG participates in a number of functions in the cells, highlighting its role as a pleiotropic protein in the cell and not only as a sporulation protein (as it was characterised earlier). Nevertheless, it remains elusive if these observations are the consequences of direct or indirect participation *spoVG* in the cell. SpoVG has a number of interacting RNA and protein molecules which still remain to be explored in detail. Furthermore, the findings of SpoVG protein-protein interaction also suggested metabolic roles of SpoVG in *B. subtilis* that maybe dependent or independent of its RNA-binding activity (Figure 8.3).

Although SpoVG binds RNA and has a pleiotropic role in *B. subtilis* (similar to many other well-understood RBPs and global regulators), it is not a global regulator in post-transcriptional gene regulation mediating the sRNA-mRNA interactions. Comparison of the significantly co-IPed RNA transcripts and the proteins co-IPed with SpoVG indicated a major role of SpoVG in the carbon metabolism, amino acid metabolism and DNA maintenance in *B. subtilis* (Figure 8.4). Perhaps, *B. subtilis* has alternate post-transcriptional gene regulation mechanisms that do not require RBPs or SpoVG, there exist other RBPs that are yet to be discovered. This study suggests exploring further to understand the other roles of RBPs and post-transcriptional gene regulation in *B. subtilis* and many other Gram-positive bacteria.

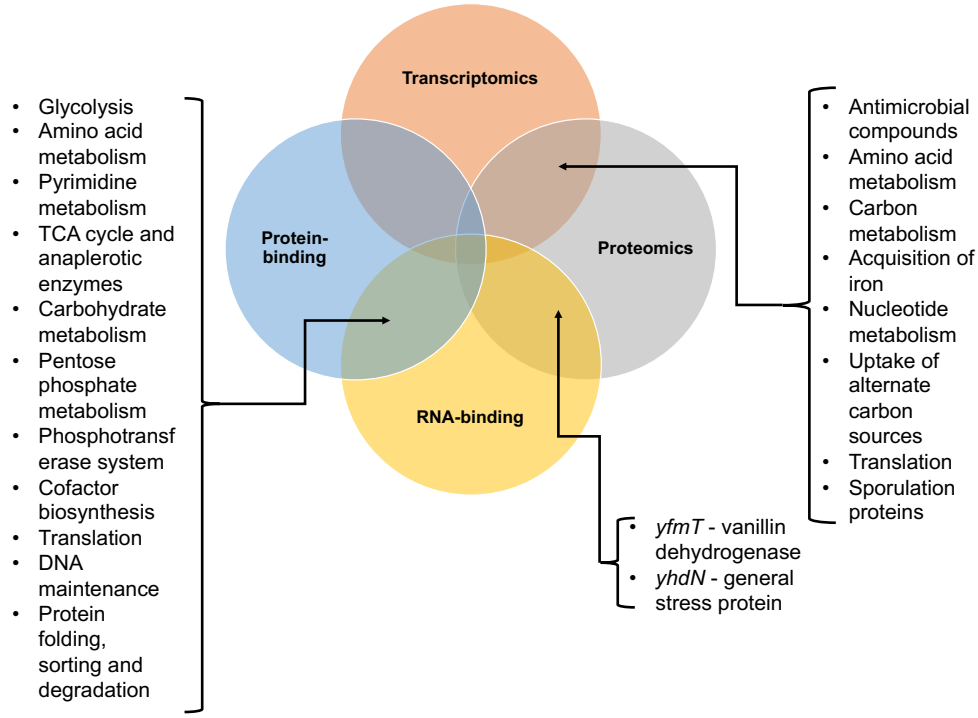


Figure 8.3: Biological processes commonly found across chapter 3,4,5, and 6

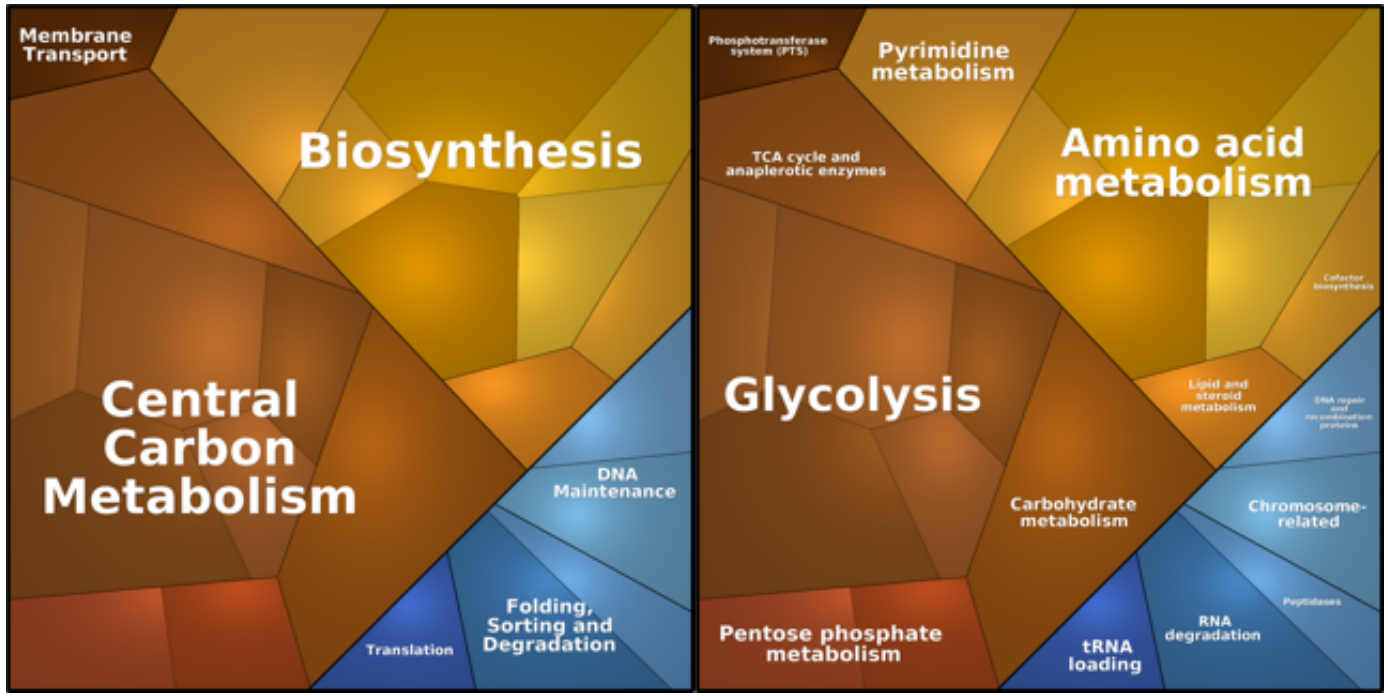


Figure 8.4: Identification of pathways common in significantly co-IPed RNAs and proteins with SpoVG

Characterisation of transcripts and proteins co-IPed with SpoVG in Chapter 5 and Chapter 6 visualised via KEGG pathway enrichment tool of proteomaps. Area of ach polygon is proportional to the average read count of the transcripts co-IPed with SpoVG in the CLASH analysis.

Bibliography

Abriouel, H. *et al.* (2011) 'Diversity and applications of Bacillus bacteriocins', *FEMS Microbiology Reviews*, 35(1), pp. 201–232. doi: 10.1111/j.1574-6976.2010.00244.x.

Ahmed, W., Zheng, K. and Liu, Z. F. (2016) 'Small non-coding RNAs: New insights in modulation of host immune response by intracellular bacterial pathogens', *Frontiers in Immunology*, 7(OCT), pp. 1–10. doi: 10.3389/fimmu.2016.00431.

Akanuma, G. *et al.* (2012) 'Inactivation of ribosomal protein genes in bacillus subtilis reveals importance of each ribosomal protein for cell proliferation and cell differentiation', *Journal of Bacteriology*, 194(22), pp. 6282–6291. doi: 10.1128/JB.01544-12.

Alén, C. and Sonenshein, A. L. (1999) 'Bacillus subtilis aconitase is an RNA-binding protein', *Proceedings of the National Academy of Sciences of the United States of America*, 96(18), pp. 10412–10417. doi: 10.1073/pnas.96.18.10412.

Ali, N. O. *et al.* (2001) 'Regulation of the acetoin catabolic pathway is controlled by sigma L in Bacillus subtilis', *Journal of Bacteriology*, 183(8), pp. 2497–2504. doi: 10.1128/JB.183.8.2497-2504.2001.

Allenby, N. E. E. *et al.* (2006) 'Phosphate starvation induces the sporulation killing factor of Bacillus subtilis', *Journal of Bacteriology*, 188(14), pp. 5299–5303. doi: 10.1128/JB.00084-06.

Andrade, J. M. *et al.* (2018) ' The RNA -binding protein Hfq is important for ribosome biogenesis and affects translation fidelity ', *The EMBO Journal*, 37(11), pp. 1–13. doi: 10.15252/embj.201797631.

Andrews, S. C., Robinson, A. K. and Rodríguez-Quiñones, F. (2003) 'Bacterial iron homeostasis', *FEMS Microbiology Reviews*, 27(2–3), pp. 215–237. doi: 10.1016/S0168-6445(03)00055-X.

Antelmann, H., Scharf, C. and Hecker, M. (2000) 'Phosphate starvation-

inducible proteins of Bacillus subtilis: Proteomics and transcriptional analysis', *Journal of Bacteriology*, 182(16), pp. 4478–4490. doi: 10.1128/JB.182.16.4478-4490.2000.

Arifuzzaman, M. *et al.* (2006) 'Large-scale identification of protein-protein interaction of Escherichia coli K-12', *Genome Research*, 16(5), pp. 686–691. doi: 10.1101/gr.4527806.

Arnaouteli, S. *et al.* (2017) 'Bifunctionality of a biofilm matrix protein controlled by redox state', *Proceedings of the National Academy of Sciences of the United States of America*, 114(30), pp. E6184–E6191. doi: 10.1073/pnas.1707687114.

Arnaouteli, S. *et al.* (2019) 'Pulcherrimin formation controls growth arrest of the Bacillus subtilis biofilm', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 116(27), pp. 13553–13562. doi: 10.1073/pnas.1903982116.

Arrieta-Ortiz, M. L. *et al.* (2015) ' An experimentally supported model of the Bacillus subtilis global transcriptional regulatory network ', *Molecular Systems Biology*. EMBO, 11(11), p. 839. doi: 10.15252/msb.20156236.

Asara, J. M. *et al.* (2008) 'A label-free quantification method by MS/MS TIC compared to SILAC and spectral counting in a proteomics screen', *Proteomics*. John Wiley & Sons, Ltd, 8(5), pp. 994–999. doi: 10.1002/pmic.200700426.

Van Assche, E. *et al.* (2015) 'RNA-binding proteins involved in post-transcriptional regulation in bacteria', *Frontiers in Microbiology*, 6(MAR). doi: 10.3389/fmicb.2015.00141.

Assis, N. G. *et al.* (2019) 'Identification of Hfq-binding RNAs in Caulobacter crescentus', *RNA Biology*. Taylor & Francis, 16(6), pp. 719–726. doi: 10.1080/15476286.2019.1593091.

Ayala, F. R., Bartolini, M. and Grau, R. (2020) 'The Stress-Responsive Alternative Sigma Factor SigB of Bacillus subtilis and Its Relatives: An Old Friend With New Functions', *Frontiers in Microbiology*, 11(September), pp. 1–20. doi: 10.3389/fmicb.2020.01761.

Azam, T. A. *et al.* (1999) 'Growth phase-dependent variation in protein composition of the Escherichia coli nucleoid', *Journal of Bacteriology*, 181(20), pp. 6361–6370. doi: 10.1128/jb.181.20.6361-6370.1999.

Babitzke, P. *et al.* (2019) 'Posttranscription Initiation Control of Gene Expression Mediated by Bacterial RNA-Binding Proteins', *Annual Review of Microbiology*, 73(1), pp. 43–67. doi: 10.1146/annurev-micro-020518-115907.

Babitzke, P., Bear, D. G. and Yanofsky, C. (1995) 'TRAP, the trp RNA-binding attenuation protein of Bacillus subtilis, is a toroid-shaped molecule that binds transcripts containing GAG or UAG repeats separated by two nucleotides', *Proceedings of the National Academy of Sciences of the United States of America*, 92(17), pp. 7916–7920. doi: 10.1073/pnas.92.17.7916.

Babitzke, P. and Romeo, T. (2007) 'CsrB sRNA family: sequestration of RNA-binding regulatory proteins', *Current Opinion in Microbiology*, 10(2), pp. 156–163. doi: 10.1016/j.mib.2007.03.007.

Baek, Y. M. *et al.* (2019) 'The bacterial endoribonuclease RNase E can cleave RNA in the absence of the RNA chaperone Hfq', *Journal of Biological Chemistry*, 294(44), pp. 16465–16478. doi: 10.1074/jbc.RA119.010105.

Bantscheff, M. *et al.* (2007) 'Quantitative mass spectrometry in proteomics: A critical review', *Analytical and Bioanalytical Chemistry*, 389(4), pp. 1017–1031. doi: 10.1007/s00216-007-1486-6.

Barnett, T. A. *et al.* (1983) 'Production by Bacillis subtilis of brown sporulation-associated pigments.', *Canadian journal of microbiology*. Canada, 29(1), pp. 96–101. doi: 10.1139/m83-015.

Bastet, L. *et al.* (2018) 'Maestro of regulation: Riboswitches orchestrate gene expression at the levels of translation, transcription and mRNA decay', *RNA Biology*. Taylor & Francis, 15(6), pp. 679–682. doi: 10.1080/15476286.2018.1451721.

Baumgardt, K. *et al.* (2018) 'The essential nature of YqfG, a YbeY homologue required for 3' maturation of Bacillus subtilis 16S ribosomal RNA is suppressed by deletion of RNase R', *Nucleic Acids Research*, 46(16), pp.

8605–8615. doi: 10.1093/nar/gky488.

Beckmann, B. M., Castello, A. and Medenbach, J. (2016) 'The expanding universe of ribonucleoproteins: of novel RNA-binding proteins and unconventional interactions', *Pflugers Archiv European Journal of Physiology*. Pflügers Archiv - European Journal of Physiology, 468(6), pp. 1029–1040. doi: 10.1007/s00424-016-1819-4.

Belitsky, B. R. and Sonenshein, A. L. (2011) 'CodY-mediated regulation of guanosine uptake in Bacillus subtilis', *Journal of Bacteriology*, 193(22), pp. 6276–6287. doi: 10.1128/JB.05899-11.

Benjamin, J. A. M. and Massé, E. (2014) 'The iron-sensing aconitase B binds its own mRNA to prevent sRNA-induced mRNA cleavage', *Nucleic Acids Research*, 42(15), pp. 10023–10036. doi: 10.1093/nar/gku649.

Bertram, R. and Schuster, C. F. (2014) 'Post-transcriptional regulation of gene expression in bacterial pathogens by toxin-antitoxin systems', *Frontiers in Cellular and Infection Microbiology*, 4(JAN), pp. 1–7. doi: 10.3389/fcimb.2014.00006.

Bijlsma, S. *et al.* (2006) 'Large-scale human metabolomics studies: A strategy for data (pre-) processing and validation', *Analytical Chemistry*, 78(2), pp. 567–574. doi: 10.1021/ac051495j.

Bischoff, M. *et al.* (2016) 'Stk1-mediated phosphorylation stimulates the DNA-binding properties of the Staphylococcus aureus SpoVG transcriptional factor', *Biochemical and Biophysical Research Communications*, 473(4), pp. 1223–1228. doi: 10.1016/j.bbrc.2016.04.044.

Bleichert, F. and Baserga, S. J. (2010) 'Ribonucleoprotein multimers and their functions', *Critical Reviews in Biochemistry and Molecular Biology*, 45(5), pp. 331–350. doi: 10.3109/10409238.2010.496772.

Boehmer, J. L. *et al.* (2010) 'The proteomic advantage: Label-free quantification of proteins expressed in bovine milk during experimentally induced coliform mastitis', *Veterinary Immunology and Immunopathology*. Elsevier B.V., 138(4), pp. 252–266. doi: 10.1016/j.vetimm.2010.10.004.

Bohannon, D. E. and Sonenshein, A. L. (1989) 'Positive regulation of

glutamate biosynthesis in Bacillus subtilis.', *Journal of bacteriology*, 171(9), pp. 4718–4727. doi: 10.1128/jb.171.9.4718-4727.1989.

Bohn, C., Rigoulay, C. and Bouloc, P. (2007) 'No detectable effect of RNA-binding protein Hfq absence in Staphylococcus aureus', *BMC Microbiology*, 7, pp. 1–9. doi: 10.1186/1471-2180-7-10.

Bose, U. *et al.* (2019) 'Proteomics: Tools of the Trade', *Adv Exp Med Biol*, 1073, pp. 1–22. doi: doi:10.1007/978-3-030-12298-0_1.

Botella, E. *et al.* (2010) 'pBaSysBioII: An integrative plasmid generating gfp transcriptional fusions for high-throughput analysis of gene expression in Bacillus subtilis', *Microbiology*, 156(6), pp. 1600–1608. doi: 10.1099/mic.0.035758-0.

Boudry, P. *et al.* (2014) 'Pleiotropic role of the RNA chaperone protein Hfq in the human pathogen Clostridium difficile', *Journal of Bacteriology*, 196(18), pp. 3234–3248. doi: 10.1128/JB.01923-14.

Bouloc, P. and Repoila, F. (2016) 'Fresh layers of RNA-mediated regulation in Gram-positive bacteria', *Current Opinion in Microbiology*. Elsevier Ltd, 30, pp. 30–35. doi: 10.1016/j.mib.2015.12.008.

Boylan, S. A. *et al.* (1988) 'Early-blocked sporulation mutations alter expression of enzymes under carbon control in Bacillus subtilis', *MGG Molecular & General Genetics*, 212(2), pp. 271–280. doi: 10.1007/BF00334696.

Boylan, S. A. *et al.* (1993) 'Stress-induced activation of the σ(B) transcription factor of Bacillus subtilis', *Journal of Bacteriology*, 175(24), pp. 7931–7937. doi: 10.1128/jb.175.24.7931-7937.1993.

Breaker, R. R. (2018) 'Riboswitches and translation control', *Cold Spring Harbor Perspectives in Biology*, 10(11). doi: 10.1101/cshperspect.a032797.

Britton, R. A. *et al.* (2002) 'Genome-wide analysis of the stationary-phase sigma factor (Sigma-H) regulon of Bacillus subtilis', *Journal of Bacteriology*, 184(17), pp. 4881–4890. doi: 10.1128/JB.184.17.4881-4890.2002.

Bromley, K. M. *et al.* (2015) 'Interfacial self-assembly of a bacterial hydrophobin', *Proceedings of the National Academy of Sciences of the*

United States of America, 112(17), pp. 5419–5424. doi: 10.1073/pnas.1419016112.

Brown, R. F. (2016) *Investigating the Evolutionary Origins and Cell Biology of Negativicutes*. University of Warwick.

Bryan, E. M., Beall, B. W. and Moran, C. P. (1996) 'A σ(E)-dependent operon subject to catabolite repression during sporulation in Bacillus subtilis', *Journal of Bacteriology*, 178(16), pp. 4778–4786. doi: 10.1128/jb.178.16.4778-4786.1996.

Buescher, J. M. *et al.* (2012) 'Global network reorganization during dynamic adaptations of Bacillus subtilis metabolism.', *Science (New York, N.Y.)*. American Association for the Advancement of Science, 335(6072), pp. 1099–103. doi: 10.1126/science.1206871.

Burbulys, D., Trach, K. A. and Hoch, J. A. (1991) 'Initiation of sporulation in B. subtilis is controlled by a multicomponent phosphorelay', *Cell*, 64(3), pp. 545–552. doi: 10.1016/0092-8674(91)90238-T.

Burke, T. P. and Portnoy, D. A. (2016) 'SpoVG Is a Conserved RNA-Binding Protein That Regulates Listeria monocytogenes Lysozyme Resistance, Virulence, and Swarming Motility', *American Academy of Microbiology*, 7(2), pp. 1–10. doi: 10.1128/mBio.00240-16.

Burton, A. T. *et al.* (2019) 'Transcriptional regulation and mechanism of sigN (ZpdN), a pBS32-encoded sigma factor in bacillus subtilis', *mBio*, 10(5), pp. 1–15. doi: 10.1128/mBio.01899-19.

Bushley, K. E. and Turgeon, B. G. (2010) 'Phylogenomics reveals subfamilies of fungal nonribosomal peptide synthetases and their evolutionary relationships', *BMC Evolutionary Biology*, 10(1), pp. 1–23. doi: 10.1186/1471-2148-10-26.

Butcher, R. A. *et al.* (2007) 'The identification of bacillaene, the product of the PksX megacomplex in Bacillus subtilis', *Proceedings of the National Academy of Sciences of the United States of America*, 104(5), pp. 1506–1509. doi: 10.1073/pnas.0610503104.

Butland, G. *et al.* (2005) 'Interaction network containing conserved and

essential protein complexes in Escherichia coli', *Nature*, 433(7025), pp. 531–537. doi: 10.1038/nature03239.

Caillet, J. *et al.* (2019) 'Identification of protein-protein and ribonucleoprotein complexes containing Hfq', *Scientific Reports*, 9(1), pp. 1–12. doi: 10.1038/s41598-019-50562-w.

Cairns, L. S. *et al.* (2013) 'A mechanical signal transmitted by the flagellum controls signalling in Bacillus subtilis', *Molecular Microbiology*, 90(1), pp. 6–21. doi: 10.1111/mmi.12342.

Calvio, C. *et al.* (2005) 'Swarming differentiation and swimming motility in Bacillus subtilis are controlled by swrA, a newly identified dicistronic operon.', *Journal of bacteriology*. American Society for Microbiology Journals, 187(15), pp. 5356–66. doi: 10.1128/JB.187.15.5356-5366.2005.

Camp, A. H. and Losick, R. (2008) 'A novel pathway of intercellular signalling in Bacillus subtilis involves a protein with similarity to a component of type III secretion channels', *Molecular Microbiology*, 69(2), pp. 402–417. doi: 10.1111/j.1365-2958.2008.06289.x.

Campo, N., Marquis, K. A. and Rudner, D. Z. (2008) 'SpoIIQ anchors membrane proteins on both sides of the sporulation septum in Bacillus subtilis', *Journal of Biological Chemistry*, 283(8), pp. 4975–4982. doi: 10.1074/jbc.M708024200.

Cangiano, G. *et al.* (2010) 'Direct and indirect control of late sporulation genes by GerR of Bacillus subtilis', *Journal of Bacteriology*, 192(13), pp. 3406–3413. doi: 10.1128/JB.00329-10.

Carpenter, P. B., Hanlon, D. W. and Ordal, G. W. (1992) 'flhF, a Bacillus subtilis flagellar gene that encodes a putative GTP-binding protein', *Molecular Microbiology*, 6(18), pp. 2705–2713. doi: 10.1111/j.1365-2958.1992.tb01447.x.

Carter, H. L. and Moran, C. P. (1986) 'New RNA polymerase σ factor under spo0 control in Bacillus subtilis', *Proceedings of the National Academy of Sciences of the United States of America*, 83(24), pp. 9438–9442. doi: 10.1073/pnas.83.24.9438.

Chastanet, A. and Losick, R. (2007) 'Engulfment during sporulation in Bacillus subtilis is governed by a multi-protein complex containing tandemly acting autolysins', *Molecular Microbiology*, 64(1), pp. 139–152. doi: 10.1111/j.1365-2958.2007.05652.x.

Chatr-aryamontri, A. *et al.* (2008) 'Protein interactions: integration leads to belief', *Trends in Biochemical Sciences*, 33(6), pp. 241–242. doi: 10.1016/j.tibs.2008.04.002.

Chen, M. *et al.* (2020) 'SpoVG Is Necessary for Sporulation in Bacillus anthracis', *Microorganisms*, 8(4), p. 548. doi: 10.3390/microorganisms8040548.

Choi, S. K. and Saier, M. H. (2005) 'Regulation of sigL expression by the catabolite control protein CcpA involves a roadblock mechanism in Bacillus subtilis: Potential connection between carbon and nitrogen metabolism', *Journal of Bacteriology*, 187(19), pp. 6856–6861. doi: 10.1128/JB.187.19.6856-6861.2005.

Christiansen, J. K. *et al.* (2004) 'The RNA-Binding Protein Hfq of Listeria monocytogenes: Role in Stress Tolerance and Virulence', *Journal of Bacteriology*, 186(11), pp. 3355–3362. doi: 10.1128/JB.186.11.3355.

Chu, F. *et al.* (2006) 'Targets of the master regulator of biofilm formation in Bacillus subtilis', *Molecular Microbiology*, 59(4), pp. 1216–1228. doi: 10.1111/j.1365-2958.2005.05019.x.

Chung, Y. S. and Dubnau, D. (1998) 'All seven comG open reading frames are required for DNA binding during transformation of competent Bacillus subtilis', *Journal of Bacteriology*, 180(1), pp. 41–45. doi: 10.1128/jb.180.1.41-45.1998.

Connelly, M. B., Young, G. M. and Sloma, A. (2004) 'Extracellular proteolytic activity plays a central role in swarming motility in Bacillus subtilis', *Journal of Bacteriology*, 186(13), pp. 4159–4167. doi: 10.1128/JB.186.13.4159-4167.2004.

Courtney, C. R., Cozy, L. M. and Kearns, D. B. (2012) 'Molecular characterization of the flagellar hook in Bacillus subtilis', *Journal of Bacteriology*, 194(17), pp. 4619–4629. doi: 10.1128/JB.00444-12.

Crowe-McAuliffe, C. *et al.* (2018) 'Structural basis for antibiotic resistance mediated by the Bacillus subtilis ABCF ATPase VmlR', *Proceedings of the National Academy of Sciences of the United States of America*, 115(36), pp. 8978–8983. doi: 10.1073/pnas.1808535115.

Cukras, A. R. *et al.* (2003) 'Ribosomal proteins S12 and S13 function as control elements for translocation of the mRNA:tRNA complex', *Molecular Cell*, 12(2), pp. 321–328. doi: 10.1016/S1097-2765(03)00275-2.

Dambach, M., Irnov, I. and Winkler, W. C. (2013) 'Association of RNAs with Bacillus subtilis Hfq', *PLoS ONE*, 8(2). doi: 10.1371/journal.pone.0055156.

Dawes, I. ., Kay, D. and Mandelstam, J. (1969) 'Sporulation in Bacillus subtilis. Establishment of a Time Scale for the Morphological Events', *Microbiology Society*, (1), pp. 1–6.

Diaconu, M. *et al.* (2005) 'Structural basis for the function of the ribosomal L7/12 stalk in factor binding and GTpase activation', *Cell*, 121(7), pp. 991–1004. doi: 10.1016/j.cell.2005.04.015.

van Dijl, J. M. and Hecker, M. (2013) 'Bacillus subtilis: from soil bacterium to super-secreting cell factory', *Microbial Cell Factories*, 12(1), p. 3. doi: 10.1186/1475-2859-12-3.

Ding, Y., Davis, B. M. and Waldor, M. K. (2004) 'Hfq is essential for Vibrio cholerae virulence and downregulates σE expression', *Molecular Microbiology*, 53(1), pp. 345–354. doi: 10.1111/j.1365-2958.2004.04142.x.

Dubnau, D. (1991) *The regulation of genetic competence in Bacillus subtilis*, *Molecular Microbiology*. doi: 10.1111/j.1365-2958.1991.tb01820.x.

Durand, S. *et al.* (2012) 'Three essential ribonucleases-RNase Y, J1, and III-control the abundance of a majority of bacillus subtilis mRNAs', *PLoS Genetics*, 8(3). doi: 10.1371/journal.pgen.1002520.

Durand, S. *et al.* (2015) 'A nitric oxide regulated small RNA controls expression of genes involved in redox homeostasis in Bacillus subtilis', *PLoS genetics*, 11(2), p. e1004957. doi: 10.1371/journal.pgen.1004957.

Durand, S. *et al.* (2019) 'Novel regulation from novel interactions: Identification of an RNA sponge that controls the levels, processing and

efficacy of the RoxS riboregulator of central metabolism in Bacillus subtilis', *bioRxiv*, p. 814905. doi: 10.1101/814905.

Eichenberger, P. *et al.* (2003) 'The σE regulon and the identification of additional sporulation genes in Bacillus subtilis', *Journal of Molecular Biology*, 327(5), pp. 945–972. doi: 10.1016/S0022-2836(03)00205-5.

Errington, J. (2003) 'Regulation of endospore formation in Bacillus subtilis', *Nature Reviews Microbiology*, 1(2), pp. 117–126. doi: 10.1038/nrmicro750.

Escorcia-Rodríguez, J. M., Tauch, A. and Freyre-González, J. A. (2020) 'Abasy Atlas v2.2: The most comprehensive and up-to-date inventory of meta-curated, historical, bacterial regulatory networks, their completeness and system-level characterization', *Computational and Structural Biotechnology Journal*, 18, pp. 1228–1237. doi: 10.1016/j.csbj.2020.05.015.

Eymann, C. *et al.* (2002) 'Bacillus subtilis functional genomics: Global characterization of the stringent response by proteome and transcriptome analysis', *Journal of Bacteriology*, 184(9), pp. 2500–2520. doi: 10.1128/JB.184.9.2500-2520.2002.

Eyraud, A. *et al.* (2014) 'A small RNA controls a protein regulator involved in antibiotic resistance in Staphylococcus aureus', *Nucleic Acids Research*, 42(8), pp. 4892–4905. doi: 10.1093/nar/gku149.

Fabret, C., Ehrlich, S. D. and Noirot, P. (2002) 'A new mutation delivery system for genome-scale approaches in Bacillus subtilis', *Molecular Microbiology*, 46(1), pp. 25–36. doi: 10.1046/j.1365-2958.2002.03140.x.

Fabret, C., Feher, V. A. and Hoch, J. A. (1999) 'Two-component signal transduction in Bacillus subtilis: How one organism sees its world', *Journal of Bacteriology*, 181(7), pp. 1975–1983. doi: 10.1128/jb.181.7.1975-1983.1999.

Fei, J. *et al.* (2008) 'Coupling of Ribosomal L1 Stalk and tRNA Dynamics during Translation Elongation', *Molecular Cell*, 30(3), pp. 348–359. doi: 10.1016/j.molcel.2008.03.012.

Ferrari, E., Jarnagin, A. S. and Schmidt, B. F. (1993) 'Commercial

Production of Extracellular Enzymes', *Bacillus subtilis and Other Gram-Positive Bacteria*, Chapter 62, pp. 917–937. doi: 10.1128/9781555818388.ch62.

Ferré-D'Amaré, A. R. (2016) 'RNA Binding: Getting Specific about Specificity', *Cell Chemical Biology*, 23(10), pp. 1177–1178. doi: 10.1016/j.chembiol.2016.10.001.

Fisher, S. (1991) 'Control Of Carbon And Nitrogen Metabolism In Bacillus Subtilis', *Annual Review of Microbiology*, 45(1), pp. 107–135. doi: 10.1146/annurev.micro.45.1.107.

Franze De Fernandez, M. T., Eoyang, L. and August, J. T. (1968) 'Factor fraction required for the synthesis of bacteriophage Qβ-RNA', *Nature*, 219(5154), pp. 588–590. doi: 10.1038/219588a0.

Fujisawa, M. *et al.* (2009) 'Characterization of Bacillus subtilis YfkE (ChaA): a calcium-specific Ca2+/H+ antiporter of the CaCA family.', *Archives of microbiology*, 191(8), pp. 649–657. doi: 10.1007/s00203-009-0494-7.

Fujita, M., González-Pastor, J. E. and Losick, R. (2005) 'High- and low-threshold genes in the Spo0A regulon of Bacillus subtilis', *Journal of Bacteriology*, 187(4), pp. 1357–1368. doi: 10.1128/JB.187.4.1357-1368.2005.

Gao, H. *et al.* (2002) 'The E1β and E2 subunits of the Bacillus subtilis pyruvate dehydrogenase complex are involved in regulation of sporulation', *Journal of Bacteriology*, 184(10), pp. 2780–2788. doi: 10.1128/JB.184.10.2780-2788.2002.

Gavin, A. C., Maeda, K. and Kühner, S. (2011) 'Recent advances in charting protein-protein interaction: Mass spectrometry-based approaches', *Current Opinion in Biotechnology*, 22(1), pp. 42–49. doi: 10.1016/j.copbio.2010.09.007.

Gerovac, M. *et al.* (2020) 'Global discovery of bacterial RNA-binding proteins by RNase-sensitive gradient profiles reports a new FinO domain protein', *RNA (New York, N.Y.)*, 26(10), pp. 1448–1463. doi: 10.1261/rna.076992.120.

van Gestel, J., Vlamakis, H. and Kolter, R. (2015) 'From Cell Differentiation to Cell Collectives: Bacillus subtilis Uses Division of Labor to Migrate', *PLoS Biology*. Public Library of Science, 13(4). doi: 10.1371/journal.pbio.1002141.

González-Pastor, J. E. (2011) 'Cannibalism: a social behavior in sporulating *Bacillus subtilis*', *FEMS Microbiology Reviews*, 35(3), pp. 415–424. doi: 10.1111/j.1574-6976.2010.00253.x.

González-Pastor, J. E., Hobbs, E. C. and Losick, R. (2003) 'Cannibalism by sporulating bacteria', *Science*. Johns Hopkins Univ. Press, 301(5632), pp. 510–513. doi: 10.1126/science.1086462.

Goodman, J. K. *et al.* (2018) 'Updates of the In-Gel Digestion Method for Protein Analysis by Mass Spectrometry', *Proteomics*, 18(23), pp. 1–5. doi: 10.1002/pmic.201800236.

Gorski, S. A., Vogel, J. and Doudna, J. A. (2017) 'RNA-based recognition and targeting: Sowing the seeds of specificity', *Nature Reviews Molecular Cell Biology*. Nature Publishing Group, 18(4), pp. 215–228. doi: 10.1038/nrm.2016.174.

Gottesman, S. (2004) 'The Small RNA Regulators of *Escherichia coli* : Roles and Mechanisms', *Annual Review of Microbiology*, 58(1), pp. 303–328. doi: 10.1146/annurev.micro.58.030603.123841.

Granneman, S. *et al.* (2009) *Identification of protein binding sites on U3 snoRNA and pre-rRNA by UV cross-linking and high-throughput analysis of cDNAs*, *PNAS*. Available at: https://www.pnas.org/content/pnas/106/24/9613.full.pdf (Accessed: 5 June 2019).

Greenberg, J. R. (1979) 'Ultraviolet light-induced crosslinking of mRNA to proteins', *Nucleic Acids Research*, 6(2), pp. 715–732. doi: 10.1093/nar/6.2.715.

Guariglia-Oropeza, V. and Helmann, J. D. (2011) 'Bacillus subtilis σ V confers lysozyme resistance by activation of two cell wall modification pathways, peptidoglycan O-Acetylation and D-alanylation of teichoic acids', *Journal of Bacteriology*, 193(22), pp. 6223–6232. doi: 10.1128/JB.06023-

11.

Guha, S., Roth, H. E. and Nierhaus, K. (1975) 'Ribosomal proteins of Bacillus subtilis vegetative and sporulating cells', *MGG Molecular & General Genetics*, 138(4), pp. 299–307. doi: 10.1007/BF00264799.

Gupta, R. S., Bhandari, V. and Naushad, H. S. (2012) 'Molecular signatures for the pvc clade (planctomycetes, verrucomicrobia, chlamydiae, and lentisphaerae) of bacteria provide insights into their evolutionary relationships', *Frontiers in Microbiology*, 3(SEP), pp. 1–19. doi: 10.3389/fmicb.2012.00327.

Guttenplan, S. B., Shaw, S. and Kearns, D. B. (2013) 'The cell biology of peritrichous flagella in Bacillus subtilis', *Molecular Microbiology*, 87(1), pp. 211–229. doi: 10.1111/mmi.12103.

Hahne, H. *et al.* (2008) 'From complementarity to comprehensiveness - Targeting the membrane proteome of growing Bacillus subtilis by divergent approaches', *Proteomics*, 8(19), pp. 4123–4136. doi: 10.1002/pmic.200800258.

Hämmerle, H. *et al.* (2014) 'Impact of Hfq on the Bacillus subtilis transcriptome', *PLoS ONE*, 9(6), p. e98661. doi: 10.1371/journal.pone.0098661.

Harwood, C. R. (1992) 'Bacillus subtilis and its relatives: molecular biological and industrial workhorses', *Trends in Biotechnology*, 10(C), pp. 247–256. doi: 10.1016/0167-7799(92)90233-L.

Harwood, C. R. and Cutting, S. M. (1990) *Molecular biological methods for Bacillus*. Wiley, Chichester, United Kingdom.

Hecker, M., Pané-Farré, J. and Völker, U. (2007) 'SigB-Dependent General Stress Response in Bacillus subtilis and Related Gram-Positive Bacteria', *Annual Review of Microbiology*, 61(1), pp. 215–236. doi: 10.1146/annurev.micro.61.080706.093445.

Hecker, M. and Völker, U. (2001) 'General stress response of Bacillus subtilis and other bacteria', *Advances in Microbial Physiology*, 44, pp. 35–91. doi: 10.1016/S0065-2911(01)44011-2.

Helder, S. *et al.* (2016) 'Determinants of affinity and specificity in RNA-binding proteins', *Current Opinion in Structural Biology*. Elsevier Ltd, 38, pp. 83–91. doi: 10.1016/j.sbi.2016.05.005.

Helmann, J. D. *et al.* (2001) 'Global Transcriptional Response of Bacillus subtilis to Heat Shock', *JOURNAL OF BACTERIOLOGY*, 183(24), pp. 7318–7328. doi: 10.1128/JB.183.24.7318-7328.2001.

Helmann, J. D. (2016) 'Bacillus subtilis extracytoplasmic function (ECF) sigma factors and defense of the cell envelope', *Current Opinion in Microbiology*. Elsevier Ltd, 30, pp. 122–132. doi: 10.1016/j.mib.2016.02.002.

Helwak, A. *et al.* (2013) 'Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding', *Cell*. Elsevier Inc., 153(3), pp. 654–665. doi: 10.1016/j.cell.2013.03.043.

Henkin, T. M. (1996) 'The role of the CcpA transcriptional regulator in carbon metabolism in Bacillus subtilis', *FEMS Microbiology Letters*, 135(1), pp. 9–15. doi: 10.1016/0378-1097(95)00370-3.

Hentze, M. W. *et al.* (2018) 'A brave new world of RNA-binding proteins', *Nature Reviews Molecular Cell Biology*. Nature Publishing Group, 19(5), pp. 327–341. doi: 10.1038/nrm.2017.130.

Hermansen, G. M. M. *et al.* (2018) 'Transcriptomic profiling of interacting nasal staphylococci species reveals global changes in gene and non-coding RNA expression', *FEMS Microbiology Letters*, 365(5), pp. 1–9. doi: 10.1093/femsle/fny004.

Herzberg, C. *et al.* (2007) 'SPINE: A method for the rapid detection and analysis of protein-protein interactions in vivo', *Proteomics*, 7(22), pp. 4032–4035. doi: 10.1002/pmic.200700491.

Hider, R. C. and Kong, X. (2010) 'Chemistry and biology of siderophores', *Natural Product Reports*, 27(5), pp. 637–657. doi: 10.1039/b906679a.

Hilden, I., Krath, B. N. and Hove-Jensen, B. (1995) 'Tricistronic operon expression of the genes gcaD (tms), which encodes N- acetylglucosamine 1-phosphate uridyltransferase, prs, which encodes phosphoribosyl

diphosphate synthetase, and ctc in vegetative cells of Bacillus subtilis', *Journal of Bacteriology*, 177(24), pp. 7280–7284. doi: 10.1128/jb.177.24.7280-7284.1995.

Hinojosa-Rebollar, R. *et al.* (1994) 'Synthesis and partial characterization of a melanin-like pigment of Bacillus subtilis', *Rev Latino Microbiol*, 35(4), pp. 399–406.

Hinojosa-Rebollar, R. E. *et al.* (1995) 'Biological flow tracers: growth and survival of Bacillus subtilis 65-8 under environmental stress', *Revista latinoamericana de microbiologia*. Mexico, 37(1), pp. 43–53.

Ho, T. D. *et al.* (2011) 'The Bacillus subtilis extracytoplasmic function σ factor σ v is induced by lysozyme and provides resistance to lysozyme', *Journal of Bacteriology*, 193(22), pp. 6215–6222. doi: 10.1128/JB.05467-11.

Hobley, L. *et al.* (2013) 'BslA is a self-assembling bacterial hydrophobin that coats the Bacillus subtilis biofilm', *Proceedings of the National Academy of Sciences of the United States of America*, 110(33), pp. 13600–13605. doi: 10.1073/pnas.1306390110.

Holberger, L. E. *et al.* (2012) 'A novel family of toxin/antitoxin proteins in Bacillus species', *FEBS Letters*. Federation of European Biochemical Societies, 586(2), pp. 132–136. doi: 10.1016/j.febslet.2011.12.020.

Holmqvist, E. *et al.* (2018) 'Global Maps of ProQ Binding In Vivo Reveal Target Recognition via RNA Structure and Stability Control at mRNA 3′ Ends', *Molecular Cell*, 70(5), pp. 971-982.e6. doi: 10.1016/j.molcel.2018.04.017.

Holmqvist, E., Berggren, S. and Rizvanovic, A. (2020) 'RNA-binding activity and regulatory functions of the emerging sRNA-binding protein ProQ', *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, 1863(9), p. 194596. doi: 10.1016/j.bbagrm.2020.194596.

Holmqvist, E. and Vogel, J. (2018) 'RNA-binding proteins in bacteria', *Nature Reviews Microbiology*. Springer US, 16(10), pp. 601–615. doi: 10.1038/s41579-018-0049-5.

Hrydziuszko, O. and Viant, M. R. (2012) 'Missing values in mass spectrometry based metabolomics: An undervalued step in the data processing pipeline', *Metabolomics*, 8, pp. 161–174. doi: 10.1007/s11306-011-0366-4.

Hu, P. *et al.* (1999) 'Sensing of nitrogen limitation by Bacillus subtilis: Comparison to enteric bacteria', *Journal of Bacteriology*, 181(16), pp. 5042–5050. doi: 10.1128/jb.181.16.5042-5050.1999.

Hudspeth, D. S. S. and Vary, P. S. (1992) 'spoVG sequence of Bacillus megaterium and Bacillus subtilis', *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression*, 1130(2), pp. 229–231. doi: 10.1016/0167-4781(92)90536-9.

Hummels, K. R. and Kearns, D. B. (2019) 'Suppressor mutations in ribosomal proteins and fliy restore Bacillus subtilis swarming motility in the absence of EF-P', *PLoS Genetics*, 15(6), pp. 1–27. doi: 10.1371/journal.pgen.1008179.

Illing, N. and Errington, J. (1991) 'The spoIIIA operon of Bacillus subtilis defines a new temporal class of mother-cell-specific sporulation genes under the control of the σE form of RNA polymerase', *Molecular Microbiology*, 5(8), pp. 1927–1940. doi: 10.1111/j.1365-2958.1991.tb00816.x.

Iosub, I. A. *et al.* (2020) 'Hfq CLASH uncovers sRNA-target interaction networks linked to nutrient availability adaptation', *eLife*, 9, pp. 1–33. doi: 10.7554/eLife.54655.

Itaya, M. *et al.* (2005) 'Combining two genomes in one cell: Stable cloning of the Synechocystis PCC6803 genome in the Bacillus subtilis 168 genome', *Proceedings of the National Academy of Sciences of the United States of America*, 102(44), pp. 15971–15976. doi: 10.1073/pnas.0503868102.

Jackson, D. W. *et al.* (2002) 'Biofilm formation and dispersal under the influence of the global regulator CsrA of Escherichia coli', *Journal of Bacteriology*, 184(1), pp. 290–301. doi: 10.1128/JB.184.1.290-301.2002.

Jahn, N. *et al.* (2012) 'BsrG/SR4 from Bacillus subtilis- the first temperature-

dependent type I toxin-antitoxin system', *Molecular Microbiology*, 83(3), pp. 579–598. doi: 10.1111/j.1365-2958.2011.07952.x.

Jahn, N. and Brantl, S. (2013) 'One antitoxin-two functions: SR4 controls toxin mRNA decay and translation', *Nucleic Acids Research*, 41(21), pp. 9870–9880. doi: 10.1093/nar/gkt735.

Jensen, K. B. and Darnell, R. B. (2008) 'CLIP: Crosslinking and immunoprecipitation of in vivo RNA targets of RNA-binding proteins', *Methods in Molecular Biology*, 488, pp. 85–98. doi: 10.1007/978-1-60327-475-3_6.

Johnson, W. C., Moran, C. P. and Losick, R. (1983) 'Two RNA polymerase sigma factors from Bacillus subtilis discriminate between overlapping promoters for a developmentally regulated gene.', *Nature*, 302, pp. 800–804. doi: 10.1038/302800a0.

Jousselin, A., Metzinger, L. and Felden, B. (2009) 'On the facultative requirement of the bacterial RNA chaperone, Hfq.', *Trends in microbiology*. Elsevier, 17(9), pp. 399–405. doi: 10.1016/j.tim.2009.06.003.

Jutras, B. L. *et al.* (2013) 'Eubacterial SpoVG Homologs Constitute a New Family of Site-Specific DNA-Binding Proteins', *PLoS ONE*, 8(6), p. e66683. doi: 10.1371/journal.pone.0066683.

Kajitani, M. *et al.* (1994) 'Regulation of the Escherichia coli hfq gene encoding the host factor for phage Q(β)', *Journal of Bacteriology*, 176(2), pp. 531–534. doi: 10.1128/jb.176.2.531-534.1994.

Kalman, S. *et al.* (1990) 'Similar organization of the sigB and spoIIA operons encoding alternate sigma factors of Bacillus subtilis RNA polymerase.', *Journal of Bacteriology*, 172(10), pp. 5575–5585. doi: 10.1128/JB.172.10.5575-5585.1990.

Karpievitch, Y. V., Dabney, A. R. and Smith, R. D. (2012) 'Normalization and missing value imputation for label-free LC-MS analysis.', *BMC bioinformatics*. BioMed Central Ltd, 13 Suppl 1(Suppl 16), p. S5. doi: 10.1186/1471-2105-13-S16-S5.

Kearns, D. B. and Losick, R. (2004) 'Swarming motility in undomesticated

Bacillus subtilis', *Molecular Microbiology*. John Wiley & Sons, Ltd (10.1111), 49(3), pp. 581–590. doi: 10.1046/j.1365-2958.2003.03584.x.

Keefer, A. B. *et al.* (2017) 'In vivo characterization of an Hfq protein encoded by the Bacillus anthracis virulence plasmid pXO1', *BMC Microbiology*. BMC Microbiology, 17(1), pp. 1–13. doi: 10.1186/s12866-017-0973-y.

Keene, J. D., Komisarow, J. M. and Friedersdorf, M. B. (2006) 'RIP-Chip: The isolation and identification of mRNAs, microRNAs and protein components of ribonucleoprotein complexes from cell extracts', *Nature Protocols*, 1(1), pp. 302–307. doi: 10.1038/nprot.2006.47.

Khanna, K. *et al.* (2019) 'The molecular architecture of engulfment during Bacillus subtilis sporulation', *eLife*, 8, pp. 1–22. doi: 10.7554/eLife.45257.

Khanna, K., Lopez-Garrido, J. and Pogliano, K. (2020) 'Shaping an Endospore: Architectural Transformations during Bacillus subtilis Sporulation', *Annual Review of Microbiology*, 74, pp. 361–386. doi: 10.1146/annurev-micro-022520-074650.

Kim, H.-H., Lee, B.-J. and Kwon, A.-R. (2010) 'Expression, crystallization, and preliminary X-ray crystallographic analysis of putative SpoVG from Staphylococcus aureus', *Archives of Pharmacal Research*, 33(8), pp. 1285–1288. doi: 10.1007/s12272-010-0820-2.

Kim, W. *et al.* (2019) 'Mechanisms for Hfq-Independent Activation of rpoS by DsrA, a Small RNA, in Escherichia coli', *Molecules and cells*, 42(5), pp. 426–439. doi: 10.14348/molcells.2019.0040.

Kimura, T. and Kobayashi, K. (2020) 'Role of Glutamate Synthase in Biofilm Formation by Bacillus subtilis', *Journal of Bacteriology*. Edited by G. O'Toole, 202(14). doi: 10.1128/JB.00120-20.

Kinsinger, R. F., Shirk, M. C. and Fall, R. (2003) 'Rapid Surface Motility in Bacillus subtilis Is Dependent on Extracellular Surfactin and Potassium Ion', *Society*, 185(18), pp. 5627–5631. doi: 10.1128/JB.185.18.5627.

Kleijn, R. J. *et al.* (2010) 'Metabolic fluxes during strong carbon catabolite repression by Malate in bacillus subtilis', *Journal of Biological Chemistry*, 285(3), pp. 1587–1596. doi: 10.1074/jbc.M109.061747.

Kobayashi, K. *et al.* (2003) 'Essential Bacillus subtilis genes', *Proceedings of the National Academy of Sciences of the United States of America*, 100(8), pp. 4678–4683. doi: 10.1073/pnas.0730515100.

Kobayashi, K. (2007) 'Gradual activation of the response regulator DegU controls serial expression of genes for flagellum formation and biofilm formation in Bacillus subtilis', *Molecular Microbiology*, 66(2), pp. 395–409. doi: 10.1111/j.1365-2958.2007.05923.x.

Kolter, R. and Greenberg, E. P. (2006) 'The superficial life of microbes', *Nature*, 441(7091), pp. 300–302. doi: 10.1038/441300a.

Koo, B. M. *et al.* (2017) 'Construction and Analysis of Two Genome-Scale Deletion Libraries for Bacillus subtilis', *Cell Systems*. Elsevier Inc., 4(3), pp. 291-305.e7. doi: 10.1016/j.cels.2016.12.013.

Kovach, A. R. *et al.* (2014) 'Recognition of U-rich RNA by Hfq from the Gram-positive pathogen Listeria monocytogenes', *Rna*, 20(10), pp. 1548–1559. doi: 10.1261/rna.044032.113.

Kovács, Á. T., Grau, R. and Pollitt, E. J. G. (2017) 'Surfing of bacterial droplets: Bacillus subtilis sliding revisited', *Proceedings of the National Academy of Sciences of the United States of America*, 114(42), p. E8802. doi: 10.1073/pnas.1710371114.

Krüger, S., Gertz, S. and Hecker, M. (1996) 'Transcriptional analysis of bglPH expression in Bacillus subtilis: Evidence for two distinct pathways mediating carbon catabolite repression', *Journal of Bacteriology*, 178(9), pp. 2637–2644. doi: 10.1128/jb.178.9.2637-2644.1996.

Kudla, G. *et al.* (2011) 'Cross-linking, ligation, and sequencing of hybrids reveals RNA-RNA interactions in yeast', *Proceedings of the National Academy of Sciences of the United States of America*. PNAS, 108(24), pp. 10010–10015. doi: 10.1073/pnas.1017386108.

Kulesus, R. R. *et al.* (2008) 'Impact of the RNA chaperone Hfq on the fitness and virulence potential of uropathogenic Escherichia coli', *Infection and Immunity*, 76(7), pp. 3019–3026. doi: 10.1128/IAI.00022-08.

Kunst, F. *et al.* (1997) 'The complete genome sequence of the Gram-

positive bacterium Bacillus subtilis', *Nature*, 390(6657), pp. 249–256. doi: 10.1038/36786.

Küster, E. *et al.* (1999) 'Mutations in catabolite control protein CcpA separating growth effects from catabolite repression', *Journal of Bacteriology*, 181(13), pp. 4125–4128. doi: 10.1128/jb.181.13.4125-4128.1999.

Kuwana, R. *et al.* (2005) 'The ylbO gene product of Bacillus subtilis is involved in the coat development and lysozyme resistance of spore', *FEMS Microbiology Letters*, 242(1), pp. 51–57. doi: 10.1016/j.femsle.2004.10.038.

Lai, S., Tremblay, J. and Déziel, E. (2009) 'Swarming motility: A multicellular behaviour conferring antimicrobial resistance', *Environmental Microbiology*, 11(1), pp. 126–136. doi: 10.1111/j.1462-2920.2008.01747.x.

Lamsa, A. *et al.* (2012) 'The Bacillus subtilis cannibalism toxin SDP collapses the proton motive force and induces autolysis', *Molecular Microbiology*, 84(3), pp. 486–500. doi: 10.1111/j.1365-2958.2012.08038.x.

de Las Rivas, J. and Fontanillo, C. (2010) 'Protein-protein interactions essentials: Key concepts to building and analyzing interactome networks', *PLoS Computational Biology*, 6(6), pp. 1–8. doi: 10.1371/journal.pcbi.1000807.

Lee, F. C. Y. and Ule, J. (2018) 'Advances in CLIP Technologies for Studies of Protein-RNA Interactions', *Molecular Cell*. doi: 10.1016/j.molcel.2018.01.005.

De Leersnyder, I. *et al.* (2018) 'Influence of growth media components on the antibacterial effect of silver ions on Bacillus subtilis in a liquid growth medium', *Scientific Reports*. Springer US, 8(1), p. 9325. doi: 10.1038/s41598-018-27540-9.

Lehnik-Habrink, M. *et al.* (2012) 'RNA degradation in Bacillus subtilis: An interplay of essential endo- and exoribonucleasesmmi', *Molecular Microbiology*, 84(6), pp. 1005–1017. doi: 10.1111/j.1365-2958.2012.08072.x.

Levdikov, V. M. *et al.* (2012) 'Structure of components of an intercellular

channel complex in sporulating Bacillus subtilis', 109(14), pp. 10–14. doi: 10.1073/pnas.1120087109.

Liang, Y.-H. *et al.* (2005) 'Protein Preparation, Crystallization and Preliminary X-Ray Crystallographic Studies of Dihydroorotase from Bacillus subtilis', *Protein & Peptide Letters*, 12(7), pp. 717–719. doi: 10.2174/0929866054696037.

Liu, X. *et al.* (2016) 'SpoVG Regulates Cell Wall Metabolism and Oxacillin Resistance in', *Antimicrobial Agents and Chemotherapy*. American Society for Microbiology, 60(6), pp. 3455–3461. doi: 10.1128/AAC.00026-16.Address.

Liu, X., Zhang, S. and Sun, B. (2016) 'SpoVG Regulates Cell Wall Metabolism and Oxacillin Resistance in Methicillin-Resistant Staphylococcus aureus Strain N315', *Antimicrobial Agents and Chemotherapy*, 60(6), pp. 3455–3461. doi: 10.1128/AAC.00026-16.

Lopez, D., Vlamakis, H. and Kolter, R. (2009) 'Generation of multiple cell types in Bacillus subtilis', *FEMS Microbiology Reviews*, 33(1), pp. 152–163. doi: 10.1111/j.1574-6976.2008.00148.x.

Lorca, G. L. *et al.* (2005) 'Catabolite Repression and Activation in Bacillus subtilis - Dependency on CcpA, HPr, and HprK', *Journal of Bacteriology*, 187(22), pp. 7826–7839. doi: 10.1128/JB.187.22.7826.

Losick, R. M. (2020) 'Bacillus subtilis: a bacterium for all seasons', *Current Biology*. Elsevier, 30(19), pp. R1146–R1150. doi: 10.1016/j.cub.2020.06.083.

Love, M. I., Huber, W. and Anders, S. (2014) 'Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2', *Genome Biology*, 15(12), pp. 1–21. doi: 10.1186/s13059-014-0550-8.

Lu, Z. *et al.* (2014) 'RIP-seq analysis of eukaryotic Sm proteins identifies three major categories of Sm-containing ribonucleoproteins', *Genome Biology*, 15(1), pp. 1–23. doi: 10.1186/gb-2014-15-1-r7.

Macek, B. *et al.* (2007) 'The Serine/Threonine/Tyrosine Phosphoproteome of the Model Bacterium Bacillus subtilis', *Molecular & Cellular Proteomics*,

6(4), pp. 697–707. doi: 10.1074/mcp.M600464-MCP200.

Mackay, J. P. *et al.* (2007) 'Protein interactions: is seeing believing?', *Trends in Biochemical Sciences*, 32(12), pp. 530–531. doi: 10.1016/j.tibs.2007.09.006.

Mäder, U. *et al.* (2002) 'Transcriptome and Proteome Analysis of Bacillus subtilis Gene Expression Modulated by Amino Acid Availability Transcriptome and Proteome Analysis of Bacillus subtilis Gene Expression Modulated by Amino Acid Availability', *Journal of bacteriology*, 184(15), pp. 4288–4295. doi: 10.1128/JB.184.15.4288.

Mäder, U. *et al.* (2012) 'SubtiWiki - A comprehensive community resource for the model organism Bacillus subtilis', *Nucleic Acids Research*, 40(D1), pp. 1278–1287. doi: 10.1093/nar/gkr923.

Maguire, B. A. *et al.* (2005) 'A protein component at the heart of an RNA machine: The importance of protein L27 for the function of the bacterial ribosome', *Molecular Cell*, 20(3), pp. 427–435. doi: 10.1016/j.molcel.2005.09.009.

Mann, M. (2006) 'Functional and quantitative proteomics using SILAC', *Nature Reviews Molecular Cell Biology*, 7(12), pp. 952–958. doi: 10.1038/nrm2067.

Marahiep, M. A., Republic, F. and Highway, K. (1991) 'Interaction of AbrB, a transcriptional regulator from Bacillus subtilis with the promoters of the transition state-activated genes tycA and spoVG', *Mol Gen Genet*, 225, pp. 347–354.

Margolis, P. S., Driks, A. and Losick, R. (1993) 'Sporulation gene spoIIB from Bacillus subtilis', *Journal of Bacteriology*, 175(2), pp. 528–540. doi: 10.1128/jb.175.2.528-540.1993.

Mars, R. A. T. *et al.* (2015) 'Small Regulatory RNA-Induced Growth Rate Heterogeneity of Bacillus subtilis', *PLoS Genetics*. Public Library of Science, 11(3). doi: 10.1371/journal.pgen.1005046.

Mars, R. A. T. *et al.* (2016) 'Regulatory RNAs in Bacillus subtilis: a Gram-Positive Perspective on Bacterial RNA-Mediated Regulation of Gene

Expression', *Microbiology and Molecular Biology Reviews*, 80(4), pp. 1029–1057. doi: 10.1128/mmbr.00026-16.

Martínez-Núñez, M. A. and López, V. E. L. y (2016) 'Nonribosomal peptides synthetases and their applications in industry', *Sustainable Chemical Processes*. Springer Nature, 4(1). doi: 10.1186/s40508-016-0057-6.

Martínez, L. C. and Vadyvaloo, V. (2014) 'Mechanisms of post-transcriptional gene regulation in bacterial biofilms', *Frontiers in Cellular and Infection Microbiology*, 5(MAR), pp. 1–15. doi: 10.3389/fcimb.2014.00038.

Matsuno, K. and Sonenshein, A. L. (1999) 'Role of SpoVG in asymmetric septation in Bacillus subtilis', *Journal of Bacteriology*, 181(11), pp. 3392–3401.

Mauri, M. and Klumpp, S. (2014) 'A Model for Sigma Factor Competition in Bacterial Cells', *PLoS Computational Biology*, 10(10), pp. 29–34. doi: 10.1371/journal.pcbi.1003845.

May, J. J., Wendrich, T. M. and Marahiel, M. A. (2001) 'The dhb Operon of Bacillus subtilis Encodes the Biosynthetic Template for the Catecholic Siderophore 2,3-Dihydroxybenzoate-Glycine-Threonine Trimeric Ester Bacillibactin', *Journal of Biological Chemistry*, 276(10), pp. 7209–7217. doi: 10.1074/jbc.M009140200.

McQuail, J. *et al.* (2020) 'The RNA-binding protein Hfq assembles into foci-like structures in nitrogen starved Escherichia coli', *Journal of Biological Chemistry*, 295(35), pp. 12355–12367. doi: 10.1074/jbc.RA120.014107.

Meißner, C., Jahn, N. and Brantl, S. (2016) 'In Vitro characterization of the type i toxin-antitoxin system bsrE/SR5 from Bacillus subtilis', *Journal of Biological Chemistry*, 291(2), pp. 560–571. doi: 10.1074/jbc.M115.697524.

Melamed, S. *et al.* (2016) 'Global Mapping of Small RNA-Target Interactions in Bacteria', *Molecular Cell*. The Authors, 63(5), pp. 884–897. doi: 10.1016/j.molcel.2016.07.026.

Melamed, S. *et al.* (2020) 'RNA-RNA Interactomes of ProQ and Hfq Reveal Overlapping and Competing Roles', *Molecular Cell*. Elsevier Inc., 77(2), pp. 411-425.e7. doi: 10.1016/j.molcel.2019.10.022.

Michna, R. H. *et al.* (2014) 'SubtiWiki-A database for the model organism Bacillus subtilis that links pathway, interaction and expression information', *Nucleic Acids Research*, 42(D1), pp. 692–698. doi: 10.1093/nar/gkt1002.

Michna, R. H. *et al.* (2016) '*Subti* Wiki 2.0—an integrated database for the model organism *Bacillus subtilis*', *Nucleic Acids Research*, 44(D1), pp. D654–D662. doi: 10.1093/nar/gkv1006.

Miethke, M. and Marahiel, M. A. (2007) 'Siderophore-Based Iron Acquisition and Pathogen Control', *Microbiology and Molecular Biology Reviews*, 71(3), pp. 413–451. doi: 10.1128/mmbr.00012-07.

Mohanty, B. K., Maples, V. F. and Kushner, S. R. (2004) 'The Sm-like protein Hfq regulates polyadenylation dependent mRNA decay in Escherichia coli', *Molecular Microbiology*, 54(4), pp. 905–920. doi: 10.1111/j.1365-2958.2004.04337.x.

Molle, V. *et al.* (2003) 'The Spo0A regulon of Bacillus subtilis', *Molecular Microbiology*, 50(5), pp. 1683–1701. doi: 10.1046/j.1365-2958.2003.03818.x.

Moon, K. and Gottesman, S. (2011) 'Competition among Hfq-binding small RNAs in Escherichia coli', *Molecular Microbiology*, 82(6), pp. 1545–1562. doi: 10.1111/j.1365-2958.2011.07907.x.

Moran Jr., C. P. *et al.* (1981) 'Promoter for a developmentally regulated gene in Bacillus subtilis', *Cell*, 25(3), pp. 783–791. doi: 0092-8674(81)90186-0 [pii].

Mordini, S. *et al.* (2013) 'The role of SwrA, DegU and PD3in fla/che expression in B. subtilis', *PLoS ONE*, 8(12), pp. 1–11. doi: 10.1371/journal.pone.0085065.

Morikawa, M. (2006) 'Beneficial biofilm formation by industrial bacteria Bacillus subtilis and related species', *Journal of Bioscience and Bioengineering*, 101(1), pp. 1–8. doi: 10.1263/jbb.101.1.

Morita, T., Maki, K. and Aiba, H. (2006) 'Erratum: RNase E-based ribonucleoprotein complexes: Mechanical basis of mRNA destabilization mediated by bacterial noncoding RNAs (Genes and Development (2005)

19, (2176-2186))', *Genes and Development*, 20(24), p. 3487. doi: 10.1101/gad.1330406.

Mota, L. J., Sarmento, L. M. and Sa-Nogueira, I. (2001) 'Control of the Arabinose Regulon in Bacillus subtilis by AraR In Vivo: Crucial Roles of Operators, Cooperativity, and DNA Looping', *Society*, 183(14), pp. 4190–4201. doi: 10.1128/JB.183.14.4190.

Mota, L. J., Tavares, P. and Sá-Noguelra, I. (1999) 'Mode of action of AraR, the key regulator of L-arabinose metabolism in Bacillus subtilis', *Molecular Microbiology*, 33(3), pp. 476–489. doi: 10.1046/j.1365-2958.1999.01484.x.

Mukherjee, S. and Kearns, D. B. (2014) 'The Structure and Regulation of Flagella in *Bacillus subtilis*', *Annual Review of Genetics*, 48(1), pp. 319–340. doi: 10.1146/annurev-genet-120213-092406.

Nagler, K. *et al.* (2016) 'Identification of differentially expressed genes during Bacillus subtilis spore outgrowth in high-salinity environments using RNA sequencing', *Frontiers in Microbiology*. Frontiers Media S.A., 7(OCT), p. 1564. doi: 10.3389/fmicb.2016.01564.

Nagórska, K. *et al.* (2008) 'Influence of the σB stress factor and yxaB, the gene for a putative exopolysaccharide synthase under σB control, on biofilm formation', *Journal of Bacteriology*, 190(10), pp. 3546–3556. doi: 10.1128/JB.01665-07.

Nakano, M. M. *et al.* (1991) 'srfA is an operon required for surfactin production, competence development, and efficient sporulation in Bacillus subtilis', *Journal of Bacteriology*, 173(5), pp. 1770–1778. doi: 10.1128/jb.173.5.1770-1778.1991.

Nakano, M. M., Xia, L. and Zuber, P. (1991) 'Transcription initiation region of the srfA operon, which is controlled by the comP-comA signal transduction system in Bacillus subtilis', *Journal of Bacteriology*, 173(17), pp. 5487–5493. doi: 10.1128/jb.173.17.5487-5493.1991.

Nakano, M. M. and Zuber, P. (1990) 'Identification of Genes Required for the Biosynthesis of the Lipopeptide Antibiotic Surfactin in Bacillus Subtilis', *Genetics and Biotechnology of Bacilli*, 170(12), pp. 397–405. doi: 10.1016/b978-0-12-274162-3.50047-5.

Nanamiya, H. and Kawamuray, F. (2010) 'Towards an elucidation of the roles of the ribosome during different growth phases in bacillus subtilis', *Bioscience, Biotechnology and Biochemistry*, 74(3), pp. 451–461. doi: 10.1271/bbb.90859.

Nannapaneni, P. *et al.* (2012) 'Defining the structure of the general stress regulon of Bacillus subtilis using targeted microarray analysis and random forest classification', *Microbiology*, 158(3), pp. 696–707. doi: 10.1099/mic.0.055434-0.

Narula, J. *et al.* (2012) 'Ultrasensitivity of the Bacillus subtilis sporulation decision', *Proceedings of the National Academy of Sciences of the United States of America*, 109(50), pp. E3513–E3522. doi: 10.1073/pnas.1213974109.

Ng Kwan Lim, E. *et al.* (2021) 'Keeping Up with RNA-Based Regulation in Bacteria: New Roles for RNA Binding Proteins', *Trends in Genetics*. Elsevier, 37(1), pp. 86–97. doi: 10.1016/j.tig.2020.09.014.

Nicholson, W. L. *et al.* (2000) 'Resistance of Bacillus Endospores to Extreme Terrestrial and Extraterrestrial Environments', *Microbiology and Molecular Biology Reviews*, 64(3), pp. 548–572. doi: 10.1128/mmbr.64.3.548-572.2000.

Nicolas, P. *et al.* (2012) 'Condition-Dependent Transcriptome Reveals High-Level Regulatory Architecture in Bacillus subtilis', *Science*, 335(6072), pp. 1103–1106. doi: 10.1126/science.1206848.

Nielsen, J. S. *et al.* (2009) 'Defining a role for Hfq in Gram-positive bacteria: Evidence for Hfq-dependent antisense regulation in Listeria monocytogenes', *Nucleic Acids Research*, 38(3), pp. 907–919. doi: 10.1093/nar/gkp1081.

Nishikiori, T. *et al.* (1986) 'Plipastatins: New Inhibitors of Phospholipase A2Produced by Bacillus Cereus BMG302-fF67 III. Structural Elucidation of Plipastatins', *The Journal of Antibiotics*, 39(6), pp. 755–761. doi: 10.7164/antibiotics.39.755.

Noone, D. *et al.* (2001) 'YkdA and YvtA, HtrA-Like Serine Proteases inBacillus subtilis, Engage in Negative Autoregulation and Reciprocal

Cross-Regulation of ykdA and yvtAGene Expression', *Journal of Bacteriology*, 183(2), pp. 654–663. doi: 10.1128/JB.183.2.654-663.2001.

Van Nues, R. *et al.* (2017) 'Kinetic CRAC uncovers a role for Nab3 in determining gene expression profiles during stress', *Nature Communications*. Springer US, 8(1), pp. 1–17. doi: 10.1038/s41467-017-00025-5.

O'Brien, J. J. *et al.* (2018) 'The effects of nonignorable missing data on label-free mass spectrometry proteomics experiments', *The Annals of Applied Statistics*, 12(4), pp. 2075–2095. doi: 10.1214/18-AOAS1144.

Ogura, M. *et al.* (2002) 'Whole-genome analysis of genes regulated by the Bacillus subtilis competence transcription factor ComK', *Journal of Bacteriology*, 184(9), pp. 2344–2351. doi: 10.1128/JB.184.9.2344-2351.2002.

Ohashi, Y. *et al.* (2003) 'Expression profiling of translation-associated genes in sporulating Bacillus subtilis and consequence of sporulation by gene inactivation', *Bioscience, Biotechnology and Biochemistry*, 67(10), pp. 2245–2253. doi: 10.1271/bbb.67.2245.

Oman, T. J. *et al.* (2011) 'Sublancin is not a lantibiotic but an S-linked glycopeptide', *Nature Chemical Biology*. Nature Publishing Group, 7(2), pp. 78–80. doi: 10.1038/nchembio.509.

Overkamp, W. and Kuipers, O. P. (2015) 'Transcriptional profile of Bacillus subtilis sigF-mutant during vegetative growth', *PLoS ONE*, 10(10), pp. 1–12. doi: 10.1371/journal.pone.0141553.

Pagliuso, A. *et al.* (2019) 'An RNA-Binding Protein Secreted by a Bacterial Pathogen Modulates RIG-I Signaling', *Cell Host and Microbe*. Cell Press, 26(6), pp. 823-835.e11. doi: 10.1016/j.chom.2019.10.004.

Panja, S., Schu, D. J. and Woodson, S. A. (2013) 'Conserved arginines on the rim of Hfq catalyze base pair formation and exchange', *Nucleic Acids Research*, 41(15), pp. 7536–7546. doi: 10.1093/nar/gkt521.

Patrick, J. E. and Kearns, D. B. (2009) 'Laboratory strains of Bacillus subtilis do not exhibit swarming motility', *Journal of Bacteriology*, 191(22), pp.

7129–7133. doi: 10.1128/JB.00905-09.

Patrick, J. E. and Kearns, D. B. (2012) 'Swarming motility and the control of master regulators of flagellar biosynthesis', *Molecular Microbiology*, 83(1), pp. 14–23. doi: 10.1111/j.1365-2958.2011.07917.x.

Perego, M. *et al.* (1991) 'The oligopeptide transport system of Bacillus subtilis plays a role in the initiation of sporulation', *Molecular Microbiology*, 5(1), pp. 173–185. doi: 10.1111/j.1365-2958.1991.tb01838.x.

Perez, A. R., Abanes-De Mello, A. and Pogliano, K. (2000) 'SpoIIB localizes to active sites of septal biogenesis and spatially regulates septal thinning during engulfment in Bacillus subtilis', *Journal of Bacteriology*, 182(4), pp. 1096–1108. doi: 10.1128/JB.182.4.1096-1108.2000.

Perez, A. R., Abanes-De Mello, A. and Pogliano, K. (2006) 'Suppression of engulfment defects in Bacillus subtilis by elevated expression of the motility regulon', *Journal of Bacteriology*. doi: 10.1128/JB.188.3.1159-1164.2006.

Petersohn, A. *et al.* (2001) 'Global analysis of the general stress response of Bacillus subtilis', *Journal of Bacteriology*, 183(19), pp. 5617–5631. doi: 10.1128/JB.183.19.5617-5631.2001.

Petschnigg, J., Snider, J. and Stagljar, I. (2011) 'Interactive proteomics research technologies: Recent applications and advances', *Current Opinion in Biotechnology*. Elsevier Ltd, 22(1), pp. 50–58. doi: 10.1016/j.copbio.2010.09.001.

Piggot, P. J. and Coote, J. G. (1976) 'Genetic aspects of bacterial endospore formation.', *Bacteriological Reviews*, 40(4), pp. 908–962. doi: 10.1128/MMBR.40.4.908-962.1976.

Potts, A. H. *et al.* (2017) 'Global role of the bacterial post-transcriptional regulator CsrA revealed by integrated transcriptomics', *Nature Communications*. Springer US, 8(1). doi: 10.1038/s41467-017-01613-1.

Potts, A. H. *et al.* (2019) 'Role of CsrA in stress responses and metabolism important for Salmonella virulence revealed by integrated transcriptomics', *PLoS ONE*, 14(1), pp. 1–30. doi: 10.1371/journal.pone.0211430.

Presecan, E. *et al.* (1997) 'The Bacilus subtilis genome from gerBC (311°)

to licR (334°)', *Microbiology*, 143(10), pp. 3313–3328. doi: 10.1099/00221287-143-10-3313.

Price, C. W. *et al.* (2001) 'Genome-wide analysis of the general stress response in Bacillus subtilis', *Molecular Microbiology*, 41(4), pp. 757–774. doi: 10.1046/j.1365-2958.2001.02534.x.

Price, C. W. (2002) 'General Stress Response', *Bacillus subtilis and Its Closest Relatives*, pp. 369–384. doi: 10.1128/9781555817992.ch26.

Provvedi, R., Chen, I. and Dubnau, D. (2001) 'NucA is required for DNA cleavage during transformation of Bacillus subtilis', *Molecular Microbiology*, 40(3), pp. 634–644. doi: 10.1046/j.1365-2958.2001.02406.x.

Quendera, A. P. *et al.* (2020) 'RNA-Binding Proteins Driving the Regulatory Activity of Small Non-coding RNAs in Bacteria', *Frontiers in Molecular Biosciences*, 7(May), pp. 1–9. doi: 10.3389/fmolb.2020.00078.

Rabhi, M. *et al.* (2011) 'The Sm-like RNA chaperone Hfq mediates transcription antitermination at Rho-dependent terminators', *EMBO Journal*, 30(14), pp. 2805–2816. doi: 10.1038/emboj.2011.192.

Rajkowitsch, L. and Schroeder, R. (2007) 'Dissecting RNA chaperone activity', *Rna*, 13(12), pp. 2053–2060. doi: 10.1261/rna.671807.

Ramanathan, M. *et al.* (2019) 'Methods to study RNA-protein interactions', 16(3), pp. 225–234. doi: 10.1038/s41592-019-0330-1.Methods.

Ramaniuk, O. *et al.* (2017) 'Kinetic modelling and meta-analysis of the B. subtilis SigA regulatory network during spore germination and outgrowth', *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, 1860(8), pp. 894–904. doi: 10.1016/j.bbagrm.2017.06.003.

Ramaniuk, O. *et al.* (2018) 'σI from Bacillus subtilis : Impact on Gene Expression and Characterization of σI -Dependent Transcription That Requires', *Journal of bacteriology*, 200(17), pp. 1–15.

Rappsilber, J., Mann, M. and Ishihama, Y. (2007) 'Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips', *Nature Protocols*, 2(8), pp. 1896–1906. doi: 10.1038/nprot.2007.261.

Ren, C. *et al.* (2012) 'Pleiotropic functions of catabolite control protein CcpA in Butanol-producing Clostridium acetobutylicum', *BMC Genomics*, 13(1). doi: 10.1186/1471-2164-13-349.

Resnekov, O. *et al.* (1995) 'Identification and characterization of sporuation gene spoVS from Bacillus subtilis', *J Bacteriol*, 177(19), pp. 5628–5635.

Resnekov, O., Driks, A. and Losick, R. (1995) 'Identification and characterization of sporulation gene spoVS from Bacillus subtilis', *Journal of Bacteriology*. doi: 10.1128/jb.177.19.5628-5635.1995.

Reva, O. N. *et al.* (2004) 'Taxonomic characterization and plant colonizing abilities of some bacteria related to Bacillus amyloliquefaciens and Bacillus subtilis', *FEMS Microbiology Ecology*, 48(2), pp. 249–259. doi: 10.1016/j.femsec.2004.02.003.

Rochat, T. *et al.* (2015) 'Tracking the elusive function of Bacillus subtilis Hfq', *PLoS ONE*, 10(4), pp. 1–18. doi: 10.1371/journal.pone.0124977.

Romeo, T. (1998) 'Global regulation by the small RNA-binding protein CsrA and the non- coding RNA molecule CsrB', *Molecular Microbiology*, 29(6), pp. 1321–1330. doi: 10.1046/j.1365-2958.1998.01021.x.

Rosenbluh, A. *et al.* (1981) 'Identification of a new developmental locus in Bacillus subtilis by construction of a deletion mutation in a cloned gene under sporulation control', *Journal of Bacteriology*, 148(1), pp. 341–351. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC216198/pdf/jbacter00263-0349.pdf (Accessed: 31 May 2019).

Röst, H. L. *et al.* (2016) 'OpenMS: A flexible open-source software platform for mass spectrometry data analysis', *Nature Methods*, 13(9), pp. 741–748. doi: 10.1038/nmeth.3959.

Rowland, B. M. *et al.* (1996) 'Sequence and genetic organization of a Bacillus subtilis operon encoding 2,3-dihydroxybenzoate biosynthetic enzymes', *Gene*, 178(1–2), pp. 119–123. doi: 10.1016/0378-1119(96)00349-6.

Saito, S., Kakeshita, H. and Nakamura, K. (2008) 'Novel small RNA-

encoding genes in the intergenic regions of Bacillus subtilis'. doi: 10.1016/j.gene.2008.09.024.

Salim, N. N. *et al.* (2012) 'Requirement of upstream Hfq-binding (ARN)x elements in glmS and the Hfq C-terminal region for GlmS upregulation by sRNAs GlmZ and GlmY', *Nucleic Acids Research*, 40(16), pp. 8021–8032. doi: 10.1093/nar/gks392.

Sandman, K., Losick, R. and Youngman, P. (1987) 'Genetic analysis of Bacillus subtilis spo mutations generated by Tn917-mediated insertional mutagenesis.', *Genetics*, 117(4), pp. 603–617. doi: 10.1128/JB.01343-06.

Santiago-Frangos, A. and Woodson, S. A. (2018) 'Hfq chaperone brings speed dating to bacterial sRNA', *Wiley Interdisciplinary Reviews: RNA*, 9(4), pp. 1–16. doi: 10.1002/wrna.1475.

Satomura, T. *et al.* (2005) 'Enhancement of glutamine utilization in Bacillus subtilis through the GlnK-GlnL two-component regulatory system', *Journal of Bacteriology*, 187(14), pp. 4813–4821. doi: 10.1128/JB.187.14.4813-4821.2005.

Sauder, A. B. and Kendall, M. M. (2018) 'After the fact(or): Posttranscriptional gene regulation in enterohemorrhagic Escherichia coli O157:H7', *Journal of Bacteriology*, pp. 1–16. doi: 10.1128/JB.00228-18.

Savage, C. R. *et al.* (2018) 'Borrelia burgdorferi SpoVG DNA- and RNA-binding protein modulates the physiology of the Lyme disease spirochete', *Journal of Bacteriology*, 200(12). doi: 10.1128/JB.00033-18.

Schnetz, K. *et al.* (1996) 'LicT, a Bacillus subtilis transcriptional antiterminator protein of the BglG family', *Journal of Bacteriology*, 178(7), pp. 1971–1979. doi: 10.1128/jb.178.7.1971-1979.1996.

Schuhmacher, J. S., Thormann, K. M. and Bange, G. (2015) 'How bacteria maintain location and number of flagella?', *FEMS Microbiology Reviews*, 39(6), pp. 812–822. doi: 10.1093/femsre/fuv034.

Schulthess, B. *et al.* (2009) 'Functional characterization of the σB-dependent yabJ-spoVG operon in Staphylococcus aureus: Role in methicillin and glycopeptide resistance', *Antimicrobial Agents and*

Chemotherapy, 53(5), pp. 1832–1839. doi: 10.1128/AAC.01255-08.

Schulthess, B. *et al.* (2011) 'The σ Bdependent yabJ-spoVG operon is involved in the regulation of extracellular nuclease, lipase, and protease expression', *Journal of Bacteriology*, 193(18), pp. 4954–4962. doi: 10.1128/JB.05362-11.

Schultz, D. *et al.* (2009) 'Deciding fate in adverse times: Sporulation and competence in Bacillus subtilis', *Proceedings of the National Academy of Sciences of the United States of America*, 106(50), pp. 21027–21034. doi: 10.1073/pnas.0912185106.

Schumann, W. (2016) 'Regulation of bacterial heat shock stimulons', *Cell Stress and Chaperones*. Cell Stress and Chaperones, pp. 959–968. doi: 10.1007/s12192-016-0727-z.

Sd-nlogueira, I. *et al.* (1997) 'Nucleotide Sequence , Genetic Organization and Expression', *Microbiology*, 143(1997), pp. 957–969.

Segall, J. and Losick, R. (1977) 'Cloned bacillus subtilis DNA containing a gene that is activated early during sporulation', *Cell*, 11(4), pp. 751–761. doi: 10.1016/0092-8674(77)90289-6.

Senesi, S. *et al.* (2004) 'Surface-Associated Flagellum Formation and Swarming Differentiation in Bacillus subtilis Are Controlled by the ifm Locus', *Journal of Bacteriology*, 186(4), pp. 1158–1164. doi: 10.1128/JB.186.4.1158-1164.2004.

Serganov, A. and Nudler, E. (2013) 'A decade of riboswitches', *Cell*. Elsevier Inc., 152(1–2), pp. 17–24. doi: 10.1016/j.cell.2012.12.024.

Sheidy, D. T. and Zielke, R. A. (2013) 'Analysis and Expansion of the Role of the Escherichia coli Protein ProQ', *PLoS ONE*, 8(10), pp. 1–11. doi: 10.1371/journal.pone.0079656.

Shemesh, M. and Chai, Y. (2013) 'A Combination of Glycerol and Manganese Promotes Biofilm Formation in Bacillus subtilis via Histidine Kinase KinD Signaling'. doi: 10.1128/JB.00028-13.

Shishkin, A. A. *et al.* (2015) 'Simultaneous generation of many RNA-seq libraries in a single reaction', *Nature Methods*. Nature Publishing Group,

12(4), pp. 323–325. doi: 10.1038/nmeth.3313.

van Sinderen, D. *et al.* (1994) 'Molecular cloning and sequence of comK, a gene required for genetic competence in Bacillus subtilis', *Molecular Microbiology*, 11(4), pp. 695–703. doi: 10.1111/j.1365-2958.1994.tb00347.x.

Smirnov, A. *et al.* (2016) 'Grad-seq guides the discovery of ProQ as a major small RNA-binding protein', *Proceedings of the National Academy of Sciences*, 113(41), pp. 11591–11596. doi: 10.1073/pnas.1609981113.

Sobrero, P. and Valverde, C. (2012) 'The bacterial protein Hfq: Much more than a mere RNA-binding factor', *Critical Reviews in Microbiology*, 38(4), pp. 276–299. doi: 10.3109/1040841X.2012.664540.

Solomon, J. *et al.* (2003) 'Isolation and Characterization of Mutants of the Bacillus subtilis Oligopeptide Permease with Altered Specificity of Oligopeptide Transport', *ScienceDaily*, 185(21), pp. 6425–6433. doi: 10.1128/JB.185.21.6425.

Somerville, G., Mikoryak, C. A. and Reitzer, L. (1999) 'Physiological characterization of Pseudomonas aeruginosa during exotoxin A synthesis: Glutamate, iron limitation, and aconitase activity', *Journal of Bacteriology*, 181(4), pp. 1072–1078. doi: 10.1128/jb.181.4.1072-1078.1999.

Soper, T. *et al.* (2010) 'Positive regulation by small RNAs and the role of Hfq', *Proceedings of the National Academy of Sciences of the United States of America*, 107(21), pp. 9602–9607. doi: 10.1073/pnas.1004435107.

Speck, E. L. and Freese, E. (1973) 'Control of metabolite secretion in Bacillus subtilis', *Journal of General Microbiology*, 78(2), pp. 261–275. doi: 10.1099/00221287-78-2-261.

Van Der Steen, J. B. and Hellingwerf, K. J. (2015) 'Activation of the General Stress Response of Bacillus subtilis by Visible Light', *Photochemistry and Photobiology*, 91(5), pp. 1032–1045. doi: 10.1111/php.12499.

Steil, L. *et al.* (2005) 'Genome-wide analysis of temporally regulated and compartment-specific gene expression in sporulating cells of Bacillus subtilis', *Microbiology*, 151(2), pp. 399–420. doi: 10.1099/mic.0.27493-0.

Stein, T. *et al.* (2004) 'Subtilosin Production by Two Bacillus subtilis Subspecies and Variance of the sbo-alb Cluster', *Applied and Environmental Microbiology*, 70(4), pp. 2349–2353. doi: 10.1128/AEM.70.4.2349-2353.2004.

Stein, T. (2005) 'Bacillus subtilis antibiotics: Structures, syntheses and specific functions', *Molecular Microbiology*, 56(4), pp. 845–857. doi: 10.1111/j.1365-2958.2005.04587.x.

Steinberg, N. *et al.* (2020) 'The extracellular matrix protein TasA is a developmental cue that maintains a motile subpopulation within Bacillus subtilis biofilms', *SCIENCE SIGNALING*, 13(eaaw8905), pp. 1–15.

Sterlini, J. M. and Mandelstam, J. (1969) 'Commitment to sporulation in *Bacillus subtilis* and its relationship to development of actinomycin resistance.', *The Biochemical journal*, 113(1), pp. 29–37.

Sternburg, E. L. and Karginov, F. V. (2020) 'Global Approaches in Studying RNA-Binding Protein Interaction Networks', *Trends in Biochemical Sciences*. Elsevier Ltd, pp. 1–11. doi: 10.1016/j.tibs.2020.03.005.

Storz, G., Vogel, J. and Wassarman, K. M. (2011) 'Regulation by Small RNAs in Bacteria: Expanding Frontiers', *Molecular Cell*, 43(6), pp. 880–891. doi: 10.1016/j.molcel.2011.08.022.

Straight, P. D. *et al.* (2007) 'A singular enzymatic megacomplex from Bacillus subtilis', *Proceedings of the National Academy of Sciences of the United States of America*. PNAS, 104(1), pp. 305–310. doi: 10.1073/pnas.0609073103.

Strauch, M. A. *et al.* (2007) 'Abh and AbrB control of Bacillus subtilis antimicrobial gene expression', *Journal of Bacteriology*, 189(21), pp. 7720–7732. doi: 10.1128/JB.01081-07.

Stülke, J. (2002) 'Control of transcription termination in bacteria by RNA-binding proteins that modulate RNA structures', *Archives of Microbiology*, 177(6), pp. 433–440. doi: 10.1007/s00203-002-0407-5.

Stülke, J. and Hillen, W. (2000) 'Regulation of Carbon Catabolism in Bacillus Species', *Annual Review of Microbiology*, 54(1), pp. 849–880. doi:

10.1146/annurev.micro.54.1.849.

Su, M. *et al.* (2015) '3-Hydroxypropionaldehyde-specific aldehyde dehydrogenase from Bacillus subtilis catalyzes 3-hydroxypropionic acid production in Klebsiella pneumoniae', *Biotechnology Letters*, 37(3), pp. 717–724. doi: 10.1007/s10529-014-1730-z.

Su, Y. *et al.* (2020) 'Bacillus subtilis: A universal cell factory for industry, agriculture, biomaterials and medicine', *Microbial Cell Factories*. BioMed Central, 19(1), pp. 1–12. doi: 10.1186/s12934-020-01436-8.

Sukhodolets, M. V. and Garges, S. (2003) 'Interaction of Escherichia coli RNA polymerase with the ribosomal protein S1 and the Sm-like ATPase Hfq', *Biochemistry*, 42(26), pp. 8022–8034. doi: 10.1021/bi020638i.

Sun, D., Cabrera-Martinez, R. M. and Setlow, P. (1991) 'Control of transcription of the Bacillus subtilis spoIIIG gene, which codes for the forespore-specific transcription factor σ(G)', *Journal of Bacteriology*, 173(9), pp. 2977–2984. doi: 10.1128/jb.173.9.2977-2984.1991.

Sutherland, B. W., Toews, J. and Kast, J. (2008) 'Utility of formaldehyde cross-linking and mass spectrometry in the study of protein-protein interactions', *Journal of Mass Spectrometry*, 43(6), pp. 699–715. doi: 10.1002/jms.1415.

Szklarczyk, D. *et al.* (2019) 'STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets', *Nucleic Acids Research*. Oxford University Press, 47(D1), pp. D607–D613. doi: 10.1093/nar/gky1131.

Takada, A. *et al.* (1997) 'DNA binding properties of the hfq gene product of Escherichia coli', *Biochemical and Biophysical Research Communications*, 236(3), pp. 576–579. doi: 10.1006/bbrc.1997.7013.

Takada, H. *et al.* (2014) 'Cell motility and biofilm formation in Bacillus subtilis are affected by the ribosomal proteins, S11 and S21', *Bioscience, Biotechnology and Biochemistry*. Taylor & Francis, 78(5), pp. 898–907. doi: 10.1080/09168451.2014.915729.

Tan, I. S. and Ramamurthi, K. S. (2014) 'Spore formation in Bacillus subtilis',

Environmental Microbiology Reports, pp. 212–225. doi: 10.1111/1758-2229.12130.

Tang, M. R. *et al.* (2006) 'Use of transcriptional profiling & bioinformatics to solve production problems', *Industrial Biotechnology*, 2(1), pp. 66–74. doi: 10.1089/ind.2006.2.66.

Tanovic, A. *et al.* (2008) 'Crystal structure of the termination module of a nonribosomal peptide synthetase', *Science*, 321(5889), pp. 659–663. doi: 10.1126/science.1159850.

Tjalsma, H. *et al.* (2000) 'Signal Peptide-Dependent Protein Transport inBacillus subtilis: a Genome-Based Survey of the Secretome', *Microbiology and Molecular Biology Reviews*, 64(3), pp. 515–547. doi: 10.1128/mmbr.64.3.515-547.2000.

Tjalsma, H. *et al.* (2004) 'Proteomics of Protein Secretion by Bacillus subtilis: Separating the "Secrets" of the Secretome', *Microbiology and Molecular Biology Reviews*, 68(2), pp. 207–233. doi: 10.1128/MMBR.68.2.207.

Tosato, V. *et al.* (1997) 'Sequence completion, identification and definition of the fengycin operon in Bacillus subtilis 168', *Microbiology*, 143(11), pp. 3443–3450. doi: 10.1099/00221287-143-11-3443.

Trach, K. and Hoch, J. A. (1989) 'The Bacillus subtilis spo0B Stage 0 sporulation operon encodes an essential GTP-binding protein', *Journal of Bacteriology*, 171(3), pp. 1362–1371. doi: 10.1128/jb.171.3.1362-1371.1989.

Tsuge, K. *et al.* (1999) 'The genes degQ, pps, and lpa-8 (sfp) are responsible for conversion of Bacillus subtilis 168 to plipastatin production', *Antimicrobial Agents and Chemotherapy*, 43(9), pp. 2183–2192. doi: 10.1128/aac.43.9.2183.

Tsuge, K., Matsui, K. and Itaya, M. (2007) 'Production of the non-ribosomal peptide plipastatin in Bacillus subtilis regulated by three relevant gene blocks assembled in a single movable DNA segment', *Journal of Biotechnology*, 129(4), pp. 592–603. doi: 10.1016/j.jbiotec.2007.01.033.

Tsui, H. T., Leung, H. E. and Winkler, M. E. (1994) 'Characterization of broadly pleiotropic phenotypes caused by an hfq insertion mutation in Escherichia coli K-12', *Molecular Microbiology*, 13(1), pp. 35–49. doi: 10.1111/j.1365-2958.1994.tb00400.x.

Turner, R. J., Lu, Y. and Switzer, R. L. (1994) 'Regulation of the Bacillus subtilis pyrimidine biosynthetic (pyr) gene cluster by an autogenous transcriptional attenuation mechanism', *Journal of Bacteriology*, 176(12), pp. 3708–3722. doi: 10.1128/jb.176.12.3708-3722.1994.

Tyanova, S. *et al.* (2016) 'The Perseus computational platform for comprehensive analysis of (prote)omics data', *Nature Methods*, 13(9), pp. 731–740. doi: 10.1038/nmeth.3901.

Tyanova, S. and Cox, J. (2018) 'Perseus: A bioinformatics platform for integrative analysis of proteomics data in cancer research', in *Methods in Molecular Biology*. doi: 10.1007/978-1-4939-7493-1_7.

Tyanova, S., Temu, T. and Cox, J. (2016) 'The MaxQuant computational platform for mass spectrometry-based shotgun proteomics', *Nature Protocols*. Nature Publishing Group, 11(12), pp. 2301–2319. doi: 10.1038/nprot.2016.136.

Uffen, R. L. and Canale-Parola, E. (1972) 'Synthesis of pulcherriminic acid by Bacillus subtilis.', *Journal of Bacteriology*, 111(1), pp. 86–93. doi: 10.1128/jb.111.1.86-93.1972.

Ul Haq, I., Müller, P. and Brantl, S. (2020) 'Intermolecular Communication in Bacillus subtilis: RNA-RNA, RNA-Protein and Small Protein-Protein Interactions', *Frontiers in Molecular Biosciences*, 7(August), pp. 1–17. doi: 10.3389/fmolb.2020.00178.

Ule, J. *et al.* (2003) 'CLIP Identifies Nova-Regulated RNA Networks in the Brain', *Science*, 302(5648), pp. 1212–1215. doi: 10.1126/science.1090095.

Ule, J. *et al.* (2005) 'CLIP: A method for identifying protein-RNA interaction sites in living cells', *Methods*, 37(4), pp. 376–386. doi: 10.1016/j.ymeth.2005.07.018.

Urdaneta, E. C. *et al.* (2019) 'Purification of cross-linked RNA-protein

complexes by phenol-toluol extraction', *Nature Communications*, 10(1). doi: 10.1038/s41467-019-08942-3.

Vlamakis, H. *et al.* (2013) 'Sticking together: building a biofilm the Bacillus subtilis way', *Nature Reviews Microbiology*, 11(3), pp. 157–168. doi: 10.1038/nrmicro2960.

VO, P. N. L. *et al.* (2021) 'Optimized expression of Hfq protein increases Escherichia coli growth', *Journal of Biological Engineering*, 15(1), p. 7. doi: 10.1186/s13036-021-00260-x.

Vogel, J. and Luisi, B. F. (2011) 'Hfq and its constellation of RNA', *Nature Reviews Microbiology*, 9(8), pp. 578–589. doi: 10.1038/nrmicro2615.

Wacker, I. *et al.* (2003) 'The regulatory link between carbon and nitrogen metabolism in Bacillus subtilis: Regulation of the gltAB operon by the catabolite control protein CcpA', *Microbiology*, 149(10), pp. 3001–3009. doi: 10.1099/mic.0.26479-0.

Wang, H. *et al.* (2014) 'Atlas of nonribosomal peptide and polyketide biosynthetic pathways reveals common occurrence of nonmodular enzymes', *Proceedings of the National Academy of Sciences of the United States of America*, 111(25), pp. 9259–9264. doi: 10.1073/pnas.1401734111.

Wang, P. Z. and Doi, R. H. (1984) 'Overlapping promoters transcribed by Bacillus subtilis sigma 55 and sigma 37 RNA polymerase holoenzymes during growth and stationary phases', *Journal of Biological Chemistry*, 259(13), pp. 8619–8625.

Wang, S. T. *et al.* (2006) 'The forespore line of gene expression in Bacillus subtilis', *Journal of Molecular Biology*, 358(1), pp. 16–37. doi: 10.1016/j.jmb.2006.01.059.

Warner, J. B. and Lolkema, J. S. (2003) 'CcpA-Dependent Carbon Catabolite Repression in Bacteria', *Microbiology and Molecular Biology Reviews*, 67(4), pp. 475–490. doi: 10.1128/mmbr.67.4.475-490.2003.

Wassarman, K. M. (2007) '6S RNA: A regulator of transcription', *Molecular Microbiology*, 65(6), pp. 1425–1431. doi: 10.1111/j.1365-

2958.2007.05894.x.

Waters, L. S. and Storz, G. (2009) 'Regulatory RNAs in Bacteria', *Cell*, 136(4), pp. 615–628. doi: 10.1016/j.cell.2009.01.043.

Wei, R. *et al.* (2018) 'Missing Value Imputation Approach for Mass Spectrometry-based Metabolomics Data', *Scientific Reports*. Springer US, 8(1), pp. 1–10. doi: 10.1038/s41598-017-19120-0.

Westermann, A. J. *et al.* (2019) 'The major RNA-binding protein ProQ impacts virulence gene expression in salmonella enterica serovar typhimurium', *mBio*, 10(1), pp. 1–21. doi: 10.1128/mBio.02504-18.

Westers, H. *et al.* (2002) 'Genome Engineering Reveals Large Dispensable Regions in Bacillus subtilis', pp. 2076–2090. doi: 10.1093/molbev/msg219.

Willkomm, D. K. and Hartmann, R. K. (2005) '6S RNA - An ancient regulator of bacterial RNA polymerase rediscovered', *Biological Chemistry*, 386(12), pp. 1273–1277. doi: 10.1515/BC.2005.144.

Wilson, T. J. G. *et al.* (1998) 'The rpfA gene of Xanthomonas campestris pathovar campestris, which is involved in the regulation of pathogenicity factor production, encodes an aconitase', *Molecular Microbiology*, 28(5), pp. 961–970. doi: 10.1046/j.1365-2958.1998.00852.x.

Wünsche, A. *et al.* (2012) 'CcpA forms complexes with CodY and RpoA in Bacillus subtilis', *FEBS Journal*, 279(12), pp. 2201–2214. doi: 10.1111/j.1742-4658.2012.08604.x.

Yakhnin, H. *et al.* (2007) ' CsrA of Bacillus subtilis regulates translation initiation of the gene encoding the flagellin protein (hag) by blocking ribosome binding ', *Molecular Microbiology*, 64(6), pp. 1605–1620. doi: 10.1111/j.1365-2958.2007.05765.x.

Yan, L. *et al.* (2003) 'Biofilm-specific cross-species induction of antimicrobial compounds in bacilli', *Applied and Environmental Microbiology*, 69(7), pp. 3719–3727. doi: 10.1128/AEM.69.7.3719-3727.2003.

Yan, S. and Wu, G. (2019) 'Proteases HtrA and HtrB for α-amylase secreted from Bacillus subtilis in secretion stress', *Cell Stress and Chaperones*. Cell Stress and Chaperones, 24(3), pp. 493–502. doi: 10.1007/s12192-019-

00985-1.

Yoshida, K. I. *et al.* (1997) 'Organization and transcription of the myo-inositol operon, iol, of Bacillus subtilis', *Journal of Bacteriology*, 179(14), pp. 4591–4598. doi: 10.1128/jb.179.14.4591-4598.1997.

Yoshida, K. I. *et al.* (1999) *Interaction of a repressor and its binding sites for regulation of the Bacillus subtilis iol divergon*, *Journal of Molecular Biology*. doi: 10.1006/jmbi.1998.2398.

Yoshida, K. I. *et al.* (2008) 'myo-inositol catabolism in Bacillus subtilis', *Journal of Biological Chemistry*, 283(16), pp. 10415–10424. doi: 10.1074/jbc.M708043200.

Zaher, H. S. and Green, R. (2010) 'Hyperaccurate and Error-Prone Ribosomes Exploit Distinct Mechanisms during tRNA Selection', *Molecular Cell*, 39(1), pp. 110–120. doi: 10.1016/j.molcel.2010.06.009.

Zakataeva, N. P. *et al.* (2012) 'Wild-type and feedback-resistant phosphoribosyl pyrophosphate synthetases from Bacillus amyloliquefaciens: Purification, characterization, and application to increase purine nucleoside production', *Applied Microbiology and Biotechnology*, 93(5), pp. 2023–2033. doi: 10.1007/s00253-011-3687-3.

Zhang, A. *et al.* (2003) 'Global analysis of small RNA and mRNA targets of Hfq', *Molecular Microbiology*, 50(4), pp. 1111–1124. doi: 10.1046/j.1365-2958.2003.03734.x.

Zhang, B. *et al.* (2016) 'Production of acetoin through simultaneous utilization of glucose, xylose, and arabinose by engineered Bacillus subtilis', *PLoS ONE*, 11(7), pp. 1–14. doi: 10.1371/journal.pone.0159298.

Zhang, Y. *et al.* (2013) 'Protein analysis by shotgun/bottom-up proteomics', *Chemical Reviews*, 113(4), pp. 2343–2394. doi: 10.1021/cr3003533.

Zhao, H. *et al.* (2017) 'Biological activity of lipopeptides from Bacillus', *Applied Microbiology and Biotechnology*, 101(15), pp. 5951–5960. doi: 10.1007/s00253-017-8396-0.

Zheng, G. *et al.* (1999) 'Genes of the sbo-alb locus of Bacillus subtilis are required for production of the antilisterial bacteriocin subtilosin', *Journal of*

Bacteriology, 181(23), pp. 7346–7355. doi: 10.1128/jb.181.23.7346-7355.1999.

Zhu, A., Ibrahim, J. G. and Love, M. I. (2019) 'Heavy-Tailed prior distributions for sequence count data: Removing the noise and preserving large differences', *Bioinformatics*, 35(12), pp. 2084–2092. doi: 10.1093/bioinformatics/bty895.

Zhu, B. and Stülke, J. (2018) 'SubtiWiki in 2018: From genes and proteins to functional network annotation of the model organism Bacillus subtilis', *Nucleic Acids Research*, 46(D1), pp. D743–D748. doi: 10.1093/nar/gkx908.

Zhu, Q. *et al.* (2019) 'Transcriptional regulation of virulence factors Spa and ClfB by the SpoVG-Rot cascade in Staphylococcus aureus', *International Journal of Medical Microbiology*. Elsevier, 309(1), pp. 39–53. doi: 10.1016/j.ijmm.2018.10.006.

Zuber, P., Healy, J. M. and Losick, R. (1987) 'Effects of plasmid propagation of a sporulation promoter on promoter utilization and sporulation in Bacillus subtilis', *Journal of Bacteriology*, 169(2), pp. 461–469. Available at: http://jb.asm.org/ (Accessed: 24 September 2019).

Zuber, P. and Losick, R. (1987) 'Role of AbrB in Spo0A- and Spo0B-dependent utilization of a sporulation promoter in Bacillus subtilis', *Journal of Bacteriology*, 169(5), pp. 2223–2230. doi: 10.1128/jb.169.5.2223-2230.1987.

Appendices

Table S.1: Common DPPs between M9 and SH5

Locus Tag	Gene	Product	Description	Function
BSU09260	*yhxA*	unknown	similar to adenosylmethionine-8-amino-7-oxononanoate aminotransferase	unknown
BSU17100	*pksC*	bacillaene synthase trans-acting acyltransferase	bacillaene synthase trans-acting acyltransferase	polyketide (bacillaene) synthesis
BSU17120	*pksE*	unknown	involved in polyketide synthesis	polyketide synthesis
BSU17170	*pksI*	unknown	involved in polyketide synthesis	polyketide synthesis
BSU18320	*ppsC*	plipastatin synthetase	plipastatin synthetase	production of the antibacterial compound plipastatin
BSU18550	*yoaC*	unknown	similar to xylulokinase	unknown
BSU22970	*ypbH*	putative adaptor protein	putative adaptor protein for ClpC-ClpP	unknown
BSU27930	*spo0B*	sporulation initiation phosphotransferase	Sporulation initiation phosphotransferase of the phosphorelay	initiation of sporulation
BSU31250	*tlpA*	methyl-accepting chemotaxis protein	membrane-bound chemotaxis receptor, methyl-accepting chemotaxis protein	control of chemotaxis
BSU31830	*yueE*	unknown	unknown	unknown
BSU32859	*yuzM*	unknown	unknown	unknown
BSU33750	*sdpA*	unknown	required for SdpC toxin maturation	maturation of the SdpC toxin
BSU33770	*sdpC*	toxin, kills non-sporulating cells	toxin, collapses the proton motive force and induces autolysis, kills non-sporulating cells, induces activity of SigW	killing of non-sporulating sister cells

Table S. 2: Common DPPs between M9 and SH2

Locus Tag	Gene	Product	Description	Function
BSU08140	*yfjD*	unknown	unknown	unknown
BSU10430	*yhxD*	unknown	general stress protein, similar to alcohol dehydrogenase,	survival of salt and ethanol stresses
BSU12900	*htrA*	serine protease Do	membrane-anchored protein quality control protease, serine protease Do	protein quality control
BSU18530	*yoaA*	unknown	similar to spermidine/spermine N-acetyltransferase	unknown
BSU19740	*yodT*	unknown	similar to adenosylmethionine-8-amino-7-oxononanoate aminotransferase	unknown
BSU27440	*glnH*	glutamine ABC transporter (binding protein)	glutamine ABC transporter (binding protein)	glutamine uptake
BSU29930	*amyX*	pullulanase	pullulanase (debranching enzyme)	starch degradation
BSU30940	*glgP*	glycogen phosphorylase	glycogen phosphorylase	glycogen biosynthesis

Table S. 3: Common DPPs between LB and M9

Locus Tag	Gene	Product	Description	Function
BSU20390	*yorG*	unknown	unknown	unknown
BSU00860	*clpC*	AAA unfoldase, ATPase subunit of the ClpC-ClpP protease	AAA unfoldase, ATPase subunit of the ClpC-ClpP protease, directs proteins phosphorylated on arginine residues to ClpP	protein degradation, positive regulator of autolysin (LytC and LytD) synthesis
BSU24870	*yqgP*	intramembrane protease	intramembrane protease	cell division or sugar metabolism

BSU18340	*ppsA*	plipastatin synthetase	plipastatin synthetase	production of the antibacterial compound plipastatin
BSU27260	*mccA*	O-acetylserine-thiol-lyase	O-acetylserine-thiol-lyase	methionine-to-cysteine conversion
BSU32770	*yusE*	unknown	similar to thioredoxin	unknown

Table S. 4: Common DPPs between LB and SH2

Locus Tag	Gene	Product	Description	Function
BSU15400	*ylmG*	unknown	unknown	unknown
BSU00550	*mfd*	transcription-repair coupling factor	transcription-repair coupling factor, eliminates genetic damage from transcriptionally active genes during sporulation, required for increased mutagenesis of lagging strand genes	promotes strand-specific DNA repair by displacing
BSU21990	*ypdQ*	unknown	similar to RNase HI	unknown
BSU04100	*kipR*	transcriptional regulator	transcriptional repressor of the kip operon	regulation of sporulation initiation
BSU28280	*leuA*	2-isopropylmalate synthase	2-isopropylmalate synthase	biosynthesis of leucine
BSU29570	*sspA*	small acid-soluble spore protein (major α-type SASP)	small acid-soluble spore protein (major α-type SASP)	protection of spore DNA

List of different media used in this study

Lysogeny Broth (LB)

For 1 L of media add following to 800 mL H_2O

1. 10 g Bacto-tryptone
2. 5 g yeast extract
3. 10 g NaCl
4. 15 g agar (for LB agar)

Adjust pH to 7.5 with 10 M NaOH, adjust volume to 1 L with dH_2O and sterilize by autoclaving. Store at 4˚C.

Casein Hydrolysate (CH) media

1. Casein hydrolysate (Oxoid L41) : 10.0 g
2. L-glutamic acid : 3.68 g
3. L-alanine : 1.25 g
4. L-asparagine : 1.39 g
5. KH_2PO_4 : 1.36 g
6. NH_4Cl : 1.34 g
7. Na_2SO_4 : 0.11 g
8. Nh_4NO_3 : 0.10 g
9. $FeCl_3.6H_2O$: 1 **mg**

- ✓ Add 800 mL H_2O, adjust the pH to 7.0 with 10 N NaOH, make up to 940 mL H_2O, aliquot 10 x 94 mL, and autoclave
- ✓ **NB:** Do **NOT** use casamino acids

On the day of culture, add following to 94 mL pre-warmed CH medium:

1. 1M $MgSO_4$ (autoclaved) : 0.04 mL
2. 100 mM $CaCl_2$ (autoclaved) : 1.00 mL
3. 50 mM $MnSO_4$ (autoclaved) : 0.20 mL
4. 2 mg/mL L-Tryptophan (filtered) : 1.00 mL

- ✓ Make up to 100 mL
- ✓ **NB:** Trp absent from casein hydrolysate

Resuspension Media (RM)

1. Sporulation Salts : 90 mL
2. 5% L-Glutamic acid, pH 7 : 4 mL (NB: adjust pH with 10 N NaOH to dissolve)
3. 100 mM $CaCl_2$: 1 mL
4. 1 M $MgSO_4$: 4 mL
5. 2 mg/mL L- Tryptophan (filtered) : 1 mL

- ✓ Make up to 100 mL if necessary

Sporulation salts:

To make 1L of sporulation salts:

1. Solution A : 1 mL
2. Solution B : 10 mL

- ✓ Make up to 1L with sterile H_2O

Solution A (100 mL):

1. $FeCl_3.6H_2O$: 0.089 g
2. $MgCl_2.6H_2O$: 0.830 g
3. $MnCl_2.4H_2O$: 1.979 g

✓ Make up to 100 mL, sterilise by autoclaving

Solution B (1L):

NH_4Cl : 53.5 g

Na_2SO_4 : 10.6 g

KH_2PO_4 : 6.8 g

NH_4NO_3 : 9.7 g

✓ Add 800mL H_2O; adjust pH to 7.0 with 1 N NaOH; make up to 1 L; filter sterilise; store at 4˚C.

Paris Media

Stocks solutions

Solution Name	Compounds	Concentration	Amount	Sterilization
PC 10X	K_2HPO_4 KH_2PO_4 Trisodium citrate	107 g/L 60 g/L 10 g/L	2X 250 mL	Steam sterilize (100˚C for 30 min)
Glucose	N/A	20%	100 mL	0.45 µ filter
K-Glutamate.H_2O	N/A	1 M	50 mL	0.20 µ filter
Fe-ammonium citrate (FAC)	N/A	0.11 g/100 mL	50 mL	0.20 µ filter
Casamino acids	N/A	2 %	100 mL	Steam
Tryptophan	N/A	2 mg/mL	50 mL	Steam
$MgSO_4$	N/A	1 M	50 mL	Autoclave

Working solutions

Stock Solutions	Diluted for 100 mL	Diluted for 50 mL	Diluted for 10 mL
PC 10X	10 mL	5 mL	1 mL
Glucose	5 mL	2.5 mL	500 µL
K-Glutamate.H_2O	2 mL	1 mL	200 µL
Fe-ammonium citrate (FAC)	200 µL	100 µL	20 µL
Casamino acids	5 mL	2.5 mL	500 µL
Tryptophan	2 mL	1 mL	200 µL
$MgSO_4$	300 µL	150 µL	30 µL