A Thesis Submitted for the Degree of PhD at the University of Warwick

Permanent WRAP URL:

http://wrap.warwick.ac.uk/154547

Copyright and reuse:
This thesis is made available online and is protected by original copyright. Please scroll down to view the document itself. Please refer to the repository record for this item for information to help you to cite it. Our policy information is available from the repository home page.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk
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

**Table of Contents** ........................................................................................................... i

**List of Tables** ................................................................................................................... v

**List of Figures** ................................................................................................................... vii

**Acknowledgements** .......................................................................................................... xi

**Declaration** ........................................................................................................................ xiii

**Abstract** .............................................................................................................................. xiv

**Abbreviations** .................................................................................................................... xv

**Chapter 1 Introduction** ..................................................................................................... 1

1.1 Bacterial Post-Transcriptional Gene Regulation ............................................................... 2

1.1.1 Regulatory RNAs in Bacteria ..................................................................................... 3

1.1.2 RNA-Binding Proteins in Bacteria ............................................................................. 6

1.2 *Bacillus subtilis* .............................................................................................................. 12

1.2.1 *B. subtilis* as a Model Species ................................................................................ 12

1.2.2 Gene Regulation in *B. subtilis* .............................................................................. 14

1.2.3 Sporulation in *B. subtilis* ....................................................................................... 16

1.2.4 RNA-Binding Proteins in *B. subtilis* ..................................................................... 19

1.3 SpoVG in *B. subtilis* and Beyond .................................................................................. 20

1.4 Summary and Thesis Structure ....................................................................................... 23

**Chapter 2 Materials and Methods** ............................................................................... 26

2.1 Bacterial Cell Biology .................................................................................................... 26

2.1.1 Growth Media ........................................................................................................... 26

2.1.2 Bacterial Strains ..................................................................................................... 27

2.1.3 Bacterial Transformations ....................................................................................... 33

2.2 Molecular Biology ....................................................................................................... 35

2.2.1 Genomic DNA Extractions .................................................................................... 35

2.2.2 Plasmid Extractions ............................................................................................... 36

2.2.3 Cloning .................................................................................................................... 36

2.2.4 Preparation of Protein Lysate ............................................................................... 41

2.2.5 SDS-PAGE and Western Blotting ......................................................................... 42

2.3 Omics Experiments ....................................................................................................... 43
2.3.1 Transcriptomics with RNAtag-Sequencing .............................................. 43
2.3.2 Total Proteomic Analysis ............................................................................. 47

2.4 RNA-Protein Interactions in \textit{B. subtilis} ...................................................... 51
2.4.1 Ultraviolet (UV) Cross-linking and Immunoprecipitation (CLIP) ............ 51
2.4.2 UV Cross-linking and Analysis of cDNA (CRAC) .................................... 51

2.5 RNA-RNA Interactions by CLASH ................................................................. 52

2.6 Protein-Protein Crosslinking and Immunoprecipitation (PPI) ................. 52
2.6.1 Cell Culturing ............................................................................................... 52
2.6.2 Crosslinking with Formaldehyde ............................................................... 52
2.6.3 Cell Lysis and Immunoprecipitation ......................................................... 53
2.6.4 Protein Gels and Western Blotting ............................................................. 54
2.6.5 Mass Spectrometry ....................................................................................... 55
2.6.6 Data Analysis ................................................................................................ 56

2.7 Phenotypic Assays ....................................................................................... 56
2.7.1 Swimming and Swarming in \textit{B. subtilis} .................................................... 56
2.7.2 Biofilms Formation on LBGM agar ......................................................... 57
2.7.3 Biofilms Formation on MSgg agar ............................................................ 57
2.7.4 TCA Precipitation ....................................................................................... 57

\textbf{Chapter 3 Transcriptomic Analysis of \(\Delta\text{spoVG}\)} ..................................... 59

3.1 Introduction ...................................................................................................... 59

3.2 Results and Discussion .................................................................................. 60
3.2.1 Growth of \(\Delta\text{spoVG}\) in M9 ................................................................. 60
3.2.2 Cell Culture and Sample Preparation for RNAtag-Seq ............................. 61
3.2.3 RNAtag-Seq Pre-Data Processing and Quality Check ............................. 61
3.2.4 Identification of Putative ncRNAs ............................................................. 69
3.2.5 Analysis of the Differentially Expressed Genes ....................................... 70
3.2.6 Gene Set Enrichment Analysis ................................................................. 85
3.2.7 KEGG Pathway Analysis of the Differentially Expressed Genes .......... 95
3.2.8 Affected Sigma Factor Regulons due to \(\Delta\text{spoVG}\) ................................. 98

3.3 Conclusion ...................................................................................................... 101

\textbf{Chapter 4 Proteomic Analysis of \(\Delta\text{spoVG}\)} ................................................. 104

4.1 Introduction ...................................................................................................... 104

4.2 Results and Discussion .................................................................................. 105
4.2.1 Growth of \(\Delta\text{spoVG}\) in LB .................................................................. 105
4.2.2 Cell Culture and Sample Preparation for LC-MS/MS .................................................. 105
4.2.3 LC-MS/MS Data Pre-Processing and Quality Analysis ........................................... 106
4.2.4 Analysis of the Differentially Produced Proteins (DPPs) .......................................... 115
4.2.5 Gene Set Enrichment Analysis ................................................................................... 130
4.2.6 KEGG Pathway Enrichment Analysis of DPPs ......................................................... 150
4.2.7 Affected Regulons in ∆spoVG ................................................................................... 154

4.3 Conclusion .................................................................................................................... 157

Chapter 5 RNA-Binding Profile of SpoVG in B. subtilis ................. 160

5.1 Introduction ................................................................................................................... 160

5.2 Results and Discussion ............................................................................................... 162
5.2.1 Construction of SpoVG-FLAG Strain ....................................................................... 162
5.2.2 RNA-Binding Assay of SpoVG-FLAG with CLIP ..................................................... 164
5.2.3 Construction of SpoVG-HTF Strain ......................................................................... 165
5.2.4 RNA-Binding Assay of SpoVG-HTF ....................................................................... 168
5.2.5 Analysis of RNAs Co-Immunoprecipitated with SpoVG ........................................ 170

5.3 Conclusions .................................................................................................................. 178

Chapter 6 Protein-Protein Interaction with SpoVG ......................... 181

6.1 Introduction ................................................................................................................... 181

6.2 Results and Discussion ............................................................................................... 183
6.2.1 Cell Culturing and Sample Handling ....................................................................... 183
6.2.2 Confirmation of Protein Cross-linking and Immunoprecipitation ............................. 184
6.2.3 Identification of Proteins Cross-linked with SpoVG via MS/MS ............................... 186
6.2.4 Characterisation of Proteins Co-Immunoprecipitated with SpoVG ......................... 189

6.3 Conclusion .................................................................................................................... 191

Chapter 7 Functional Characterisation of ∆spoVG ......................... 194

7.1 Introduction ................................................................................................................... 194
7.1.1 Motility in B. subtilis ............................................................................................... 195
7.1.2 Biofilm Formation in B. subtilis ............................................................................... 196
7.1.3 Secretion in B. subtilis ............................................................................................. 197
7.1.4 Growth of B. subtilis in Different Carbon Sources .................................................. 198

7.2 Results and Discussion ............................................................................................... 199
7.2.1 Motility and Biofilm in ∆spoVG ............................................................................... 199
7.2.2 Secretion in ∆spoVG ............................................................................................... 208
7.2.3 Growth of ∆spoVG in Alternate Carbon Sources .................................................... 212
7.3 Conclusion..................................................................................................................214

Chapter 8 Discussion ......................................................................................................217

8.1 Investigating the Global Transcriptome of ΔspoVG ..................................................220
8.2 Exploring the Global Proteome of ΔspoVG .................................................................222
8.3 RNA-Binding Activity of SpoVG ..............................................................................224
8.4 Protein-Protein Interaction with SpoVG .................................................................227
8.5 Phenotypic Analysis of ΔspoVG ..............................................................................230
8.6 Limitations and Future Work .....................................................................................234
8.7 Summary ..................................................................................................................235

Bibliography ..................................................................................................................238

Appendices .....................................................................................................................279

List of different media used in this study ....................................................................... 282

Lysogeny Broth (LB) ........................................................................................................ 282
Casein Hydrolysate (CH) media .................................................................................... 282
Resuspension Media (RM) ............................................................................................. 283
Paris Media ..................................................................................................................... 285
List of Tables

Table 1.1: Overview of sigma factors in B. subtilis .................................................. 15
Table 2.1: B. subtilis strains used in this study ......................................................... 28
Table 2.2: Plasmids used in this study ................................................................. 38
Table 2.3: Primers used in this study ....................................................................... 39
Table 3.1: Depth and coverage outputs from RNAtag-Seq ........................................ 63
Table 3.2: Summary of differentially expressed putative non-coding RNAs (ncRNAs) identified in different growth conditions ......................................... 69
Table 3.3: Significantly DEGs in ∆spoVG in M9 ...................................................... 74
Table 3.4: Top 10 significantly DEGs in ∆spoVG in SH2 .......................................... 77
Table 3.5: Top 10 significantly DEGs in ∆spoVG in SH5 .......................................... 79
Table 3.6: Significantly differentially expressed genes (DEGs) in ∆spoVG vs WT found commonly in M9, SH2 and SH5 .......................................................... 82
Table 3.7: Significantly enriched gene sets identified in ∆spoVG from the transcriptomic data ........................................................................................................... 87
Table 3.8: Summary of KEGG pathways enriched in ∆spoVG in B. subtilis 96
Table 4.1: DPPs with logFC > 2 and logFC < -2 in ∆spoVG ...................................... 115
Table 4.2: Significant DPPs in LB in ∆spoVG ............................................................. 119
Table 4.3: Significant DPPs in M9 in ∆spoVG ........................................................... 121
Table 4.4: Significant DPPs in SH2 in ∆spoVG ......................................................... 124
Table 4.5: Significant DPPs in SH5 in ∆spoVG ......................................................... 127
Table 4.6: Common DPPs in all the growth conditions examined in this study ........ 131
Table 4.7: Commonly changed DPPs with opposite log fold change (logFC) values in SH2 vs SH5 ............................................................ 132
Table 4.8: Significantly enriched Category 3 (SubtiWiki) gene sets identified in ∆spoVG from the proteomic data ................................................................. 135
Table 4.9: Significantly enriched antimicrobial proteins in ΔspoVG in M9 .................................................. 138

Table 4.10: Proteomic profile of decreased ribosomal proteins in ΔspoVG in SH2 .................................................................................................................. 144

Table 4.11: Summary of the KEGG pathways enriched in due to the DPPs in ΔspoVG in all the growth conditions ............................................................................................................ 152

Table 4.12: Nonribosomal proteins affected in ΔspoVG in LB and M9 ..... 153

Table 5.1: Top 20 co-immunoprecipitated (co-IPed) RNAs with SpoVG-HTF in this study ................................................................................................................................. 171

Table 6.1: Selection of proteins co-IPed with SpoVG ............................................. 188
List of Figures

Figure 1.1: Schematic of the different types of regulatory RNAs in bacteria 5

Figure 1.2: Overview of several properties of B. subtilis ........................................... 13

Figure 1.3: Stages of sporulation in B. subtilis ................................................................. 18

Figure 1.4: Overview of roles and function of SpoVG in B. subtilis and other bacteria ....................................................................................................................................................... 22

Figure 1.5: Level of transcription of spoVG across some conditions compared to........................................................................................................................................................................... 24

Figure 1.6: Approach taken in this thesis to study global omics of ΔspoVG in B. subtilis ........................................................................................................................................................................................................ 25

Figure 2.1: Schematic of the deletion system used to generate the ΔspoVG mutant .................................................................................................................................................................................................. 31

Figure 2.2: Schematic of construction of the spoVG complement strain (BSB1 ΔspoVG::spoVG) .................................................................................................................................................. 32

Figure 3.1: ΔspoVG shows similar growth to the WT in M9 supplemented with glucose .................................................................................................................................................................................................. 62

Figure 3.2: Schematic of the steps taken in the differential gene expression analysis of ΔspoVG ........................................................................................................................................................................... 63

Figure 3.3: Deletion of spoVG results in differential expression of genes .......... 65

Figure 3.4: The biological replicates of the WT and ΔspoVG cluster differently based on the type of growth media used .......................................................................................................................................................................................... 66

Figure 3.5: The replicates of WT and ΔspoVG transcripts cluster separately based on their genotypes in each growth media .............................................................................................................................................................................. 68

Figure 3.6: Deletion of spoVG produced statistically significantly DEGs in each growth media ......................................................................................................................................................................................................................... 70

Figure 3.7: Volcano plots of show the up-regulated and down-regulated genes in ΔspoVG in each of the (A) M9, (B) SH2 and (C) SH5 ....................... 72
Figure 3.8: The transcriptomic analysis of the ∆spoVG strain identified commonly altered significantly DEGs across all the growth conditions. ................................................................. 81

Figure 3.9: GSEA of the ∆spoVG transcriptomic data identifies significantly enriched gene sets in each growth condition .......................... 86

Figure 3.10: Commonly identified gene sets in ∆spoVG under different growth conditions ........................................................................ 88

Figure 3.11: Significantly enriched KEGG pathways identified in the transcriptomic analysis of ∆spoVG in B. subtilis .......................... 97

Figure 3.12: Sigma factor regulons affected in ∆spoVG in the transcriptomic analysis ........................................................................ 99

Figure 4.1: ∆spoVG shows identical growth to the WT in LB broth ......... 106

Figure 4.2: Schematic of the proteomic analysis of the DPPs in ∆spoVG. 107

Figure 4.3: Normal distribution of the log2 transformed protein LFQ intensities identified in LC-MS/MS experiment ....................... 108

Figure 4.4: Pearson correlation analysis of the proteomes from the WT and ∆spoVG samples compared condition wise showed a good correlation amongst the samples ......................................................... 111

Figure 4.5: The biological replicates of the WT and ∆spoVG cluster differently based on the growth media used ........................................... 113

Figure 4.6: The replicates of WT and ∆spoVG proteome cluster separately based on their genotypes......................................................... 114

Figure 4.7: Distribution of the DPPs in all the growth conditions ............ 116

Figure 4.8: Overview of the DPPs (padj<0.05) in ∆spoVG in each condition investigated ................................................................. 117

Figure 4.9: The ∆spoVG common DPPs across all the growth conditions 129

Figure 4.10: GSEA of the significantly DPPs in ∆spoVG ..................... 133

Figure 4.11: Biological processes affected in SH2 in ∆spoVG ............... 143

Figure 4.12: Biological pathways affected in SH5 .............................. 147

Figure 4.13: The KEGG enriched pathways due to the DPPs in ∆spoVG .. 151
Figure 4.14: Sigma factors regulons affected in ΔspoVG ............................. 156

Figure 5.1: Western blot confirms the production of SpoVG-FLAG in B. subtilis ................................................................. 163

Figure 5.2: CLIP identifies SpoVG as an RBP in B. subtilis in vivo .............. 165

Figure 5.3: Western blot confirms the production of SpoVG-HTF in B. subtilis in LB ............................................................. 167

Figure 5.4: SpoVG-HTF binds to RNA transcripts (Courtesy: Dr. Sander Granneman – University of Edinburgh) .............................. 169

Figure 5.5: Functional characterisation of the transcripts bound to SpoVG .................................................................................. 176

Figure 6.1: Schematic of the major steps in the PPI study with SpoVG-HTF and identification of the co-IPed proteins. ......................... 184

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. ................................................................. 185

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 ................................................................ 186

Figure 6.4: Significantly co-IPed protein with SpoVG belong to several KEGG pathways ................................................................ 190

Figure 6.5: Experimentally identified protein-protein association network of SpoVG ........................................................................ 191

Figure 7.1: The ΔspoVG (3610) cells have reduced swarming motility and form distinctive “channel” like morphology on M9 agar .......... 199

Figure 7.2: The ΔspoVG (3610) swarming cells have restricted growth and do not produce pink pigment on M9 agar plates after 24 hours .... 201

Figure 7.3: (Picture courtesy: Nicola Stanley-Wall Lab) The biofilms of the ΔspoVG (3610) are smaller and have more “wrinkles” on MSgg agar 202

Figure 7.4: The swimming phenotype of ΔspoVG (168) is reduced on M9 agar ............................................................................. 203
Figure 7.5: The biofilms of ΔspoVG (3610) have an expanded morphology and produce less brown pigment on LBGM agar ................................. 207

Figure 7.6: Secretion of α-amylase in ΔspoVG (168 and 3610) is decreased in LBA-starch............................................................................. 210

Figure 7.7: Secretion of total proteins in ΔspoVG (3610) is reduced in LB 211

Figure 7.8: Growth of ΔspoVG (168) is affected in Paris media in response to various carbon sources......................................................... 213

Figure 8.1: Comparison of transcriptomic and proteomic changes in the ΔspoVG mutant in M9, SH2 and SH5........................................... 232

Figure 8.2: Summary of phenotypes identified with ΔspoVG in this thesis .......................................................................................... 233

Figure 8.3: Biological processes commonly found across chapter 3, 4, 5, and 6 .................................................................................. 236

Figure 8.4: Identification of pathways common in significantly co-IPed RNAs and proteins with SpoVG ............................................. 237
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 Tauqueer 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 Scholars 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 Ramakison 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. Krishnamegh 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 ΔspoVG 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, ΔspoVG was functionally characterised based on the -omics data from this thesis and novel phenotypes of ΔspoVG 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 MnSO₄)
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
MS²  MS/MS spectra
ncRNA  Non-coding RNA
nm  Nanometre
NRP  Non-ribosomal peptide
nt  Nucleotide
O/N  Overnight
OD₆₀₀  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
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>s</td>
<td>Seconds</td>
</tr>
<tr>
<td>SDS-PAGE</td>
<td>Sodium dodecyl sulphate–polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>SH2</td>
<td>Sporulation hour 2</td>
</tr>
<tr>
<td>SH5</td>
<td>Sporulation hour 5</td>
</tr>
<tr>
<td>sRNA</td>
<td>Small non-coding RNA</td>
</tr>
<tr>
<td>TCA</td>
<td>Trichloroacetic acid</td>
</tr>
<tr>
<td>TCA cycle</td>
<td>Tricarboxylic acid cycle</td>
</tr>
<tr>
<td>TF</td>
<td>Transcription factor</td>
</tr>
<tr>
<td>UMP</td>
<td>Uridine monophosphate</td>
</tr>
<tr>
<td>UTR</td>
<td>Untranslated region</td>
</tr>
<tr>
<td>UV</td>
<td>Ultraviolet</td>
</tr>
<tr>
<td>V</td>
<td>Volt</td>
</tr>
<tr>
<td>WB</td>
<td>Western blot</td>
</tr>
<tr>
<td>WT</td>
<td>Wild-type</td>
</tr>
<tr>
<td>µ</td>
<td>Micro</td>
</tr>
</tbody>
</table>
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 ΔspoVG 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 ∆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](image)

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 $\lambda = 254$ 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 carbohydrates 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*](image-url)
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-Rodriguez, 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*

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

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 \textit{B. subtilis} (Ul Haq, Müller and Brantl, 2020). The lack of a global regulator of post-transcriptional gene regulation in \textit{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 \textit{L. monocytogenes}, \textit{S. aureus} and \textit{B. burgdorferi}. Therefore, in this thesis I present a multi-omics approach in characterising SpoVG for its RNA-binding activity and regulatory effects in \textit{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 ∆spoVG in \textit{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 ∆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 \textit{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 omics of ΔspoVG in *B. subtilis*

Next, Chapter 7 functionally characterises the ΔspoVG 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

<table>
<thead>
<tr>
<th>Strains</th>
<th>Description</th>
<th>Resistance</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>BSB1 (168, trp+)</td>
<td>Laboratory strain wild-type <em>B. subtilis</em></td>
<td>None</td>
<td>(Nicolas et al., 2012)</td>
</tr>
<tr>
<td>BSB1 spoVG::phl (ΔspoVG)</td>
<td>Chromosomal deletion of spoVG by a reciprocal recombination and replacement with the phl gene (encoding for phleomycin)</td>
<td>Phleomycin</td>
<td>Emma L Denham</td>
</tr>
<tr>
<td>BSB1 spoVG::phl, amyE::spoVG (ΔspoVG::spoVG)</td>
<td>Chromosomal complement of spoVG at amyE locus</td>
<td>Tetracycline</td>
<td>This study</td>
</tr>
<tr>
<td>168, trp+ SpoVG-Flag</td>
<td>Chromosomal integration of SpoVG-FLAG at amyE locus in the ΔspoVG strain</td>
<td>Tetracycline</td>
<td>This study</td>
</tr>
<tr>
<td>168, trp+ SpoVG-HTF</td>
<td>Chromosomal integration of SpoVG-HTF at amyE locus in the ΔspoVG strain</td>
<td>Tetracycline</td>
<td>This study</td>
</tr>
<tr>
<td>NCIB 3610</td>
<td>Undomesticated strain wild-type <em>B. subtilis</em></td>
<td>None</td>
<td>BGSC</td>
</tr>
<tr>
<td>NCIB 3610 ΔspoVG</td>
<td>Chromosomal deletion of ΔspoVG</td>
<td>Phleomycin</td>
<td>This study</td>
</tr>
<tr>
<td>NCIB 3610 ΔspoVG::spoVG</td>
<td>Chromosomal complement of spoVG at amyE locus</td>
<td>Tetracycline</td>
<td>This study</td>
</tr>
</tbody>
</table>
2.1.2.1 Construction of $\Delta$spoVG

The $\Delta$spoVG knockout deletion strain was constructed according to the Standard Operating Procedure (SOP) BaSynthetic (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 $\Delta$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 $\Delta$spoVG::spoVG

For constructing the complement strain of spoVG ($\Delta$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 $\Delta$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.
Chapter 2 | Materials and Methods

(A) Selection of deletion

1) PCR
2) Joining PCR
3) Transformation
4) ΔspoVG mutant

(B) Direct repeats

Primers

Phleo3

Phleo5

Chromosome

PhleoR

cl

P_{upp}

P_{sacB}

P_{sak}

PgeoR

upp

spoVG

NeoR

PhleoR

NeoS

ΔspoVG mutant

Mahapatra, Swati | PhD Thesis | 2020
Figure 2.1: Schematic of the deletion system used to generate the $\Delta spoVG$ 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 $\Delta spoVG$ 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 $\Delta spoVG$ deletion mutant.
Figure 2.2: Schematic of construction of the spoVG complement strain (BSB1 ΔspoVG::spoVG)

The schematic shows the mechanism of complementation of spoVG into the ΔspoVG 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 (TetR); and 5’ and 3’ homologous sequence of amyE gene. (B) Represents the homologous recombination between the plasmid spoVG_pRMC and the amyE locus in ΔspoVG mutant. (C) Illustrates the construct of the complemented BSB1 ΔspoVG::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 MgCl₂, 10 mM MgSO₄, 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 MgCl₂ and 10 mM MgSO₄.7H₂O) 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 OD₆₀₀ 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 OD₆₀₀ 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 inverted 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 NgoMIV (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 Ascl 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 Ascl 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

<table>
<thead>
<tr>
<th>Plasmids</th>
<th>Description</th>
<th>Resistance</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>pFLAG (ECE263)</td>
<td>Cloning vector with tagging genes with 3X-FLAG. Does not integrate into B. subtilis</td>
<td>Chloramphenicol in <em>E. coli</em> and Tetracycline in <em>B. subtilis</em></td>
<td>BGSC</td>
</tr>
<tr>
<td>pRMC</td>
<td>Cloning vector for the complementation of deletion mutants. Integrates into the <em>amyE</em> locus of <em>B. subtilis</em></td>
<td>Ampicillin in <em>E. coli</em> and Tetracycline in <em>B. subtilis</em></td>
<td>Ruben AT Mars (Mars et al., 2015)</td>
</tr>
<tr>
<td>pRMC-spoVG-FLAG</td>
<td>pRMC containing <em>spoVG-FLAG</em></td>
<td>Tetracycline</td>
<td>This study</td>
</tr>
<tr>
<td>pRMC-spoVG-HTF</td>
<td>pRMC containing <em>spoVG-HTF</em></td>
<td>Tetracycline</td>
<td>This study</td>
</tr>
</tbody>
</table>
Table 2.3: Primers used in this study

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

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Primer description</th>
<th>Primer sequence (5' → 3')</th>
<th>Type</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>spoVG_FLAG_FP</td>
<td>3X FLAG plasmid to spoVG</td>
<td>ataaggaggaactatggtggATTTACCTTATGCCCGAAAATGAAG</td>
<td>FP</td>
<td>This study</td>
</tr>
<tr>
<td>spoVG_FLAG_RP</td>
<td>spoVG to 3X FLAG plasmid</td>
<td>caccatcatgatctttataatcgGGAGAAGCTCCAGCTCTTC</td>
<td>RP</td>
<td>This study</td>
</tr>
<tr>
<td>spoVG_pRMC_FP</td>
<td>pRMC to spoVG</td>
<td>tgcttacagggattccggATTTACCTTATCCCGAGAATG</td>
<td>FP</td>
<td>This study</td>
</tr>
<tr>
<td>spoVG_pRMC_RP</td>
<td>3X FLAG to pRMC</td>
<td>atatcactagattgggctggCGACCGAGCGCAGCGAGT</td>
<td>RP</td>
<td>This study</td>
</tr>
<tr>
<td>Bac27F</td>
<td>Amplifies 16S rRNA</td>
<td>AGAGTTTGGATCMTGGCTCAG</td>
<td>FP</td>
<td>Illumina</td>
</tr>
<tr>
<td>Univ1492R</td>
<td>Amplifies 16S rRNA</td>
<td>CGGTTACCTTGGTACGCTT</td>
<td>RP</td>
<td>Illumina</td>
</tr>
<tr>
<td>AR2</td>
<td>Barcoded adapter bind to the 5' of RNA</td>
<td>TACACGACGCTCTTCCGAT</td>
<td>FP</td>
<td>Illumina</td>
</tr>
<tr>
<td>3Tr3</td>
<td>Barcoded adapter bind to the 3' of RNA</td>
<td>AGATCGGAAGAGCACACGTCTG</td>
<td>RP</td>
<td>Illumina</td>
</tr>
<tr>
<td>P0_spoVG</td>
<td>Deletion of spoVG</td>
<td>gtgtggact tgtgactctc</td>
<td>FP</td>
<td>Emma L Denham</td>
</tr>
<tr>
<td>P1_spoVG</td>
<td>Deletion of spoVG</td>
<td>acgctcttgattaataccgta</td>
<td>FP</td>
<td>Emma L Denham</td>
</tr>
<tr>
<td>P2_spoVG</td>
<td>Deletion of spoVG</td>
<td>CGACCTGCAGCCATGCAAGCTctagttcaccaccttttc</td>
<td>RP</td>
<td>Emma L Denham</td>
</tr>
<tr>
<td>Primer/Expression</td>
<td>Description</td>
<td>Sequence</td>
<td>Tag</td>
<td>Lab</td>
</tr>
<tr>
<td>-------------------</td>
<td>-------------</td>
<td>----------</td>
<td>-----</td>
<td>-------------</td>
</tr>
<tr>
<td>P3_spoVG</td>
<td>Deletion of spoVG</td>
<td>CGAGCTCGAATTCACTGCGTCGcaagactgctgaaaggg</td>
<td>FP</td>
<td>Emma L Denham</td>
</tr>
<tr>
<td>P4_spoVG</td>
<td>Deletion of spoVG</td>
<td>catctcgcgaacagttgaac</td>
<td>RP</td>
<td>Emma L Denham</td>
</tr>
<tr>
<td>phleo3</td>
<td>Phleomycin deletion cassette</td>
<td>AGCTTGCATGCCTGCGAGGTGC</td>
<td>FP</td>
<td>Noirot lab</td>
</tr>
<tr>
<td>phleo5</td>
<td>Phleomycin deletion cassette</td>
<td>CGACGGCCAGTGAATTCGAGCTGC</td>
<td>RP</td>
<td>Noirot lab</td>
</tr>
<tr>
<td>spoVG_pRMC_comp_FP</td>
<td>Inserts spoVG into pRMC</td>
<td>atacactagattggcttgAAAGTGATTCTGAGAG</td>
<td>FP</td>
<td>This study</td>
</tr>
<tr>
<td>spoVG_pRMC_comp_RP</td>
<td>Inserts spoVG into pRMC</td>
<td>tgcgtacaggtcttgAAAGTTTGAATACAGAATTAAAAAG</td>
<td>RP</td>
<td>This study</td>
</tr>
<tr>
<td>pRMC_seq_F</td>
<td>pRMC sequencing primer</td>
<td>GACCGGCGTCAGGATCT</td>
<td>FP</td>
<td>Denham lab</td>
</tr>
<tr>
<td>pRMC_seq_R</td>
<td>pRMC sequencing primer</td>
<td>GTCGACGTCATATGGAATCC</td>
<td>RP</td>
<td>Denham lab</td>
</tr>
</tbody>
</table>
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 Ascl. 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 LB-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 OD₆₀₀ 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 transblotter (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 ΔspoVG 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 MgCl₂, 20 mM NaN₃). For M9, cells were harvested at OD₆₀₀ 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 Na₂EDTA, 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:choleform (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 ΔspoVG 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 ΔspoVG 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 ΔspoVG 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 ΔspoVG 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 imputed 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 imputed 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 ΔspoVG 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 (~OD600 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°C. 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<sub>600</sub> 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<sub>4</sub>) (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<sub>600</sub> 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 ΔspoVG

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* (Pigott and Coote,
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 ΔspoIIBΔspoVG 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 ΔspoIIBΔspoVG 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 ΔspoVG 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 ΔspoVG 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 ΔspoVG 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 (ΔspoVG::spoVG) at the amyE locus (construction described in Chapter 2, section 2.1.2.2) was also included. The ΔspoVG strain showed a very similar growth pattern to the WT (Figure 3.1). A slight increase in growth was present in the ΔspoVG and ΔspoVG::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 ΔspoVG 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: ΔspoVG shows similar growth to the WT in M9 supplemented with glucose

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

3.2.3.1 Data quality analysis

For the differential gene expression analyses, the read counts from ΔspoVG 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 log2 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 - ΔspoVG. Biological replicate numbers of sample are shown from 1-3.

<table>
<thead>
<tr>
<th>Condition_strain_replicate</th>
<th>Percentage of total reads aligned to <em>B. subtilis</em> NC_000964.3 (%)</th>
<th>Number of reads mapped</th>
</tr>
</thead>
<tbody>
<tr>
<td>M9_DEL_1</td>
<td>98.40</td>
<td>4,985,717</td>
</tr>
<tr>
<td>M9_DEL_2</td>
<td>98.03</td>
<td>3,107,568</td>
</tr>
<tr>
<td>M9_DEL_3</td>
<td>98.62</td>
<td>3,906,328</td>
</tr>
<tr>
<td>M9_W_1</td>
<td>98.62</td>
<td>6,243,859</td>
</tr>
<tr>
<td>M9_W_2</td>
<td>98.57</td>
<td>4,018,904</td>
</tr>
<tr>
<td>M9_W_3</td>
<td>98.57</td>
<td>3,162,411</td>
</tr>
<tr>
<td>SH2_DEL_1</td>
<td>96.64</td>
<td>3,223,257</td>
</tr>
<tr>
<td>SH2_DEL_2</td>
<td>97.88</td>
<td>3,651,905</td>
</tr>
<tr>
<td>SH2_DEL_3</td>
<td>97.99</td>
<td>4,227,724</td>
</tr>
<tr>
<td>SH2_W_1</td>
<td>94.86</td>
<td>1,554,197</td>
</tr>
<tr>
<td>SH2_W_2</td>
<td>97.29</td>
<td>1,479,861</td>
</tr>
<tr>
<td>SH2_W_3</td>
<td>97.03</td>
<td>932,054</td>
</tr>
<tr>
<td>SH5_DEL_1</td>
<td>95.36</td>
<td>4,184,141</td>
</tr>
<tr>
<td>SH5_DEL_2</td>
<td>97.86</td>
<td>5,609,482</td>
</tr>
<tr>
<td>SH5_DEL_3</td>
<td>96.8</td>
<td>2,769,349</td>
</tr>
<tr>
<td>SH5_W_1</td>
<td>98.57</td>
<td>6,119,117</td>
</tr>
<tr>
<td>SH5_W_2</td>
<td>96.47</td>
<td>2,856,831</td>
</tr>
<tr>
<td>SH5_W_3</td>
<td>97.71</td>
<td>3,831,957</td>
</tr>
</tbody>
</table>

Figure 3.2: Schematic of the steps taken in the differential gene expression analysis of ΔspoVG.

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 log₂ 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 ∆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 ∆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 ∆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 ∆*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 ΔspoVG 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.
Chapter 3 | Transcriptomics

Transcriptomic Analysis (vsd) — M9

(A) PCA Plot

PC1: 74% variance
PC2: 13% variance

media: M9

genotype: wt, ko
Figure 3.5: The replicates of WT and ΔspoVG 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 \( \Delta \)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

<table>
<thead>
<tr>
<th>Putative ncRNAs identified</th>
<th>M9</th>
<th>SH2</th>
<th>SH5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total no. of putative ncRNAs</td>
<td>32</td>
<td>7</td>
<td>19</td>
</tr>
<tr>
<td>Total no. of up-regulated putative ncRNAs</td>
<td>10</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>Total no. of down-regulated putative ncRNAs</td>
<td>22</td>
<td>5</td>
<td>13</td>
</tr>
</tbody>
</table>
3.2.5 Analysis of the Differentially Expressed Genes

To identify the genes differentially expressed in ∆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 ∆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 ∆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](image)

(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.
Chapter 3 | Transcriptomics

Volcano Plot

(A)

LFC (ΔspoVG vs WT)

−log10 padj (M9)

padj < 0.05
FALSE
TRUE
NA
Figure 3.7: Volcano plots show the up-regulated and down-regulated genes in $\Delta$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 $\triangle$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 $\triangle$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 $\triangle$spoVG in M9.
Table 3.3: Significantly DEGs in \( \Delta^{spoVG} \) in M9

Top 10 up-regulated (A) and down-regulated (B) DEGs in \( \Delta^{spoVG} \) in M9 are arranged in a decreasing and ascending order of LFC values respectively.

(A)

<table>
<thead>
<tr>
<th>Locus Tag</th>
<th>Gene</th>
<th>Product</th>
<th>Description</th>
<th>Function</th>
<th>LFC</th>
<th>padj</th>
</tr>
</thead>
<tbody>
<tr>
<td>new_3788276_3788397_c</td>
<td>S1427</td>
<td>intergenic region</td>
<td>new RNA feature, intergenic region</td>
<td>NA</td>
<td>3.410</td>
<td>2.66E-72</td>
</tr>
<tr>
<td>BSU20350</td>
<td>yorK</td>
<td>single-strand DNA-specific exonuclease</td>
<td>single-strand DNA-specific exonuclease</td>
<td>unknown</td>
<td>3.160</td>
<td>3.66E-05</td>
</tr>
<tr>
<td>BSU21290</td>
<td>yomN</td>
<td>unknown</td>
<td>unknown</td>
<td>unknown</td>
<td>2.333</td>
<td>6.08E-03</td>
</tr>
<tr>
<td>BSU21300</td>
<td>yomM</td>
<td>unknown</td>
<td>unknown</td>
<td>unknown</td>
<td>2.303</td>
<td>2.45E-03</td>
</tr>
<tr>
<td>BSU24740</td>
<td>corA</td>
<td>unknown</td>
<td>general stress protein, similar to magnesium transporter</td>
<td>unknown</td>
<td>2.058</td>
<td>2.91E-03</td>
</tr>
<tr>
<td>new_2246149_2246571</td>
<td>S804</td>
<td>3'UTR of yomM</td>
<td>new RNA feature, 3'UTR of yomM</td>
<td>NA</td>
<td>2.023</td>
<td>1.28E-02</td>
</tr>
<tr>
<td>BSU21190</td>
<td>yomX</td>
<td>unknown</td>
<td>unknown</td>
<td>unknown</td>
<td>1.995</td>
<td>9.06E-04</td>
</tr>
<tr>
<td>new_1116857_1117712_c</td>
<td>S365</td>
<td>intergenic region</td>
<td>new RNA feature, intergenic region</td>
<td>NA</td>
<td>1.786</td>
<td>4.04E-04</td>
</tr>
<tr>
<td>BSUB_NCR1117109</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>1.775</td>
<td>6.98E-04</td>
</tr>
<tr>
<td>BSU00500</td>
<td>gcaD</td>
<td>bifunctional N-acetylglucosamine-1-phosphate uridylyltransferase/glucosamine-1-phosphate acetyltransferase</td>
<td>bifunctional N-acetylglucosamine-1-phosphate uridylyltransferase/glucosamine-1-phosphate acetyltransferase</td>
<td>cell wall metabolism</td>
<td>1.760</td>
<td>1.22E-38</td>
</tr>
<tr>
<td>Locus Tag</td>
<td>Gene</td>
<td>Product</td>
<td>Description</td>
<td>Function</td>
<td>LFC</td>
<td>padj</td>
</tr>
<tr>
<td>-------------------------</td>
<td>---------</td>
<td>---------</td>
<td>----------------------------------------------------------------------------</td>
<td>--------------------------------</td>
<td>---------</td>
<td>-----------</td>
</tr>
<tr>
<td>BSUB_NCR2273724</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>-3.558</td>
<td>1.95E-30</td>
</tr>
<tr>
<td>BSU17170</td>
<td>pksI</td>
<td>unknown</td>
<td>involved in polyketide synthesis</td>
<td>polyketide synthesis</td>
<td>-3.336</td>
<td>2.28E-21</td>
</tr>
<tr>
<td>new_2273515_2273748</td>
<td>bsrG</td>
<td>unknown</td>
<td>hydrophobic class I toxin</td>
<td>stability of prophage SPβ</td>
<td>-3.249</td>
<td>1.70E-26</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>hydrophobic toxin; interferes with cell envelope biosynthesis, causes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>membrane invaginations together with delocalization of the cell wall</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>biosynthesis machinery, triggers autolysis, and inhibits protein</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>synthesis machinery, triggers autolysis, and inhibits protein biosynthesis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BSU17160</td>
<td>pksH</td>
<td>unknown</td>
<td>involved in polyketide synthesis</td>
<td>polyketide synthesis</td>
<td>-3.163</td>
<td>2.11E-29</td>
</tr>
<tr>
<td>BSU17140</td>
<td>pksF</td>
<td>unknown</td>
<td>involved in polyketide synthesis</td>
<td>polyketide synthesis</td>
<td>-3.090</td>
<td>1.53E-32</td>
</tr>
<tr>
<td>BSUB_NCR2069983</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>-3.036</td>
<td>1.18E-22</td>
</tr>
<tr>
<td>BSU17150</td>
<td>pksG</td>
<td>unknown</td>
<td>involved in polyketide synthesis</td>
<td>polyketide synthesis</td>
<td>-3.032</td>
<td>2.53E-26</td>
</tr>
<tr>
<td>BSU17130</td>
<td>acpK</td>
<td>acyl carrier protein</td>
<td>acyl carrier protein</td>
<td>polyketide biosynthesis</td>
<td>-3.004</td>
<td>3.11E-13</td>
</tr>
<tr>
<td>new_820648_820823</td>
<td>S254</td>
<td>independent transcript</td>
<td>new RNA feature, independent</td>
<td>osmoprotection</td>
<td>-2.952</td>
<td>1.65E-04</td>
</tr>
<tr>
<td>new_2069824_2069977</td>
<td>SR5</td>
<td>antisense RNA to bsrE</td>
<td>antisense RNA, controls</td>
<td>Control of BsrE</td>
<td>-2.901</td>
<td>3.55E-21</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>expression of the BsrE toxin</td>
<td>expression of the BsrE toxin</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
3.2.5.2 DEGs in SH2

In SH2 161 DEGs were affected in response to \(\Delta spoVG\), 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 \textit{de novo} biosynthesis of UMP, is an integral membrane protein and a regulator of protein PyrR (Turner, Lu and Switzi, 1994; Hahne et al., 2008). In addition to \(pyrC\) many other genes of this operon were also up-regulated in \(\Delta spoVG\) in SH2. S811 was the most down-regulated gene in SH2 in \(\Delta spoVG\) (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 \(\Delta spoVG\). 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 $\Delta$spoVG in SH2

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

NA = not available

<table>
<thead>
<tr>
<th>Locus Tag</th>
<th>Gene</th>
<th>Product</th>
<th>Description</th>
<th>Function</th>
<th>LFC</th>
<th>padj</th>
</tr>
</thead>
<tbody>
<tr>
<td>BSU15500</td>
<td>pyrC</td>
<td>dihydroorotase</td>
<td>dihydroorotase</td>
<td>pyrimidine biosynthesis</td>
<td>10.834</td>
<td>1.63E-08</td>
</tr>
<tr>
<td>new_1630017_1630089</td>
<td>S568</td>
<td>intergenic region</td>
<td>new RNA feature, intergenic region</td>
<td>NA</td>
<td>8.611</td>
<td>2.73E-04</td>
</tr>
<tr>
<td>new_3788276_3788397_c</td>
<td>S1427</td>
<td>intergenic region</td>
<td>new RNA feature, intergenic region</td>
<td>NA</td>
<td>8.517</td>
<td>9.26E-134</td>
</tr>
<tr>
<td>BSU15560</td>
<td>pyrE</td>
<td>orotate phosphoribosyltransferase</td>
<td>orotate phosphoribosyltransferase</td>
<td>pyrimidine biosynthesis</td>
<td>7.551</td>
<td>4.60E-23</td>
</tr>
<tr>
<td>BSU15520</td>
<td>pyrAB</td>
<td>carbamoyl-phosphate synthetase</td>
<td>carbamoyl-phosphate synthetase</td>
<td>pyrimidine biosynthesis</td>
<td>6.905</td>
<td>2.32E-10</td>
</tr>
<tr>
<td>BSU15540</td>
<td>pyrD</td>
<td>dihydroorotic acid dehydrogenase</td>
<td>dihydroorotic acid dehydrogenase</td>
<td>pyrimidine biosynthesis</td>
<td>6.773</td>
<td>8.03E-25</td>
</tr>
<tr>
<td>BSU15550</td>
<td>pyrF</td>
<td>orotidine-5-phosphate decarboxylase</td>
<td>orotidine 5-phosphate decarboxylase</td>
<td>pyrimidine biosynthesis</td>
<td>6.772</td>
<td>5.88E-24</td>
</tr>
<tr>
<td>BSU15530</td>
<td>pyrK</td>
<td>dihydroorotic acid dehydrogenase</td>
<td>dihydroorotic acid dehydrogenase</td>
<td>pyrimidine biosynthesis</td>
<td>6.704</td>
<td>2.30E-25</td>
</tr>
<tr>
<td>new_3535821_3536011</td>
<td>S1314</td>
<td>5'UTR of sacB</td>
<td>new RNA feature, 5'UTR of sacB</td>
<td>6.351</td>
<td>2.98E-04</td>
<td></td>
</tr>
<tr>
<td>BSU15510</td>
<td>pyrAA</td>
<td>carbamoyl-phosphate synthetase</td>
<td>carbamoyl-phosphate synthetase</td>
<td>pyrimidine biosynthesis</td>
<td>6.138</td>
<td>2.93E-20</td>
</tr>
<tr>
<td>Locus Tag</td>
<td>Gene</td>
<td>Product</td>
<td>Description</td>
<td>Function</td>
<td>LFC</td>
<td>padj</td>
</tr>
<tr>
<td>------------------------</td>
<td>------------</td>
<td>-----------------</td>
<td>------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>-------------------</td>
<td>------</td>
<td>----------</td>
</tr>
<tr>
<td>BSUB_NCR2273724</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>-4.704</td>
<td>9.83E-23</td>
</tr>
<tr>
<td>new_2273801_2274302</td>
<td>S811</td>
<td>independent transcript</td>
<td>new RNA feature, independent transcript</td>
<td>NA</td>
<td>-4.589</td>
<td>1.98E-20</td>
</tr>
<tr>
<td>new_2273515_2273748</td>
<td>bsrG</td>
<td>hydrophobic class I toxin</td>
<td>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</td>
<td>stability of prophage SP-β</td>
<td>-4.498</td>
<td>9.93E-22</td>
</tr>
<tr>
<td>new_1386984_1387035_c</td>
<td>S472</td>
<td>5'UTR of ispA</td>
<td>new RNA feature, 5'UTR of ispA</td>
<td>NA</td>
<td>-3.064</td>
<td>5.33E-04</td>
</tr>
<tr>
<td>new_2100447_2100579</td>
<td>S744</td>
<td>5'UTR of dhaS</td>
<td>new RNA feature, 5'UTR of dhaS</td>
<td>NA</td>
<td>-2.890</td>
<td>2.33E-23</td>
</tr>
<tr>
<td>BSU37350</td>
<td>sboA</td>
<td>subtilosin A</td>
<td>subtilosin A</td>
<td>antimicrobial peptide</td>
<td>-2.806</td>
<td>6.86E-07</td>
</tr>
<tr>
<td>new_3239640_3239929</td>
<td>S1208</td>
<td>5'UTR of nupN</td>
<td>new RNA feature, 5'UTR of nupN</td>
<td>-2.704</td>
<td>2.44E-09</td>
<td></td>
</tr>
<tr>
<td>BSU13190</td>
<td>ispA</td>
<td>intracellular serine protease</td>
<td>intracellular serine protease</td>
<td>protein degradation</td>
<td>-2.289</td>
<td>5.98E-09</td>
</tr>
<tr>
<td>BSU19310</td>
<td>dhaS</td>
<td>aldehyde dehydrogenase (NAD)</td>
<td>3-hydroxypropionaldehyde-specific aldehyde dehydrogenase (NAD)</td>
<td>unknown</td>
<td>-2.286</td>
<td>2.23E-13</td>
</tr>
<tr>
<td>BSU39320</td>
<td>yxiB</td>
<td>unknown</td>
<td>unknown</td>
<td>unknown</td>
<td>-2.234</td>
<td>2.13E-02</td>
</tr>
</tbody>
</table>
Table 3.5: Top 10 significantly DEGs in ∆spoVG in SH5

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

<table>
<thead>
<tr>
<th>Locus Tag</th>
<th>Gene</th>
<th>Product</th>
<th>Description</th>
<th>Function</th>
<th>LFC</th>
<th>padj</th>
</tr>
</thead>
<tbody>
<tr>
<td>new_3788276_3788397_c</td>
<td>S1427</td>
<td>intergenic region</td>
<td>new RNA feature, intergenic region</td>
<td>NA</td>
<td>7.256</td>
<td>7.80E-135</td>
</tr>
<tr>
<td>BSU00500</td>
<td></td>
<td></td>
<td></td>
<td>cell wall metabolism</td>
<td>3.244</td>
<td>6.05E-42</td>
</tr>
<tr>
<td>BSU24620</td>
<td>tasA</td>
<td>major component of biofilm matrix</td>
<td>major component of biofilm matrix, forms amyloid fibers</td>
<td>biofilm formation</td>
<td>2.180</td>
<td>2.45E-06</td>
</tr>
<tr>
<td>BSU00510</td>
<td></td>
<td></td>
<td></td>
<td>biosynthesis of histidine</td>
<td>2.058</td>
<td>3.24E-21</td>
</tr>
<tr>
<td>BSU06230</td>
<td></td>
<td></td>
<td></td>
<td>myo-inositol uptake</td>
<td>1.895</td>
<td>1.57E-12</td>
</tr>
<tr>
<td>BSU34300</td>
<td>epsH</td>
<td>undecaprenyl (UnDP) priming UDP-N-acetylglucosamine transferase</td>
<td>undecaprenyl priming UDP-N-acetylglucosamine transferase</td>
<td>synthesis of extracellular poly-N-acetylglucosamine</td>
<td>1.860</td>
<td>3.55E-02</td>
</tr>
<tr>
<td>new_4007659_4007722_c</td>
<td>S1504</td>
<td>5'UTR of yxiT/2</td>
<td>new RNA feature, 5'UTR of yxiT/2</td>
<td></td>
<td>1.815</td>
<td>1.59E-03</td>
</tr>
<tr>
<td>BSU39030</td>
<td></td>
<td></td>
<td></td>
<td>unknown</td>
<td>1.737</td>
<td>2.71E-04</td>
</tr>
<tr>
<td>BSU24670</td>
<td></td>
<td></td>
<td></td>
<td>genetic competence, DNA uptake</td>
<td>1.586</td>
<td>3.99E-02</td>
</tr>
<tr>
<td>BSU39029</td>
<td></td>
<td></td>
<td></td>
<td>unknown</td>
<td>1.576</td>
<td>3.67E-04</td>
</tr>
<tr>
<td>Locus Tag</td>
<td>Gene</td>
<td>Product</td>
<td>Description</td>
<td>Function</td>
<td>LFC</td>
<td>padj</td>
</tr>
<tr>
<td>-----------------</td>
<td>-------------</td>
<td>---------</td>
<td>----------------------------------------------------------------------------</td>
<td>---------------------------------</td>
<td>---------</td>
<td>---------</td>
</tr>
<tr>
<td>BSUB_NCR2273724</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>-3.685</td>
<td>2.29E-12</td>
</tr>
<tr>
<td>BSUB_NCR2069983</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>-3.499</td>
<td>2.90E-08</td>
</tr>
<tr>
<td>new_2069824_2069977</td>
<td>SR5</td>
<td>antisense RNA to bsrE</td>
<td>antisense RNA, controls expression of the BsrE toxin</td>
<td>control of BsrE toxin expression</td>
<td>-3.399</td>
<td>1.95E-08</td>
</tr>
<tr>
<td>new_2273515_2273748</td>
<td>bsrG</td>
<td>hydrophobic class I toxin</td>
<td>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</td>
<td>stability of prophage SP-beta</td>
<td>-3.047</td>
<td>7.45E-10</td>
</tr>
<tr>
<td>new_3510674_3510779</td>
<td>S1309</td>
<td>5'UTR of lutP</td>
<td>new RNA feature, 5'UTR of lutP</td>
<td>NA</td>
<td>-1.791</td>
<td>8.06E-03</td>
</tr>
<tr>
<td>BSU28099</td>
<td>yszA</td>
<td>unknown</td>
<td>unknown</td>
<td>unknown</td>
<td>-1.720</td>
<td>6.99E-04</td>
</tr>
<tr>
<td>new_2273801_2274302</td>
<td>S811</td>
<td>independent transcript</td>
<td>new RNA feature, independent transcript</td>
<td>NA</td>
<td>-1.490</td>
<td>1.41E-03</td>
</tr>
<tr>
<td>BSU36940</td>
<td>ywlD</td>
<td>unknown</td>
<td>unknown</td>
<td>unknown</td>
<td>-1.477</td>
<td>1.30E-02</td>
</tr>
<tr>
<td>new_1383187_1383294_c</td>
<td>S469</td>
<td>intergenic region</td>
<td>new RNA feature, intergenic region</td>
<td>NA</td>
<td>-1.451</td>
<td>3.45E-02</td>
</tr>
<tr>
<td>new_1571801_1571902_c</td>
<td>S551</td>
<td>5'UTR of ylbJ</td>
<td>new RNA feature, 5'UTR of ylbJ</td>
<td>NA</td>
<td>-1.413</td>
<td>3.03E-02</td>
</tr>
</tbody>
</table>
3.2.5.4 Common DEGs in all the conditions

To identify any commonly identified DEGs in $\Delta$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 $\Delta$spoVG strain identified commonly altered significantly DEGs across all the growth conditions.](image)

Venn diagram represents the overlap of the significantly differentially expressed genes (padj < 0.05) in the $\Delta$spoVG mutant compared to the WT in different growth conditions (M9, SH2 and SH5).
Table 3.6: Significantly differentially expressed genes (DEGs) in ΔspoVG vs WT found commonly in M9, SH2 and SH5

NA = not available

<table>
<thead>
<tr>
<th>Locus Tag</th>
<th>Gene</th>
<th>Product</th>
<th>Description</th>
<th>Function</th>
<th>M9</th>
<th>SH2</th>
<th>SH5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>LFC</td>
<td>padj</td>
<td>LFC</td>
</tr>
<tr>
<td><strong>Up-regulated in all the conditions</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>new_3788276_3788397_c</td>
<td>S1427</td>
<td>intergenic region</td>
<td>new RNA feature, intergenic region, flanked by S1426 downstream and upp upstream</td>
<td>NA</td>
<td>3.410</td>
<td>2.66E-72</td>
<td>8.517</td>
</tr>
<tr>
<td>BSU00500</td>
<td>gcaD</td>
<td>bifunctional N-acetylglucosamine-1-phosphate</td>
<td>bifunctional N-acetylglucosamine-1-phosphate uridylyltransferase/glucosamine-1-phosphate acetyltransferase</td>
<td>cell wall metabolism</td>
<td>1.760</td>
<td>1.22E-38</td>
<td>3.459</td>
</tr>
<tr>
<td>BSU00510</td>
<td>prs</td>
<td>phosphoribosylpyrophosphate synthetase</td>
<td>phosphoribosylpyrophosphate synthetase, universally conserved protein</td>
<td>phosphoribosylpyrophosphate synthesis (biosynthesis of histidine)</td>
<td>1.640</td>
<td>4.62E-35</td>
<td>3.076</td>
</tr>
<tr>
<td>BSU38400</td>
<td>epr</td>
<td>minor extracellular serine protease</td>
<td>minor extracellular serine protease, involved in control of swarming motility</td>
<td>protein degradation</td>
<td>1.063</td>
<td>7.58E-05</td>
<td>0.942</td>
</tr>
</tbody>
</table>
### Down-regulated in all the conditions

<table>
<thead>
<tr>
<th>Gene ID</th>
<th>Gene Name</th>
<th>Description</th>
<th>Regulation Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>BSUB_NCR2273724</td>
<td>bsrG</td>
<td>hydrophobic class I toxin</td>
<td>stability of prophage SPβ</td>
</tr>
<tr>
<td></td>
<td></td>
<td>hydrophobic toxin; interferes with cell envelope biosynthesis, causes</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>membrane invaginations together with delocalization of the cell wall synthesis</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>machinery, triggers autolysis, and inhibits protein biosynthesis</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BSUB_NCR2069983</td>
<td>SR5</td>
<td>antisense RNA to bsrE</td>
<td>control of BsrE toxin expression</td>
</tr>
<tr>
<td></td>
<td></td>
<td>antisense RNA, controls expression of the BsrE toxin</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>new_2273801_2274302</td>
<td>S811</td>
<td>independent transcript</td>
<td>new RNA feature, independent transcript, potential 3’UTR of bsrG, antisense to yokL and asrG</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BSU31550</td>
<td>nupO</td>
<td>ABC transporter for guanosine (ATP-binding protein)</td>
<td>uptake of guanosine</td>
</tr>
</tbody>
</table>

**Table notes:**
- The table lists genes that were down-regulated in all the conditions.
- BSUB_NCR2273724: new_2273515_2273748
- BSUB_NCR2069983: new_2069824_2069977
- new_2273801_2274302
- BSU31550: nupO

**Regulation Information:**
- Stability of prophage SPβ: Down-regulated
- Control of BsrE toxin expression: Down-regulated
- Uptake of guanosine: Down-regulated
### Mixed LFC

<table>
<thead>
<tr>
<th>Gene ID</th>
<th>Feature</th>
<th>Description</th>
<th>Fold Change</th>
<th>Genotype</th>
<th>Control 1</th>
<th>Control 2</th>
<th>Control 3</th>
<th>Control 4</th>
<th>Control 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>new_3740106_3740205_c</td>
<td>3'UTR of ssbB</td>
<td>new RNA feature, 3'UTR of ssbB</td>
<td>NA</td>
<td>-1.447</td>
<td>1.73E-14</td>
<td>1.881</td>
<td>2.44E-09</td>
<td>1.232</td>
<td>2.54E-06</td>
</tr>
<tr>
<td>BSU36310_copy2</td>
<td>NA</td>
<td>new RNA feature, 3'UTR of ssbB</td>
<td>NA</td>
<td>-1.419</td>
<td>9.53E-17</td>
<td>1.899</td>
<td>6.89E-11</td>
<td>1.221</td>
<td>9.86E-06</td>
</tr>
<tr>
<td>BSU24690</td>
<td>comGE</td>
<td>minor pseudopilin DNA transport machinery, minor pseudopilin</td>
<td>-0.880</td>
<td>3.68E-02</td>
<td>1.805</td>
<td>1.33E-02</td>
<td>1.160</td>
<td>7.86E-03</td>
<td></td>
</tr>
<tr>
<td>BSU24710</td>
<td>comGC</td>
<td>major pseudopilin DNA uptake, major component of the pseudopilus</td>
<td>-0.858</td>
<td>8.17E-03</td>
<td>2.310</td>
<td>2.56E-04</td>
<td>1.247</td>
<td>2.71E-04</td>
<td></td>
</tr>
<tr>
<td>BSU24730</td>
<td>comGA</td>
<td>traffic ATPase late competence protein, traffic ATPase, binds and transports transforming DNA, required for growth arrest of competent cells</td>
<td>-0.804</td>
<td>1.38E-03</td>
<td>2.369</td>
<td>9.50E-11</td>
<td>1.236</td>
<td>3.89E-04</td>
<td></td>
</tr>
<tr>
<td>BSU24720</td>
<td>comGB</td>
<td>DNA transport machinery polytopic membrane protein, DNA transport machinery</td>
<td>-0.767</td>
<td>1.06E-03</td>
<td>2.055</td>
<td>2.63E-10</td>
<td>1.387</td>
<td>4.58E-05</td>
<td></td>
</tr>
<tr>
<td>BSU24700</td>
<td>comGD</td>
<td>minor pseudopilin DNA transport machinery, minor pseudopilin</td>
<td>-0.751</td>
<td>4.70E-02</td>
<td>2.370</td>
<td>6.00E-04</td>
<td>1.065</td>
<td>1.27E-03</td>
<td></td>
</tr>
<tr>
<td>BSU35480</td>
<td>degV</td>
<td>fatty acid binding protein fatty acid binding protein</td>
<td>-0.638</td>
<td>5.52E-03</td>
<td>1.287</td>
<td>1.41E-02</td>
<td>1.269</td>
<td>3.34E-04</td>
<td></td>
</tr>
<tr>
<td>new_368894_371728_c</td>
<td>intergenic region</td>
<td>new RNA feature, intergenic region, flanked by yckB downstream and nin upstream, antisense to BSU_03385, yckC, yckD and yckE</td>
<td>NA</td>
<td>-0.583</td>
<td>1.07E-02</td>
<td>2.179</td>
<td>1.48E-11</td>
<td>1.102</td>
<td>1.11E-03</td>
</tr>
<tr>
<td>BSU12440</td>
<td>phrA</td>
<td>phosphatase (RapA) inhibitor response regulator aspartate phosphatase (RapA) inhibitor, control of the phosphorelay</td>
<td>-0.522</td>
<td>4.94E-02</td>
<td>-1.246</td>
<td>1.00E-02</td>
<td>0.994</td>
<td>4.63E-03</td>
<td></td>
</tr>
</tbody>
</table>
3.2.6 Gene Set Enrichment Analysis

To understand the biological changes in the transcriptome of ΔspoVG 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 △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 ΔspoVG from the transcriptomic data

<table>
<thead>
<tr>
<th>GSEA - Category 3</th>
<th>padj</th>
<th>Total genes in pathway</th>
<th>Genes up-regulated</th>
<th>Genes down-regulated</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>M9</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Biosynthesis of antibacterial compounds</td>
<td>2.42E-03</td>
<td>50</td>
<td>1</td>
<td>49</td>
</tr>
<tr>
<td>General stress proteins (controlled by SigB)</td>
<td>2.42E-03</td>
<td>151</td>
<td>127</td>
<td>24</td>
</tr>
<tr>
<td>Miscellaneous metabolic pathways</td>
<td>2.42E-03</td>
<td>73</td>
<td>15</td>
<td>58</td>
</tr>
<tr>
<td>Prophage 1</td>
<td>2.42E-03</td>
<td>15</td>
<td>6</td>
<td>9</td>
</tr>
<tr>
<td>Toxins, antitoxins and immunity against toxins</td>
<td>2.42E-03</td>
<td>17</td>
<td>0</td>
<td>17</td>
</tr>
<tr>
<td>Heat shock proteins</td>
<td>8.07E-03</td>
<td>27</td>
<td>21</td>
<td>6</td>
</tr>
<tr>
<td>Phosphorelay</td>
<td>1.21E-02</td>
<td>44</td>
<td>17</td>
<td>27</td>
</tr>
<tr>
<td>Cell envelope stress proteins (controlled by SigM, V, W, X, Y)</td>
<td>1.97E-02</td>
<td>133</td>
<td>65</td>
<td>68</td>
</tr>
<tr>
<td>ABC transporters</td>
<td>2.55E-02</td>
<td>170</td>
<td>65</td>
<td>105</td>
</tr>
<tr>
<td>Quorum sensing</td>
<td>2.66E-02</td>
<td>22</td>
<td>7</td>
<td>15</td>
</tr>
<tr>
<td>Acquisition of iron</td>
<td>2.75E-02</td>
<td>44</td>
<td>16</td>
<td>28</td>
</tr>
<tr>
<td>Biosynthesis/ acquisition of nucleotides</td>
<td>3.63E-02</td>
<td>63</td>
<td>51</td>
<td>12</td>
</tr>
<tr>
<td><strong>SH2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Biosynthesis/ acquisition of nucleotides</td>
<td>6.10E-03</td>
<td>56</td>
<td>35</td>
<td>21</td>
</tr>
<tr>
<td>Genetic competence</td>
<td>6.10E-03</td>
<td>57</td>
<td>40</td>
<td>17</td>
</tr>
<tr>
<td>Phosphorylation on an Arg residue</td>
<td>1.22E-02</td>
<td>102</td>
<td>75</td>
<td>27</td>
</tr>
<tr>
<td><strong>SH5</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Genetic competence</td>
<td>1.22E-02</td>
<td>55</td>
<td>48</td>
<td>7</td>
</tr>
<tr>
<td>General stress proteins (controlled by SigB)</td>
<td>2.03E-02</td>
<td>154</td>
<td>127</td>
<td>27</td>
</tr>
<tr>
<td>Motility and chemotaxis</td>
<td>2.03E-02</td>
<td>68</td>
<td>64</td>
<td>4</td>
</tr>
<tr>
<td>Biosynthesis/ acquisition of amino acids</td>
<td>2.74E-02</td>
<td>145</td>
<td>41</td>
<td>104</td>
</tr>
<tr>
<td>Sporulation proteins</td>
<td>3.17E-02</td>
<td>485</td>
<td>148</td>
<td>337</td>
</tr>
<tr>
<td>Quorum sensing</td>
<td>3.66E-02</td>
<td>21</td>
<td>18</td>
<td>3</td>
</tr>
<tr>
<td>Utilization of specific carbon sources</td>
<td>3.66E-02</td>
<td>219</td>
<td>142</td>
<td>77</td>
</tr>
<tr>
<td>Utilization of lipids</td>
<td>4.42E-02</td>
<td>30</td>
<td>20</td>
<td>10</td>
</tr>
</tbody>
</table>
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 ΔspoVG 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 ∆*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 ∆*spoVG* (Figure 3.9). In this gene set, 17 genes were identified in ∆*spoVG* (Table 3.7). Noticeably, all of the affected toxin/antitoxin genes were down-regulated in ∆*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 ∆*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 ΔspoVG 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 ΔspoVG. 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 ∆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 ∆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 ΔspoVG (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 ∆*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 downregulated (Table 3.7). Among these were the genes of *dbh* operon and the *yvmC-cypX* operon. The *dbh* 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 ∆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 ∆*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 ∆*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 ΔspoVG in *B. subtilis*

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

<table>
<thead>
<tr>
<th>KEGG Pathway – M9</th>
<th>padj</th>
<th>No. of gene in pathway (k)</th>
<th>No. of genes enriched (x)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biosynthesis of siderophore group nonribosomal peptides</td>
<td>1.22E-03</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>Nonribosomal peptide structures</td>
<td>2.26E-03</td>
<td>9</td>
<td>6</td>
</tr>
<tr>
<td>Biosynthesis of secondary metabolites - unclassified</td>
<td>2.61E-02</td>
<td>10</td>
<td>5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>KEGG Pathway – SH2</th>
<th>padj</th>
<th>No. of gene in pathway (k)</th>
<th>No. of genes enriched (x)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pyrimidine metabolism</td>
<td>9.09E-06</td>
<td>49</td>
<td>9</td>
</tr>
<tr>
<td>Ribosome</td>
<td>3.26E-05</td>
<td>57</td>
<td>9</td>
</tr>
<tr>
<td>Quorum sensing</td>
<td>3.39E-03</td>
<td>68</td>
<td>7</td>
</tr>
<tr>
<td>ABC transporters</td>
<td>7.33E-03</td>
<td>140</td>
<td>10</td>
</tr>
<tr>
<td>Histidine metabolism</td>
<td>1.43E-02</td>
<td>18</td>
<td>3</td>
</tr>
<tr>
<td>beta-Alanine metabolism</td>
<td>2.65E-02</td>
<td>9</td>
<td>2</td>
</tr>
<tr>
<td>Valine, leucine and isoleucine degradation</td>
<td>2.79E-02</td>
<td>23</td>
<td>3</td>
</tr>
<tr>
<td>Arginine and proline metabolism</td>
<td>3.48E-02</td>
<td>25</td>
<td>3</td>
</tr>
<tr>
<td>Metabolic pathways</td>
<td>3.88E-02</td>
<td>598</td>
<td>25</td>
</tr>
<tr>
<td>Lysine degradation</td>
<td>3.89E-02</td>
<td>11</td>
<td>2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>KEGG Pathway – SH5</th>
<th>padj</th>
<th>No. of gene in pathway (k)</th>
<th>No. of genes enriched (x)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inositol phosphate metabolism</td>
<td>1.72E-02</td>
<td>15</td>
<td>8</td>
</tr>
</tbody>
</table>
Figure 3.11: Significantly enriched KEGG pathways identified in the transcriptomic analysis of \( \Delta \text{spo VG} \) in \textit{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 \textit{pps} operon and all the genes were down-regulated. This result was comparable to the down-regulation and enrichment of genes involved in the \textit{Biosynthesis of antimicrobial compounds} (section 3.2.6.1). All the genes in \textit{Pyrimidine metabolism} in SH2 were up-regulated in \( \Delta \text{spo VG} \) and were comparable to the \textit{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 \( \Delta \text{spo VG} \) in SH2 (Figure 3.11). These observations suggested a role of \textit{spo VG} in nucleotide and amino acid metabolism in \textit{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 ΔspoVG

To identify the effect of ΔspoVG 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 ΔspoVG (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 $\Delta$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 ΔspoVG 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 ΔspoVG (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 ΔspoVG 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 $\Delta$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 $\Delta$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 $\Delta$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 $\Delta$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 $\Delta$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 $\Delta$spoVG in M9 when compared with the WT. A slight increase in the growth of $\Delta$spoVG and $\Delta$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 reglon 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 ∆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 ΔspoVG

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 ΔspoVG 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 ΔspoVG mutation in the *B. subtilis* 168, *trp*+ strain. This chapter discusses the identification of the differentially produced proteins (DPPs) in ΔspoVG 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 ∆spoVG in LB

To study the effect of ∆spoVG on the growth of the bacteria, the mutant strain was grown in LB and compared to the WT. A complemented strain (ΔspoVG::spoVG) was also included. The growth curves showed no growth defect in ΔspoVG (Figure 4.1). However, a mild growth defect in ΔspoVG 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 ΔspoVG 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.

![Growth Curve](image)

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

The growth curves represent the growth of WT, ΔspoVG and ΔspoVG::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).
4.2.3.1 Data filtering and transformation

The MS² identified 2,866 proteins for each of the growth conditions for both the WT and ΔspoVG 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 log₂ 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 ΔspoVG 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 \( \Delta \text{spoVG} \) 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 \( \text{spoVG} \) 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 \( \Delta \text{spoVG} \) suggested that the experiments were highly reproducible.

For the sporulation conditions, the strains of the \( \Delta \text{spoVG} \) and WT showed R values between 0.91-0.96 (Figure 4.4 (C and D)), indicating some changes in the proteome of \( \Delta \text{spoVG} \) in both the sporulation conditions tested. Further, the correlation among the biological replicates of the WT and the \( \Delta \text{spoVG} \) 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 \( \Delta \text{spoVG} \).

### 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).
Figure 4.4: Pearson correlation analysis of the proteomes from the WT and ΔspoVG 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 = ΔspoVG; 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 imputed 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 \( \Delta spoVG \) 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 \( \Delta spoVG \) 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 \( \Delta spoVG \) 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 \(\Delta spoVG\) 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: \(W = WT\); \(D = \Delta spoVG\)
Figure 4.6: The replicates of WT and ΔspoVG proteome cluster separately based on their genotypes

Clustering of proteomes of the WT and the ΔspoVG 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 (○) = ΔspoVG.
4.2.4 Analysis of the Differentially Produced Proteins (DPPs)

To identify the DPPs in $\Delta$spoVG, comparisons were made between the protein intensities of $\Delta$spoVG vs the WT control. DPPs with adjusted $p$-value ($p_{adj}$) < 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 $-\log_{10}(p_{adj})$ were plotted against respective logFC values on Volcano plots (Figure 4.7). The analysis identified both significantly increased and decreased differentially produced proteins in $\Delta$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 $\Delta$spoVG

<table>
<thead>
<tr>
<th>Growth condition</th>
<th>Proteins with logFC &lt; -2 (Decreased)</th>
<th>Proteins with logFC &gt; 2 (Increased)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LB</td>
<td>36</td>
<td>33</td>
</tr>
<tr>
<td>M9</td>
<td>46</td>
<td>34</td>
</tr>
<tr>
<td>SH2</td>
<td>131</td>
<td>60</td>
</tr>
<tr>
<td>SH5</td>
<td>173</td>
<td>32</td>
</tr>
</tbody>
</table>
Figure 4.7: Distribution of the DPPs in all the growth conditions

The Volcano plots of -log₁₀ 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).
4.2.4.1 DPPs in LB

In the LB condition, 92 proteins were differentially produced in ΔspoVG, among which 49 increased and 43 decreased in the mutant. The number of DPPs recorded in ΔspoVG 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 ΔspoVG (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 ∆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 ∆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 ∆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 ∆spoVG

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

<table>
<thead>
<tr>
<th>Locus Tag</th>
<th>Gene</th>
<th>Product</th>
<th>Description</th>
<th>Function</th>
<th>logFC</th>
<th>padj</th>
</tr>
</thead>
<tbody>
<tr>
<td>BSU08080</td>
<td>acoC</td>
<td>acetoin dehydrogenase E2 component (dihydrolipoamide acetyltransferase)</td>
<td>acetoin dehydrogenase E2 component (dihydrolipoamide acetyltransferase)</td>
<td>acetoin utilization</td>
<td>10.966</td>
<td>2.15E-03</td>
</tr>
<tr>
<td>BSU10350</td>
<td>yhfS</td>
<td>unknown</td>
<td>similar to acetyl-CoA C-acetyltransferase</td>
<td>unknown</td>
<td>10.870</td>
<td>5.40E-19</td>
</tr>
<tr>
<td>BSU13819</td>
<td>ykzS</td>
<td>unknown</td>
<td>unknown</td>
<td>unknown</td>
<td>10.459</td>
<td>5.85E-04</td>
</tr>
<tr>
<td>BSU32770</td>
<td>yusE</td>
<td>unknown</td>
<td>similar to thioredoxin</td>
<td>unknown</td>
<td>9.649</td>
<td>4.92E-04</td>
</tr>
<tr>
<td>BSU05600</td>
<td>ydgE</td>
<td>unknown</td>
<td>putative N-acetyltransferase</td>
<td>unknown</td>
<td>8.954</td>
<td>1.17E-13</td>
</tr>
<tr>
<td>BSU15400</td>
<td>ylmG</td>
<td>unknown</td>
<td>unknown</td>
<td>unknown</td>
<td>8.849</td>
<td>2.23E-14</td>
</tr>
<tr>
<td>BSU31500</td>
<td>yuxK</td>
<td>unknown</td>
<td>unknown</td>
<td>unknown</td>
<td>8.741</td>
<td>1.29E-03</td>
</tr>
<tr>
<td>BSU39080</td>
<td>licT</td>
<td>transcriptional antiterminator (BglG family)</td>
<td>transcriptional antiterminator of the bglP-bglH-yxiiE operon and bglS</td>
<td>control of beta-glucan and beta-glucoside utilization</td>
<td>8.648</td>
<td>2.46E-16</td>
</tr>
<tr>
<td>BSU20390</td>
<td>yorG</td>
<td>unknown</td>
<td>unknown</td>
<td>unknown</td>
<td>8.645</td>
<td>1.89E-20</td>
</tr>
<tr>
<td>BSU23040</td>
<td>fer</td>
<td>ferredoxin</td>
<td>ferredoxin</td>
<td>electron transfer</td>
<td>8.549</td>
<td>2.34E-02</td>
</tr>
<tr>
<td>Locus Tag</td>
<td>Gene</td>
<td>Product</td>
<td>Description</td>
<td>Function</td>
<td>logFC</td>
<td>padj</td>
</tr>
<tr>
<td>-----------</td>
<td>------</td>
<td>------------------------------</td>
<td>-------------------------------------------------------</td>
<td>-----------------------------------</td>
<td>--------</td>
<td>------------</td>
</tr>
<tr>
<td>BSU34930</td>
<td>hisZ</td>
<td>histidyl-tRNA synthetase</td>
<td>histidyl-tRNA synthetase</td>
<td>translation</td>
<td>-10.070</td>
<td>1.75E-13</td>
</tr>
<tr>
<td>BSU21990</td>
<td>ypdQ</td>
<td>unknown</td>
<td>similar to RNase H1</td>
<td>unknown</td>
<td>-9.550</td>
<td>1.55E-18</td>
</tr>
<tr>
<td>BSU15630</td>
<td>ylnF</td>
<td>probably precorrin-2 dehydrogenase</td>
<td>probably precorrin-2 dehydrogenase</td>
<td>siroheme biosynthesis, sulfite reduction</td>
<td>-9.370</td>
<td>1.48E-14</td>
</tr>
<tr>
<td>BSU02990</td>
<td>opuAB</td>
<td>glycine betaine and arsenobetaine [ABC transporter] (permease)</td>
<td>glycine betaine and arsenobetaine [ABC transporter] (permease)</td>
<td>compatible solute transport</td>
<td>-9.228</td>
<td>1.58E-14</td>
</tr>
<tr>
<td>BSU02650</td>
<td>pcp</td>
<td>pyrrolidone-carboxylate peptidase</td>
<td>pyrrolidone-carboxylate peptidase</td>
<td>removal of the N-terminal pyroglutamyl group from peptides</td>
<td>-8.828</td>
<td>7.41E-13</td>
</tr>
<tr>
<td>BSU29690</td>
<td>acuA</td>
<td>Gcn5-related N-acetyltransferase</td>
<td>protein acetylase for the control of AcsA activity</td>
<td>control of AcsA activity</td>
<td>-8.791</td>
<td>1.42E-14</td>
</tr>
<tr>
<td>BSU35200</td>
<td>yvkB</td>
<td>unknown</td>
<td>similar to transcriptional regulator (TetR family)</td>
<td>unknown</td>
<td>-8.346</td>
<td>2.03E-23</td>
</tr>
<tr>
<td>BSU32610</td>
<td>frlB</td>
<td>fructoselysine-6-P-glycosidase</td>
<td>fructoselysine-6-P-glycosidase</td>
<td>metabolism of aminoacylated fructose</td>
<td>-8.314</td>
<td>1.42E-11</td>
</tr>
<tr>
<td>BSU31060</td>
<td>gbsA</td>
<td>glycine betaine-aldehyde dehydrogenase</td>
<td>glycine betaine-aldehyde dehydrogenase, glycine betaine synthesis</td>
<td>osmoprotection</td>
<td>-8.291</td>
<td>3.61E-02</td>
</tr>
<tr>
<td>BSU28800</td>
<td>araA</td>
<td>L-arabinose isomerase</td>
<td>L-arabinose isomerase</td>
<td>arabinose utilization</td>
<td>-8.164</td>
<td>2.58E-16</td>
</tr>
</tbody>
</table>
### Table 4.3: Significant DPPs in M9 in \( \Delta spoVG \)

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

<table>
<thead>
<tr>
<th>Locus Tag</th>
<th>Gene</th>
<th>Product</th>
<th>Description</th>
<th>Function</th>
<th>logFC</th>
<th>padj</th>
</tr>
</thead>
<tbody>
<tr>
<td>BSU15090</td>
<td>gerR</td>
<td>unknown</td>
<td>probably DNA-binding protein, regulates transcription of some spore coat genes</td>
<td>spore coat formation and resistance of spores to lysozyme</td>
<td>12.435</td>
<td>1.19E-09</td>
</tr>
<tr>
<td>BSU11450</td>
<td>oppC</td>
<td>oligopeptide ABC transporter (permease)</td>
<td>oligopeptide ABC transporter (permease)</td>
<td>initiation of sporulation, competence development</td>
<td>9.632</td>
<td>5.71E-13</td>
</tr>
<tr>
<td>BSU30940</td>
<td>glgP</td>
<td>glycogen phosphorylase</td>
<td>glycogen phosphorylase</td>
<td>glycogen biosynthesis</td>
<td>9.308</td>
<td>3.89E-12</td>
</tr>
<tr>
<td>BSU12610</td>
<td>xkdG</td>
<td>unknown</td>
<td>PBSX prophage</td>
<td>unknown</td>
<td>9.064</td>
<td>1.95E-12</td>
</tr>
<tr>
<td>BSU20390</td>
<td>yorG</td>
<td>unknown</td>
<td>unknown</td>
<td>unknown</td>
<td>8.992</td>
<td>2.41E-20</td>
</tr>
<tr>
<td>BSU37030</td>
<td>racA</td>
<td>cell division protein</td>
<td>cell division protein; attaches the chromosome to the cell pole at the onset of sporulation</td>
<td>unknown</td>
<td>8.733</td>
<td>3.65E-03</td>
</tr>
<tr>
<td>BSU19740</td>
<td>yodT</td>
<td>unknown</td>
<td>similar to adenosylmethionine-8-amino-7-oxononanoate aminotransferase</td>
<td>unknown</td>
<td>8.642</td>
<td>9.05E-15</td>
</tr>
<tr>
<td>BSU05610</td>
<td>vmlR</td>
<td>ABCF ATPase</td>
<td>F-type ATP-binding cassette protein, allosterically dissociates antibiotics (virginiamycin M, lincomycin) from their ribosomal binding sites</td>
<td>dissociation of antibiotics (virginiamycin M, lincomycin) from the ribosome</td>
<td>8.548</td>
<td>1.37E-03</td>
</tr>
<tr>
<td>BSU08140</td>
<td>yflD</td>
<td>unknown</td>
<td>unknown</td>
<td>unknown</td>
<td>8.276</td>
<td>3.43E-03</td>
</tr>
<tr>
<td>BSU10810</td>
<td>yisP</td>
<td>farnesyl diphosphate phosphatase</td>
<td>farnesyl diphosphate phosphatase, production of farnesol</td>
<td>control of KinC activity</td>
<td>7.180</td>
<td>2.31E-02</td>
</tr>
<tr>
<td>Locus Tag</td>
<td>Gene</td>
<td>Product</td>
<td>Description</td>
<td>Function</td>
<td>logFC</td>
<td>padj</td>
</tr>
<tr>
<td>-------------</td>
<td>------</td>
<td>----------------------------------</td>
<td>-----------------------------------------------------</td>
<td>---------------------------------------</td>
<td>--------</td>
<td>------------</td>
</tr>
<tr>
<td>BSU24150</td>
<td>mmgC</td>
<td>acyl-CoA dehydrogenase</td>
<td>acyl-CoA dehydrogenase</td>
<td>mother cell metabolism</td>
<td>-10.568</td>
<td>4.33E-16</td>
</tr>
<tr>
<td>BSU17120</td>
<td>pksE</td>
<td>unknown</td>
<td>involved in polyketide synthesis</td>
<td>polyketide synthesis</td>
<td>-10.092</td>
<td>8.22E-17</td>
</tr>
<tr>
<td>BSU27260</td>
<td>mccA</td>
<td>O-acetylserine-thiol-lyase</td>
<td>O-acetylserine-thiol-lyase</td>
<td>methionine-to-cysteine conversion</td>
<td>-9.847</td>
<td>2.02E-19</td>
</tr>
<tr>
<td>BSU39680</td>
<td>ill</td>
<td>inosose isomerase, converts 2KMI to 1-keto-D-chiro-inositol</td>
<td>inosose isomerase, converts 2KMI to 1-keto-D-chiro-inositol</td>
<td>myo-inositol catabolism</td>
<td>-9.397</td>
<td>7.91E-11</td>
</tr>
<tr>
<td>BSU01980</td>
<td>skfH</td>
<td>unknown</td>
<td>unknown</td>
<td>unknown</td>
<td>-9.307</td>
<td>7.60E-16</td>
</tr>
<tr>
<td>BSU29690</td>
<td>acuA</td>
<td>Gcn5-related N-acetyltransferase</td>
<td>protein acetylase for the control of AcsA activity</td>
<td>control of AcsA activity</td>
<td>-9.007</td>
<td>8.70E-15</td>
</tr>
<tr>
<td>BSU10430</td>
<td>yhdD</td>
<td>unknown</td>
<td>general stress protein, similar to alcohol dehydrogenase,</td>
<td>survival of salt and ethanol stresses</td>
<td>-8.959</td>
<td>1.00E-16</td>
</tr>
<tr>
<td>BSU22100</td>
<td>kdgA</td>
<td>2-dehydro-3-deoxy-phosphogluconate aldolase</td>
<td>2-dehydro-3-deoxy-phosphogluconate aldolase</td>
<td>utilization of galacturonic acid</td>
<td>-8.758</td>
<td>1.00E-03</td>
</tr>
<tr>
<td>BSU27930</td>
<td>spoOB</td>
<td>sporulation initiation phosphotransferase</td>
<td>sporulation initiation phosphotransferase of the phosphorelay</td>
<td>initiation of sporulation</td>
<td>-8.680</td>
<td>5.95E-04</td>
</tr>
<tr>
<td>BSU33490</td>
<td>cadA</td>
<td>cadmium transporting ATPase</td>
<td>cadmium transporting ATPase, resistance to cadmium</td>
<td>cadmium export</td>
<td>-8.546</td>
<td>8.87E-16</td>
</tr>
</tbody>
</table>
4.2.4.3 DPPs in SH2

Amongst all the growth conditions investigated in this study, SH2 produced the greatest number of DPPs in $\Delta spoVG$ (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 $\Delta spoVG$. 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 $\Delta spoVG$ 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 $\Delta spoVG$ (Figure 4.8). In SH5, 255 DPPs were decreased and 301 DPPs were increased in $\Delta spoVG$. The top 10 most decreased and increased DPPs in $\Delta spoVG$ 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 \( \Delta spoVG \)

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

**(A)**

<table>
<thead>
<tr>
<th>Locus Tag</th>
<th>Gene</th>
<th>Product</th>
<th>Description</th>
<th>Function</th>
<th>logFC</th>
<th>padj</th>
</tr>
</thead>
<tbody>
<tr>
<td>BSU02450</td>
<td>glnL</td>
<td>two-component response regulator</td>
<td>two-component response regulator, regulation of the ( glsA-glnT ) operon</td>
<td>regulation of the ( glsA-glnT ) operon</td>
<td>13.645</td>
<td>2.61E-03</td>
</tr>
<tr>
<td>BSU24580</td>
<td>yqhH</td>
<td>unknown</td>
<td>similar to SNF2 helicase</td>
<td>unknown</td>
<td>10.572</td>
<td>1.95E-16</td>
</tr>
<tr>
<td>BSU17140</td>
<td>pksF</td>
<td>unknown</td>
<td>involved in polyketide synthesis</td>
<td>polyketide synthesis</td>
<td>10.150</td>
<td>3.05E-05</td>
</tr>
<tr>
<td>BSU24430</td>
<td>spolIIA A</td>
<td>AAA protease</td>
<td>AAA protease, required for SigG activation</td>
<td>activation of SigG</td>
<td>10.010</td>
<td>3.49E-20</td>
</tr>
<tr>
<td>BSU19689</td>
<td>yokU</td>
<td>unknown</td>
<td>unknown</td>
<td>unknown</td>
<td>9.758</td>
<td>7.21E-18</td>
</tr>
<tr>
<td>BSU31500</td>
<td>yuxK</td>
<td>unknown</td>
<td>unknown</td>
<td>unknown</td>
<td>9.750</td>
<td>1.06E-04</td>
</tr>
<tr>
<td>BSU10430</td>
<td>yhxD</td>
<td>unknown</td>
<td>general stress protein, similar to alcohol dehydrogenase,</td>
<td>survival of salt and ethanol stresses</td>
<td>9.723</td>
<td>9.92E-18</td>
</tr>
<tr>
<td>BSU06230</td>
<td>ioIT</td>
<td>major transporter for inositol</td>
<td>major transporter for inositol</td>
<td>myo-inositol uptake</td>
<td>9.572</td>
<td>3.95E-04</td>
</tr>
<tr>
<td>BSU11500</td>
<td>spx</td>
<td>transcriptional regulator Spx</td>
<td>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</td>
<td>negative and positive regulator of many genes</td>
<td>9.488</td>
<td>2.56E-05</td>
</tr>
<tr>
<td>BSU19730</td>
<td>yodS</td>
<td>unknown</td>
<td>similar to butyrate-acetoacetate CoA-transferase</td>
<td>unknown</td>
<td>9.274</td>
<td>1.81E-10</td>
</tr>
<tr>
<td>Locus Tag</td>
<td>Gene</td>
<td>Product</td>
<td>Description</td>
<td>Function</td>
<td>logFC</td>
<td>padj</td>
</tr>
<tr>
<td>-----------</td>
<td>-------</td>
<td>--------------------------</td>
<td>---------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------</td>
<td>---------</td>
<td>-------</td>
</tr>
<tr>
<td>BSU25410</td>
<td>rpsU</td>
<td>ribosomal protein S21</td>
<td>ribosomal protein</td>
<td>translation</td>
<td>-12.511</td>
<td>7.40E-17</td>
</tr>
<tr>
<td>BSU22010</td>
<td>ypcP</td>
<td>5'→3' exonuclease</td>
<td>5'→3' exonuclease</td>
<td>protection of developing and dormant spores against UV DNA damage</td>
<td>-11.152</td>
<td>2.23E-15</td>
</tr>
<tr>
<td>BSU39430</td>
<td>deoR</td>
<td>transcriptional repressor</td>
<td>transcriptional repressor of the dra-nupC-pdp operon</td>
<td>regulation of deoxyribonucleotide utilization</td>
<td>-11.019</td>
<td>1.42E-05</td>
</tr>
<tr>
<td>BSU35410</td>
<td>flgK</td>
<td>flagellar hook-filament junction proteins</td>
<td>flagellar hook-filament junction proteins</td>
<td>motility and chemotaxis</td>
<td>-10.884</td>
<td>1.55E-04</td>
</tr>
<tr>
<td>BSU05810</td>
<td>gmuB</td>
<td>glucomannan-specific lichenan-specific Phosphotransferase system (PTS), EIIB component</td>
<td>glucomannan-specific permease of the PTS, EIIB of the PTS</td>
<td>glucomannan uptake and phosphorylation</td>
<td>-10.739</td>
<td>2.83E-20</td>
</tr>
<tr>
<td>BSU11660</td>
<td>tenl</td>
<td>thiazole tautomerase</td>
<td>thiazole tautomerase</td>
<td>biosynthesis of thiamine</td>
<td>-10.492</td>
<td>1.63E-18</td>
</tr>
<tr>
<td>BSU14150</td>
<td>ykuN</td>
<td>flavodoxin</td>
<td>flavodoxin, binds FMN, replaces ferredoxin under conditions of iron limitation, probably involved in electron transfer to nitric oxide synthase</td>
<td>electron transfer</td>
<td>-10.377</td>
<td>1.82E-12</td>
</tr>
<tr>
<td>BSU10310</td>
<td>yhfO</td>
<td>unknown</td>
<td>predicted acyltransferase</td>
<td>unknown</td>
<td>-10.285</td>
<td>3.13E-18</td>
</tr>
<tr>
<td>BSU04410</td>
<td>ydbB</td>
<td>unknown</td>
<td>unknown</td>
<td>unknown</td>
<td>-10.258</td>
<td>1.47E-04</td>
</tr>
<tr>
<td>BSU29580</td>
<td>thil</td>
<td>sulfuryl transferase</td>
<td>sulfuryl transferase, biosynthesis of 4-thiouridine in tRNA</td>
<td>biosynthesis of 4-thiouridine in tRNA</td>
<td>-10.170</td>
<td>8.71E-17</td>
</tr>
</tbody>
</table>
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 ∆spoVG across all the growth conditions (Table 4.6). Whereas the production of DhaS (aldehyde dehydrogenase (NAD)) was decreased in ∆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)

<table>
<thead>
<tr>
<th>Locus Tag</th>
<th>Gene</th>
<th>Product</th>
<th>Description</th>
<th>Function</th>
<th>logFC</th>
<th>padj</th>
</tr>
</thead>
<tbody>
<tr>
<td>BSU02450</td>
<td>glnL</td>
<td>two-component response regulator</td>
<td>two-component response regulator, regulation of the (glsA-glnT) operon</td>
<td>regulation of the (glsA-glnT) operon</td>
<td>11.905</td>
<td>7.57E-03</td>
</tr>
<tr>
<td>BSU17080</td>
<td>pksA</td>
<td>transcriptional regulator (TetR family)</td>
<td>transcriptional regulator (TetR family), not involved in the regulation of the pks operon!</td>
<td>unknown</td>
<td>10.820</td>
<td>5.75E-20</td>
</tr>
<tr>
<td>BSU35030</td>
<td>nagR</td>
<td>transcriptional regulator (GntR family)</td>
<td>transcriptional regulator (GntR family)</td>
<td>regulator of the (nagA-nagB-nagR) operon</td>
<td>9.832</td>
<td>7.15E-18</td>
</tr>
<tr>
<td>BSU02290</td>
<td>psd</td>
<td>phosphatidylserine decarboxylase</td>
<td>phosphatidylserine decarboxylase</td>
<td>biosynthesis of phospholipids</td>
<td>9.557</td>
<td>2.66E-04</td>
</tr>
<tr>
<td>BSU34700</td>
<td>psdA</td>
<td>ABC transporter (ATP-binding protein)</td>
<td>ABC transporter (ATP-binding protein) for the export of lipid II-binding lantibiotics, such as nisin and gallidermin</td>
<td>export of toxic peptides</td>
<td>9.390</td>
<td>8.06E-05</td>
</tr>
<tr>
<td>BSU22510</td>
<td>yjc</td>
<td>unknown</td>
<td>unknown</td>
<td>unknown</td>
<td>9.331</td>
<td>1.05E-04</td>
</tr>
<tr>
<td>BSU24730</td>
<td>comGA</td>
<td>traffic ATPase</td>
<td>late competence protein, traffic ATPase, binds and transports transforming DNA, required for growth arrest of competent cells</td>
<td>genetic competence</td>
<td>9.135</td>
<td>1.83E-25</td>
</tr>
<tr>
<td>BSU34800</td>
<td>cwo</td>
<td>endopeptidase-type autolysin</td>
<td>D,L-endopeptidase-type autolysin, primary autolytic pathway for cell elongation</td>
<td>cell wall synthesis, cell elongation</td>
<td>8.641</td>
<td>4.13E-11</td>
</tr>
<tr>
<td>BSU06970</td>
<td>rhlL</td>
<td>rhamnose oligosaccharide ABC transporter</td>
<td>ABC transporter (sugar-binding protein) for rhamnose oligosaccharides</td>
<td>uptake of rhamnose oligosaccharides</td>
<td>8.449</td>
<td>8.16E-05</td>
</tr>
<tr>
<td>Locus Tag</td>
<td>Gene</td>
<td>Product</td>
<td>Description</td>
<td>Function</td>
<td>logFC</td>
<td>padj</td>
</tr>
<tr>
<td>------------</td>
<td>------</td>
<td>----------------------------------</td>
<td>----------------------------------------------------</td>
<td>-----------------------------------</td>
<td>-------</td>
<td>------------</td>
</tr>
<tr>
<td>BSU25410</td>
<td>rpsU</td>
<td>ribosomal protein S21</td>
<td>ribosomal protein</td>
<td>translation</td>
<td>-11.298</td>
<td>1.91E-16</td>
</tr>
<tr>
<td>BSU35750</td>
<td>tagA</td>
<td>UDP-N-acetyl-D-mannosamine transferase</td>
<td>UDP-N-acetyl-D-mannosamine transferase</td>
<td>biosynthesis of teichoic acid</td>
<td>-10.590</td>
<td>2.50E-16</td>
</tr>
<tr>
<td>BSU02600</td>
<td>cwlJ</td>
<td>spore coat protein, cell wall hydrolase</td>
<td>spore coat protein, cell wall hydrolase</td>
<td>spore germination</td>
<td>-10.543</td>
<td>2.51E-18</td>
</tr>
<tr>
<td>BSU26950</td>
<td>yraG</td>
<td>unknown</td>
<td>forespore-specific sporulation protein, similar to spore coat protein</td>
<td>unknown</td>
<td>-10.284</td>
<td>3.12E-21</td>
</tr>
<tr>
<td>BSU18410</td>
<td>ggt</td>
<td>gamma-glutamyltransferase</td>
<td>gamma-glutamyltransferase</td>
<td>degradation of poly-glutamate capsules</td>
<td>-10.162</td>
<td>7.21E-18</td>
</tr>
<tr>
<td>BSU39430</td>
<td>deoR</td>
<td>transcriptional repressor</td>
<td>transcriptional repressor of the dra-nupC-pdp operon</td>
<td>regulation of deoxyribonucleotide utilization</td>
<td>-10.016</td>
<td>3.43E-05</td>
</tr>
<tr>
<td>BSU36110</td>
<td>chrS</td>
<td>transcriptional repressor (Lrp family)</td>
<td>transcriptional repressor (Lrp family) of the chrS-ywrB-ywrA operon</td>
<td>regulation of chromate export</td>
<td>-9.775</td>
<td>1.49E-21</td>
</tr>
<tr>
<td>BSU05810</td>
<td>gmuB</td>
<td>glucomannan-specific lichenan-specific PTS, EIIB component</td>
<td>glucomannan-specific permease of the PTS, EIIB of the PTS</td>
<td>glucomannan uptake and phosphorylation</td>
<td>-9.722</td>
<td>3.64E-20</td>
</tr>
<tr>
<td>BSU17990</td>
<td>sspO</td>
<td>small acid-soluble spore protein (minor)</td>
<td>small acid-soluble spore protein (minor)</td>
<td>protection of spore DNA</td>
<td>-9.684</td>
<td>1.83E-25</td>
</tr>
<tr>
<td>BSU10310</td>
<td>yhfO</td>
<td>unknown</td>
<td>predicted acyltransferase</td>
<td>unknown</td>
<td>-9.589</td>
<td>3.46E-18</td>
</tr>
</tbody>
</table>
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
\( \Delta \text{spoVG} \) 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 \( \Delta \text{spoVG} \). 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 \( \Delta \text{spoVG} \) 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 \emph{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 \( \text{padj} < 0.05 \) 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, \emph{Biosynthesis of antibacterial compounds} and \emph{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.

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

<table>
<thead>
<tr>
<th>Locus tag</th>
<th>Gene</th>
<th>Product</th>
<th>Function</th>
<th>SH2</th>
<th>SH5</th>
</tr>
</thead>
<tbody>
<tr>
<td>BSU17330</td>
<td>miaA</td>
<td>tRNA isopentenylpyrophosphate transferase</td>
<td>tRNA modification</td>
<td>6.796</td>
<td>1.75E-02</td>
</tr>
<tr>
<td>BSU03200</td>
<td>putB</td>
<td>proline dehydrogenase (L-proline, NAD)</td>
<td>proline utilization</td>
<td>1.074</td>
<td>1.23E-03</td>
</tr>
<tr>
<td>BSU00150</td>
<td>dgk</td>
<td>deoxyguanosine kinase</td>
<td>purine salvage and interconversion</td>
<td>0.874</td>
<td>1.68E-02</td>
</tr>
<tr>
<td>BSU05390</td>
<td>ydfF</td>
<td>unknown</td>
<td>unknown</td>
<td>0.758</td>
<td>4.50E-02</td>
</tr>
<tr>
<td>BSU09710</td>
<td>bmrC</td>
<td>multidrug ABC transporter (ATP-binding protein)</td>
<td>multiple antibiotic resistance</td>
<td>0.516</td>
<td>3.65E-02</td>
</tr>
<tr>
<td>BSU38310</td>
<td>ywbI</td>
<td>transcriptional regulator (LysR family)</td>
<td>control of expression of the ywbH-ywbG operon</td>
<td>0.427</td>
<td>2.50E-02</td>
</tr>
<tr>
<td>BSU40950</td>
<td>yyaC</td>
<td>sporulation protein</td>
<td>unknown</td>
<td>-7.747</td>
<td>7.40E-17</td>
</tr>
<tr>
<td>BSU06370</td>
<td>pbuG</td>
<td>hypoxanthin/ guanine permease</td>
<td>hypoxanthine and guanine uptake</td>
<td>-9.663</td>
<td>2.83E-20</td>
</tr>
</tbody>
</table>
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 ΔspoVG

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

### 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 ΔspoVG (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 ΔspoVG 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-nogueira *et al.*, 1997). *B. subtilis* can grow on L-arabinose as the sole carbon and energy source (Mota, Tavares and Sá-Nogueira, 1999; Mota, Sarmento and Sa-Nogueira, 2001; Zhang *et al.*, 2016). In this study, AraA decreased by logFC ~ -8.2 in ΔspoVG 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 ΔspoVG in LB (Yoshida *et al.*, 1997, 1999, 2008).
Table 4.8: Significantly enriched Category 3 (SubtiWiki) gene sets identified in ΔspoVG from the proteomic data

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

<table>
<thead>
<tr>
<th>Cat3 Gene Set</th>
<th>Total proteins identified in the gene set</th>
<th>Enrichment of the gene set (padj)</th>
<th>Proteins increased</th>
<th>Proteins decreased</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>LB</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Utilization of specific carbon sources</td>
<td>154</td>
<td>2.96E-02</td>
<td>49</td>
<td>69</td>
</tr>
<tr>
<td>Heat shock proteins</td>
<td>25</td>
<td>2.96E-02</td>
<td>21</td>
<td>4</td>
</tr>
<tr>
<td>PBSX prophage</td>
<td>7</td>
<td>2.96E-02</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td><strong>M9</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Miscellaneous metabolic pathways</td>
<td>58</td>
<td>5.55E-03</td>
<td>14</td>
<td>35</td>
</tr>
<tr>
<td>Biosynthesis of antibacterial compounds</td>
<td>34</td>
<td>5.55E-03</td>
<td>2</td>
<td>29</td>
</tr>
<tr>
<td>General stress proteins (controlled by SigB)</td>
<td>94</td>
<td>4.07E-02</td>
<td>20</td>
<td>60</td>
</tr>
<tr>
<td><strong>SH2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Translation</td>
<td>183</td>
<td>1.11E-02</td>
<td>100</td>
<td>77</td>
</tr>
<tr>
<td><strong>SH5</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sporulation proteins</td>
<td>260</td>
<td>1.11E-02</td>
<td>69</td>
<td>186</td>
</tr>
</tbody>
</table>
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 ∆*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 ∆*spoVG*. PksE and SkfH were also identified as some of the most decreased proteins in ∆*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 \textit{pks}, \textit{skf} and \textit{ppsABCDE} operon proteins were in agreement with the previous observations from the transcriptomic data where the expressions of the \textit{pks}, \textit{skf} and \textit{pps} genes were down-regulated in \textit{ΔspoVG} in M9 (Chapter 3). The similar direction of change in these proteins indicated a role of SpoVG in the production of antimicrobial compounds in \textit{B. subtilis}. In addition to the antimicrobial compounds, decrease in the levels of toxin proteins such as the \textit{SdpABC} (logFC -3.0, -7.2, -5.2 respectively) and the \textit{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 \textit{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 ΔspoVG in M9

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

<table>
<thead>
<tr>
<th>Locus Tag</th>
<th>Gene</th>
<th>Product</th>
<th>Description</th>
<th>Function</th>
<th>logFC</th>
<th>padj</th>
</tr>
</thead>
<tbody>
<tr>
<td>BSU01980</td>
<td><em>skfH</em></td>
<td>unknown</td>
<td>unknown</td>
<td>unknown</td>
<td>-9.307</td>
<td>7.60E-16</td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>skf</em> operon</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BSU17100</td>
<td><em>pksC</em></td>
<td>bacillaene synthase trans-acting acyltransferase</td>
<td>bacillaene synthase trans-acting acyltransferase</td>
<td>polyketide (bacillaene) synthesis</td>
<td>-1.634</td>
<td>6.45E-04</td>
</tr>
<tr>
<td>BSU17120</td>
<td><em>pksE</em></td>
<td>unknown</td>
<td>involved in polyketide synthesis</td>
<td>polyketide synthesis</td>
<td>-10.092</td>
<td>8.22E-17</td>
</tr>
<tr>
<td>BSU17170</td>
<td><em>pksI</em></td>
<td>unknown</td>
<td>involved in polyketide synthesis</td>
<td>polyketide synthesis</td>
<td>-8.262</td>
<td>5.94E-03</td>
</tr>
<tr>
<td>BSU17190</td>
<td><em>pksL</em></td>
<td>polyketide synthase of type I</td>
<td>polyketide synthase of type I</td>
<td>polyketide synthesis</td>
<td>-2.467</td>
<td>2.24E-05</td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>ppsABCDE</em> operon</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BSU18320</td>
<td><em>ppsC</em></td>
<td>plipastatin synthetase</td>
<td>plipastatin synthetase</td>
<td>production of the antibacterial compound plipastatin</td>
<td>-2.126</td>
<td>2.51E-05</td>
</tr>
<tr>
<td>BSU18340</td>
<td><em>ppsA</em></td>
<td>plipastatin synthetase</td>
<td>plipastatin synthetase</td>
<td>production of the antibacterial compound plipastatin</td>
<td>-2.044</td>
<td>1.07E-02</td>
</tr>
</tbody>
</table>
In addition to the changed levels of antimicrobial compounds, the proteomic analysis of $\Delta$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 $\Delta$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).
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 ΔspoVG in M9 (Chapter 3). It was also identified that the expression of sigB significantly increased in ΔspoVG in M9 (Chapter 3). However, in this proteomic study of ΔspoVG 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 ΔspoVG.

The decreased levels of several GSPs in this proteomic study compared to the up-regulated GSR genes in the transcriptomic analysis of ΔspoVG 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 ΔspoVG in SH2 (Figure 4.10). In this gene set, a total of 183 proteins were detected most of which were increased in ΔspoVG (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 Kawamura, 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 correlates 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 ΔspoVG 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 ΔspoVG during sporulation conditions maybe helpful in highlighting its role in translation.
Figure 4.11: Biological processes affected in SH2 in ΔspoVG

Voronoi treemap of all the (A) increased (B) decreased proteins in ΔspoVG. The area of the polygon is proportional to the absolute logFC of the DPPs.
Table 4.10: Proteomic profile of decreased ribosomal proteins in ΔspoVG in SH2

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

<table>
<thead>
<tr>
<th>Locus Tag</th>
<th>Gene</th>
<th>Product</th>
<th>Description</th>
<th>Function</th>
<th>logFC</th>
<th>padj</th>
</tr>
</thead>
<tbody>
<tr>
<td>BSU25410</td>
<td>rpsU</td>
<td>ribosomal protein S21</td>
<td>ribosomal protein</td>
<td>translation</td>
<td>-12.511</td>
<td>7.40E-17</td>
</tr>
<tr>
<td>BSU27940</td>
<td>rpmA</td>
<td>ribosomal protein L27 (BL24)</td>
<td>ribosomal protein</td>
<td>translation</td>
<td>-8.161</td>
<td>4.68E-03</td>
</tr>
<tr>
<td>BSU24900</td>
<td>rpmGA</td>
<td>ribosomal protein L33a</td>
<td>ribosomal protein L33a</td>
<td>translation</td>
<td>-7.970</td>
<td>5.25E-10</td>
</tr>
<tr>
<td>BSU01400</td>
<td>rpmJ</td>
<td>ribosomal protein L36 (ribosomal protein B)</td>
<td>ribosomal protein</td>
<td>translation</td>
<td>-7.144</td>
<td>3.25E-07</td>
</tr>
<tr>
<td>BSU40500</td>
<td>rplL</td>
<td>ribosomal protein L9</td>
<td>ribosomal protein</td>
<td>translation</td>
<td>-2.569</td>
<td>7.71E-09</td>
</tr>
<tr>
<td>BSU01240</td>
<td>rpmC</td>
<td>ribosomal protein L29</td>
<td>ribosomal protein</td>
<td>translation</td>
<td>-2.409</td>
<td>3.47E-07</td>
</tr>
<tr>
<td>BSU01200</td>
<td>rpsS</td>
<td>ribosomal protein S19 (BS19)</td>
<td>ribosomal protein</td>
<td>translation</td>
<td>-2.303</td>
<td>4.67E-05</td>
</tr>
<tr>
<td>BSU29660</td>
<td>rpsD</td>
<td>ribosomal protein S4 (BS4)</td>
<td>ribosomal protein</td>
<td>translation</td>
<td>-2.083</td>
<td>2.09E-05</td>
</tr>
<tr>
<td>BSU01190</td>
<td>rplB</td>
<td>ribosomal protein L2 (BL2)</td>
<td>ribosomal protein L2, required for assembly of RplP into the 50S subunit of the ribosome</td>
<td>translation</td>
<td>-2.044</td>
<td>1.04E-05</td>
</tr>
<tr>
<td>BSU01230</td>
<td>rplP</td>
<td>ribosomal protein L16</td>
<td>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</td>
<td>translation</td>
<td>-1.998</td>
<td>2.16E-05</td>
</tr>
</tbody>
</table>
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 ΔspoVG suggested a potential role of SpoVG in the later stages of sporulation.
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 ΔspoVG. In the late stages of sporulation (SH5), a decreased level of Spo0B would have been expected. Therefore, the increased level of Spo0B in ΔspoVG 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 ΔspoVG. 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 ΔspoVG. The null mutation of spoVS when combined with the double null mutant ΔspoVGΔspollB alleviates the sporulation blockage due to the ΔspoVGΔspollB 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 spollB (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), SpolIIA (dissolution of the septal cell wall) (logFC = -9.34); SpolIIAA (logFC = 10.01) and SpolIIAG (activation of SigG) (logFC = -0.16), SpolIIJ (membrane insertion of proteins and protein secretion) (logFC = -9.36), SpolIIA (required for dissolution of the septal cell wall) (logFC = -0.76) and SpolIIJ (forespore encasement by the spore coat) (logFC = 1.64). SpolID is known to form a multi-protein complex with SpolIM and SpolIP (MPD complex) which 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 SpolIM is dependent on SpolIB (Chastanet and Losick, 2007).

SpolIIA and SpolIIAG are both encoded from the spolIIA operon (an operon with 7 to 8 genes) (Illing and Errington, 1991). SpolIIA is a protease and is required for the activation of sigma factor SigG (Camp and Losick, 2008). Whereas, SpolIIAG is a component of the SpolIIA-SpolIIQ type III secretion system that resides in the forespore membrane (Illing and Errington, 1991; Camp and Losick, 2008). SpolIIQ is a forespore protein and is a component of the SpolIIAH-SpolIIQ 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 SpolIE (Camp and Losick, 2008; Campo, Marquis and Rudner, 2008; Levdiskov et al., 2012). SpolIIJ is a vegetative protein and is encoded from yidC1 which form an operon rpmH-mpA-yidC1-jag or yidC1-jag. SpolIIJ 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 SpolIIAH-SpolIIQ 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 ΔspoVG, 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 ΔspoVG.
Figure 4.13: The KEGG enriched pathways due to the DPPs in $\Delta$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 $\Delta$spoVG) and the amino acid racemase (increased in $\Delta$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 \( \Delta \text{spoVG} \) in all the growth conditions

<table>
<thead>
<tr>
<th>KEGG pathways</th>
<th>padj</th>
<th>Total proteins in pathway</th>
<th>Proteins enriched</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>LB</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Histidine metabolism</td>
<td>3.05E-02</td>
<td>18</td>
<td>3</td>
</tr>
<tr>
<td>Nonribosomal peptide structures</td>
<td>4.52E-02</td>
<td>9</td>
<td>2</td>
</tr>
<tr>
<td><strong>M9</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nonribosomal peptide structures</td>
<td>4.62E-04</td>
<td>9</td>
<td>4</td>
</tr>
<tr>
<td><strong>SH2</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ribosome</td>
<td>1.25E-03</td>
<td>57</td>
<td>28</td>
</tr>
<tr>
<td>RNA polymerase</td>
<td>2.88E-02</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td><strong>SH5</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Valine, leucine and isoleucine degradation</td>
<td>2.20E-03</td>
<td>23</td>
<td>12</td>
</tr>
<tr>
<td>Arginine and proline metabolism</td>
<td>1.64E-02</td>
<td>25</td>
<td>11</td>
</tr>
<tr>
<td>Taurine and hypotaurine metabolism</td>
<td>1.94E-02</td>
<td>8</td>
<td>5</td>
</tr>
</tbody>
</table>
Table 4.12: Nonribosomal proteins affected in ΔspoVG 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.

<table>
<thead>
<tr>
<th>Locus Tag</th>
<th>Gene</th>
<th>Product</th>
<th>Description</th>
<th>Function</th>
<th>logFC</th>
<th>padj</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>LB</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BSU18340</td>
<td>ppsA</td>
<td>plipastatin synthetase</td>
<td>plipastatin synthetase</td>
<td>production of the antibacterial compound plipastatin</td>
<td>1.64</td>
<td>4.70E-02</td>
</tr>
<tr>
<td>BSU34430</td>
<td>racX</td>
<td>amino acid racemase</td>
<td>amino acid racemase, production of non-canonical D-amino acids</td>
<td>production of non-canonical D-amino acids</td>
<td>-5.57</td>
<td>2.62E-02</td>
</tr>
<tr>
<td><strong>M9</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BSU18320</td>
<td>ppsC</td>
<td>plipastatin synthetase</td>
<td>plipastatin synthetase</td>
<td>production of the antibacterial compound plipastatin</td>
<td>-2.13</td>
<td>2.51E-05</td>
</tr>
<tr>
<td>BSU18340</td>
<td>ppsA</td>
<td>plipastatin synthetase</td>
<td>plipastatin synthetase</td>
<td>production of the antibacterial compound plipastatin</td>
<td>-2.04</td>
<td>1.07E-02</td>
</tr>
</tbody>
</table>
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 \( \Delta spoVG \) 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 \textit{B. subtilis} which belongs to the core metabolic pathways of the cell (Fisher, 1991). In \textit{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 \textit{spoVG} is repressed (Zuber and Losick, 1987) and the nitrogen metabolising genes is decreased (Boylan \textit{et al.}, 1988; Hu \textit{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 \textit{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 \textit{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 \( \Delta spoVG \) suggested a role of SpoVG in the amino acid and nitrogen metabolism in \textit{B. subtilis}. Although the mechanisms behind these changes are yet to be understood.

\subsection*{4.2.7 Affected Regulons in \( \Delta spoVG \)}

The results obtained so far strongly suggested that several regulons are perturbed in the absence of SpoVG in \textit{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 ΔspoVG (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 ∆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 genes.
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 ∆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 ∆spoVG (Figure 4.14 (D)). This suggested a role of SpoVG in the resistance to lysozyme in SH5. Previous study by Burke and Portnoy, (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 ∆spoVG in *B. subtilis* in LB, M9, SH2 and SH5 growth conditions. The results showed that 5 studied. Maximal proteins were changed in SH2, where almost 29.4% of the total proteome was affected in ∆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 \textit{B. subtilis}. Significant changes in the proteomes of \textit{\Delta 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, \textit{General stress proteins (controlled by SigB)}, Translation, \textit{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 \textit{\Delta 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 \textit{\Delta spoVG}. Investigations in other Gram-positive bacteria such as \textit{L. monocytogenes} and \textit{S. aureus} have also found SpoVG to have a role in the central carbon metabolism (Bischoff \textit{et al.}, 2016; Burke and Portnoy, 2016). A similar observation from this study, suggested participation of SpoVG in the carbon metabolism of \textit{B. subtilis} as well.

Major changes were also observed in several sigma factor regulons in \textit{\Delta 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 \textit{\Delta 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 \textit{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 ∆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 ∆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 ΔspoVG 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 ($\lambda = 254$ 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-$\gamma$ATP 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*](image)

(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-$\gamma$ATP 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 $\Delta$*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-$\gamma$ATP 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 $\sim$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

\textit{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 \textit{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 \textit{srfAA}, \textit{srfAB}, \textit{srfAC} RNAs (Table 5.1). These transcripts belong to the \textit{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 \textit{B. subtilis} (Zhao \textit{et al.}, 2017). In addition to these three RNAs, \textit{comS} and \textit{srfAD} transcript (other members of \textit{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.

<table>
<thead>
<tr>
<th>Locus tag</th>
<th>Gene</th>
<th>Average read counts (bp) (WT)</th>
<th>Average read counts (bp) (SpoVG-HTF)</th>
<th>Product</th>
<th>Description</th>
<th>Function</th>
<th>Gene length (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BSU03490</td>
<td>srfAB</td>
<td>1,681</td>
<td>1,32,007</td>
<td>surfactin synthetase / competence</td>
<td>surfactin synthetase / competence</td>
<td>antibiotic synthesis</td>
<td>10,761</td>
</tr>
<tr>
<td>BSU03480</td>
<td>srfAA</td>
<td>1,205</td>
<td>1,09,515</td>
<td>surfactin synthetase / competence</td>
<td>surfactin synthetase / competence</td>
<td>antibiotic synthesis</td>
<td>10,764</td>
</tr>
<tr>
<td>BSU03510</td>
<td>srfAC</td>
<td>807</td>
<td>52,629</td>
<td>surfactin synthetase / competence</td>
<td>surfactin synthetase / competence</td>
<td>antibiotic synthesis</td>
<td>3,822</td>
</tr>
<tr>
<td>BSU39230</td>
<td>wapA</td>
<td>278</td>
<td>32,226</td>
<td>cell wall-associated protein precursor</td>
<td>cell wall-associated protein precursor, contact-dependent growth inhibition protein</td>
<td>intercellular competition</td>
<td>7,002</td>
</tr>
<tr>
<td>BSU11240</td>
<td>carB</td>
<td>703</td>
<td>18,339</td>
<td>carbamoyl-phosphate transferase-arginine (subunit B)</td>
<td>carbamoyl-phosphate transferase-arginine (subunit B)</td>
<td>biosynthesis of arginine</td>
<td>3,090</td>
</tr>
<tr>
<td>BSU01080</td>
<td>rpoC</td>
<td>136</td>
<td>17,155</td>
<td>RNA polymerase, β subunit</td>
<td>RNA polymerase, β subunit</td>
<td>transcription</td>
<td>3,597</td>
</tr>
<tr>
<td>BSU17210</td>
<td>pksN</td>
<td>58</td>
<td>15,545</td>
<td>polyketide synthase of type I</td>
<td>polyketide synthase of type I</td>
<td>polyketide synthesis</td>
<td>16,464</td>
</tr>
<tr>
<td>BSU12430</td>
<td>rapA</td>
<td>262</td>
<td>13,810</td>
<td>response regulator aspartate phosphatase</td>
<td>response regulator aspartate phosphatase, dephosphorylates Spo0F-P, control of the phosphorelay</td>
<td>control of sporulation initiation</td>
<td>1,134</td>
</tr>
<tr>
<td>-----------</td>
<td>-------</td>
<td>------</td>
<td>--------</td>
<td>------------------------------------------</td>
<td>---------------------------------------------------------------------------------</td>
<td>---------------------------------</td>
<td>--------</td>
</tr>
<tr>
<td>BSU14860</td>
<td>pycA</td>
<td>107</td>
<td>13,547</td>
<td>pyruvate carboxylase</td>
<td>pyruvate carboxylase</td>
<td>replenishment of the oxaloacetate pool</td>
<td>3,444</td>
</tr>
<tr>
<td>BSU03210</td>
<td>putC</td>
<td>165</td>
<td>13,505</td>
<td>1-pyrroline-5-carboxylate dehydrogenase</td>
<td>1-pyrroline-5-carboxylate dehydrogenase</td>
<td>proline utilization</td>
<td>1,545</td>
</tr>
<tr>
<td>BSU35360</td>
<td>hag</td>
<td>134</td>
<td>13,069</td>
<td>flagellin protein</td>
<td>flagellin protein, about 20,000 subunits make up one flagellum</td>
<td>motility and chemotaxis</td>
<td>912</td>
</tr>
<tr>
<td>BSU18310</td>
<td>ppsD</td>
<td>50</td>
<td>12,977</td>
<td>plipastatin synthetase</td>
<td>plipastatin synthetase</td>
<td>production of the antibacterial compound plipastatin</td>
<td>10,809</td>
</tr>
<tr>
<td>BSU17180</td>
<td>pksJ</td>
<td>49</td>
<td>12,804</td>
<td>polyketide synthase</td>
<td>polyketide synthase</td>
<td>polyketide synthesis</td>
<td>15,135</td>
</tr>
<tr>
<td>BSU18000</td>
<td>citB</td>
<td>122</td>
<td>12,300</td>
<td>aconitase, trigger enzyme</td>
<td>aconitase, trigger enzyme</td>
<td>TCA cycle</td>
<td>2,727</td>
</tr>
<tr>
<td>BSU00500</td>
<td>gcaD</td>
<td>107</td>
<td>11,616</td>
<td>bifunctional N-acetylglucosamine-1-phosphate</td>
<td>bifunctional N-acetylglucosamine-1-phosphate uridytransferase or glucosamine-1-phosphate acetyltransferase</td>
<td>cell wall metabolism</td>
<td>1,368</td>
</tr>
<tr>
<td>BSU17220</td>
<td>pksR</td>
<td>33</td>
<td>11,324</td>
<td>polyketide synthase</td>
<td>polyketide synthase</td>
<td>polyketide synthesis</td>
<td>7,629</td>
</tr>
<tr>
<td>Strain</td>
<td>Gene</td>
<td>Start</td>
<td>Length</td>
<td>Description</td>
<td>Function</td>
<td>IC50</td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>------</td>
<td>-------</td>
<td>--------</td>
<td>----------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------</td>
<td>-------</td>
<td></td>
</tr>
<tr>
<td>BSU11430</td>
<td>oppA</td>
<td>123</td>
<td>11,310</td>
<td>oligopeptide ABC transporter (binding protein)</td>
<td>initiation of sporulation, competence development</td>
<td>1,635</td>
<td></td>
</tr>
<tr>
<td>BSU37370</td>
<td>albA</td>
<td>90</td>
<td>11,054</td>
<td>radical S-adenosylmethionine enzyme</td>
<td>antilisterial bacteriocin (subtilosin) production</td>
<td>1,344</td>
<td></td>
</tr>
<tr>
<td>BSU17190</td>
<td>pksL</td>
<td>35</td>
<td>10,708</td>
<td>polyketide synthase of type I</td>
<td>polyketide synthase</td>
<td>13,614</td>
<td></td>
</tr>
<tr>
<td>BSU37280</td>
<td>narG</td>
<td>44</td>
<td>10,488</td>
<td>nitrate reductase (alpha subunit)</td>
<td>nitrate respiration, nitrogen assimilation</td>
<td>3,684</td>
<td></td>
</tr>
</tbody>
</table>
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 ∆*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 $\Delta$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)
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 ΔspoVG 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 ΔspoVG 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 ∆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 (MS²) 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.
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 byproduct 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 log2 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.

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

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 ∆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 ∆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 ΔspoVG 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
ΔspoVG 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 ∆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 \( \Delta spoVG \) 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 \( \Delta spoVG \) 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 \( \Delta spoVG \) 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 \( \Delta spoVG \) 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, tricarboxylic acid (TCA) cycle or other modes of carbon metabolism.

The proteomics data of \( \Delta spoVG \) 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 ΔspoVG 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 ΔspoVG 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 ΔspoVG

7.2.1.1 The ΔspoVG cells swarm less on M9 agar

Swarming motility of the ΔspoVG 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 ΔspoVG mutant produced a channel like pattern on the surface of the agar. The complemented strain (ΔspoVG::spoVG) was able to restore the swarming deficiency.

![Image of swarming motility](image)

Figure 7.1: The ΔspoVG (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, ΔspoVG and complemented ΔspoVG::spoVG after overnight incubation.
7.2.1.2 $\Delta$spoVG has restricted growth on M9 swarming plates and lacks pigment formation

The $\Delta$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 $\Delta$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 $\Delta$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 $\Delta$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 $\Delta$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 $\Delta$spoVG in M9 (Chapter 3). However, the proteins products were not differentially produced in the proteomics data. The gene set enrichment analysis (GSEA) in $\Delta$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 ΔspoVG (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, ΔspoVG and ΔspoVG::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 ΔspoVG plates.
7.2.1.3 Biofilms of ΔspoVG are smaller and have more “wrinkles” on MSgg agar

Since the ΔspoVG 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, ΔspoVG 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 ΔspoVG 3610 on MSgg agar after 5 days of incubation (Figure 7.3). The restricted biofilms of ΔspoVG 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 ΔspoVG 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](Image) (Picture courtesy: Nicola Stanley-Wall Lab) The biofilms of the ΔspoVG (3610) are smaller and have more “wrinkles” on MSgg agar

The 3610 strain of the WT, ΔspoVG and spoVG::spoVG were grown on MSgg and inoculated on MSgg agar (1.5%) (Amaouteli 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 ΔspoVG 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 ΔspoVG 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 ΔspoVG 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, ΔspoVG and the complement were set up on M9 agar and monitored for up to 16 hours.

The results revealed that the ΔspoVG 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 ΔspoVG 3610 in M9 (Figure 7.1). Also, the complemented strain was able to restore the WT phenotype.

![Figure 7.4: The swimming phenotype of ΔspoVG (168) is reduced on M9 agar](image)

The swimming phenotype of the 168 strains of the WT, ΔspoVG and ΔspoVG::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 ΔspoVG 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 ΔspoVG 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 ΔspoVG 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 ΔspoVG forms expanded biofilms and produces less pigment on LB agar

Since ΔspoVG 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 ΔspoVG 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 ΔspoVG 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 ΔspoVG.

The results from this experiment revealed that the ΔspoVG (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 ΔspoVG (3610) produced less brown pigment compared to the WT and the complemented strains (Figure 7.5 (B)). The enhanced biofilm formation of ΔspoVG 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 ΔspoVG was decreased in both the LB and MSgg. These opposite results of the biofilm formation in LB vs M9 in ΔspoVG 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) 

<table>
<thead>
<tr>
<th></th>
<th>Day 1</th>
<th>Day 2</th>
<th>Day 3</th>
<th>Day 5.5</th>
<th>Day 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT 3610</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\Delta spoVG)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\Delta spoVG::spoVG)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 7.5: The biofilms of ΔspoVG (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, ΔspoVG and complement ΔspoVG::spoVG on LBGM agar (1.5%) plates. (A) The top view of the plates showing the expanded biofilms of the ΔspoVG. The colonies of ΔspoVG 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 ΔspoVG. A noticeable change in colour in ΔspoVG can be observed after day 5.
7.2.2 Secretion in ∆spoVG

7.2.2.1 ∆spoVG 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 ∆spoVG is affected in LB, the ability of ∆spoVG to secrete amylase was tested by the iodine/starch assay using both the 3610 and the 168, strains of WT, ∆spoVG and the complement (∆spoVG::spoVG).

The results from this experiment showed that ∆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 ∆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 ∆spoVG would further enable to quantify these changes in the secretome quantitatively.
Figure 7.6: Secretion of α-amylase in ΔspoVG (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 ΔspoVG::spoVG (ΔspoVG<sup>C</sup>) 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 ∆spoVG in LB

In addition to the α-amylase starch assay, the secreted proteins were precipitated and were analysed by SDS-PAGE. The ∆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 ∆spoVG and also the previous secretion experiments. The complemented strain restored the phenotype of the WT.

![Figure 7.7: Secretion of total proteins in ∆spoVG (3610) is reduced in LB](image)

The 3610 strains of the WT, ∆spoVG and ∆spoVG::spoVG were grown in LB till OD₆₀₀ 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 ΔspoVG in Alternate Carbon Sources

To test the ability of ΔspoVG 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 ΔspoVG 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 ΔspoVG. Milder changes in growth were observed with fructose (0.3%), xylose (0.3%) and MI (10 mM). In xylose (0.3%), the growth of ΔspoVG was decreased between 10-14 hours time point compared to the WT and complemented strain. Also, all of the growth differences in ΔspoVG were observed post the 6 hours time point. Of all the conditions, a maximal change in the growth of ΔspoVG was observed in galactose (Figure 7.8). To better understand these behaviours of ΔspoVG, further experiments would be needed.
Figure 7.8: Growth of ΔspoVG (168) is affected in Paris media in response to various carbon sources

Each growth curve shows the growth of ΔspoVG in Paris media supplemented with different sources of carbon (indicated on each curve) and grown in a microtiter well for 14 hours. The WT, ΔspoVG and the complement strain ΔspoVG::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 ∆spoVG when it was tested for its capacity to form biofilm on LBGM agar. When tested for the capacity to form biofilms, the ∆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 ∆spoVG mutant. It was shown that, there was an overall decrease in the ability of ∆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 ∆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 ΔspoVG 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 ΔspoVG 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 RBP with condition-dependent roles, or existence of sRNA-mRNA base-pairing without the need of RBP.

Recent investigations to explore the roles of RBP 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 ΔspoVG

Chapter 3 presents the first global transcriptomic analysis of the ΔspoVG 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 ΔspoVG, 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 \( \Delta \text{spoVG} \) 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 \( \Delta \text{spoVG} \). These observations indicated a role of spoVG in the synthesis of antimicrobial peptides in \( B. \text{subtilis} \). 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 \( \Delta \text{spoVG} \) 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 ΔspoVG. The SigI regulon in M9 was the most affected regulon where ~44% genes (all down-regulated) were affected due to ΔspoVG. The genes of SigI sigma factors are involved in iron metabolism (Ramanuk 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 ΔspoVG. 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 ΔspoVG

Further, in order to investigate the changes in the proteome levels due to deletion, the first global proteomic quantification of the ΔspoVG 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 ΔspoVG 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 ΔspoVG 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 data 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 ΔspoVG in SH2. Additionally, Ribosome was also a significantly enriched KEGG pathway in the transcriptomic analysis of ΔspoVG. 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 ΔspoVG (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 ΔspoVG (Chapter 3 and 4). For example, in both the transcriptome and the proteome of ΔspoVG, 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 ∆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 ∆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 ΔspoVG. 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 multifunctional 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 \textit{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 \( \Delta\text{spoVG} \)

Finally, the transcriptome (Chapter 3) and proteome (Chapter 4) profiles obtained from \( \text{spoVG} \) 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 \( \Delta\text{spoVG} \) 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 \( \Delta\text{spoVG} \), the phenotypes of \( \Delta\text{spoVG} \) 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 \textit{B. subtilis}. The investigations revealed that the \( \Delta\text{spoVG} \) has a reduced swimming and swarming motility in M9. Also, it has a retarded capacity to form biofilms on MSgg agar. The \( \Delta\text{spoVG} \) swarming cells on M9 agar also lacked the formation of pink pigment (pulcherrimin) compared to its WT and complemented strain. Additionally, the \( \Delta\text{spoVG} \) swarming cells has a restricted growth even after 24 hours of incubation. Similar experiments in nutrient rich media LBGM identified that \( \Delta\text{spoVG} \) 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, ∆spoVG showed decreased α-amylase production suggesting a deficiency in its secretion system. Comparison of the total secreted proteins of ∆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 ∆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 ∆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 ∆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 ∆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 ∆*spoVG* in this thesis
8.6 Limitations and Future Work

The thesis explored the total transcriptome and proteome of the ΔspoVG 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 ΔspoVG 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 ΔspoVG 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 ΔspoVG, 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 reference point for the ΔspoVG. From the -omics data analysis, a comparison of the ΔspoVG strain in SH5 vs SH2 shall also be made to identify the differentially expressed genes and differentially produced proteins in ΔspoVG in SH5 compared to ΔspoVG 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 ΔspoVG 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 may be 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 each polygon is proportional to the average read count of the transcripts co-IPed with SpoVG in the CLASH analysis.


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


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


Jahn, N. et al. (2012) ‘BsrG/SR4 from Bacillus subtilis- the first temperature-


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.


Mackay, J. P. et al. (2007) ‘Protein interactions: is seeing believing?’, 


Mahapatra, Swati | PhD Thesis | 2020

Bibliography


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.


complexes by phenol-toluol extraction’, *Nature Communications*, 10(1). doi: 10.1038/s41467-019-08942-3.


## Appendices

### Table S.1: Common DPPs between M9 and SH5

<table>
<thead>
<tr>
<th>Locus Tag</th>
<th>Gene</th>
<th>Product</th>
<th>Description</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>BSU09260</td>
<td>yhxA</td>
<td>unknown</td>
<td>similar to adenosylmethionine-8-amino-7-oxononanoate aminotransferase</td>
<td>unknown</td>
</tr>
<tr>
<td>BSU17100</td>
<td>pksC</td>
<td>bacillaene synthase trans-acting acyltransferase</td>
<td>bacillaene synthase trans-acting acyltransferase</td>
<td>polyketide (bacillaene) synthesis</td>
</tr>
<tr>
<td>BSU17120</td>
<td>pksE</td>
<td>unknown</td>
<td>involved in polyketide synthesis</td>
<td>polyketide synthesis</td>
</tr>
<tr>
<td>BSU17170</td>
<td>pksI</td>
<td>unknown</td>
<td>involved in polyketide synthesis</td>
<td>polyketide synthesis</td>
</tr>
<tr>
<td>BSU18320</td>
<td>ppsC</td>
<td>pipastatin synthetase</td>
<td>pipastatin synthetase</td>
<td>production of the antibacterial compound pipastatin</td>
</tr>
<tr>
<td>BSU18550</td>
<td>yoaC</td>
<td>unknown</td>
<td>similar to xylulokinase</td>
<td>unknown</td>
</tr>
<tr>
<td>BSU22970</td>
<td>ypbH</td>
<td>putative adaptor protein</td>
<td>putative adaptor protein for ClpC-ClpP</td>
<td>unknown</td>
</tr>
<tr>
<td>BSU27930</td>
<td>spo0B</td>
<td>sporulation initiation phosphotransferase</td>
<td>Sporulation initiation phosphotransferase of the phosphorelay</td>
<td>initiation of sporulation</td>
</tr>
<tr>
<td>BSU31250</td>
<td>tipA</td>
<td>methyl-accepting chemotaxis protein</td>
<td>membrane-bound chemotaxis receptor, methyl-accepting chemotaxis protein</td>
<td>control of chemotaxis</td>
</tr>
<tr>
<td>BSU31830</td>
<td>yueE</td>
<td>unknown</td>
<td>unknown</td>
<td>unknown</td>
</tr>
<tr>
<td>BSU32859</td>
<td>yuzM</td>
<td>unknown</td>
<td>unknown</td>
<td>unknown</td>
</tr>
<tr>
<td>BSU33750</td>
<td>sdpA</td>
<td>unknown</td>
<td>required for SdpC toxin maturation</td>
<td>maturation of the SdpC toxin</td>
</tr>
<tr>
<td>BSU33770</td>
<td>sdpC</td>
<td>toxin, kills non-sporulating cells</td>
<td>toxin, collapses the proton motive force and induces autolysis, kills non-sporulating cells, induces activity of SigW</td>
<td>killing of non-sporulating sister cells</td>
</tr>
</tbody>
</table>
Table S. 2: Common DPPs between M9 and SH2

<table>
<thead>
<tr>
<th>Locus Tag</th>
<th>Gene</th>
<th>Product</th>
<th>Description</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>BSU08140</td>
<td>yfjD</td>
<td>unknown</td>
<td>unknown</td>
<td>unknown</td>
</tr>
<tr>
<td>BSU10430</td>
<td>yhxD</td>
<td>unknown</td>
<td>general stress protein, similar to alcohol dehydrogenase</td>
<td>survival of salt and ethanol stresses</td>
</tr>
<tr>
<td>BSU12900</td>
<td>htrA</td>
<td>serine protease Do</td>
<td>membrane-anchored protein quality control protease, serine protease Do</td>
<td>protein quality control</td>
</tr>
<tr>
<td>BSU18530</td>
<td>yoaA</td>
<td>unknown</td>
<td>similar to spermidine/spermine N-acetyltransferase</td>
<td>unknown</td>
</tr>
<tr>
<td>BSU19740</td>
<td>yodT</td>
<td>unknown</td>
<td>similar to adenosylmethionine-8-amino-7-oxononanoate aminotransferase</td>
<td>unknown</td>
</tr>
<tr>
<td>BSU27440</td>
<td>glnH</td>
<td>glutamine ABC transporter (binding protein)</td>
<td>glutamine ABC transporter (binding protein)</td>
<td>glutamine uptake</td>
</tr>
<tr>
<td>BSU29930</td>
<td>amyX</td>
<td>pullulanase</td>
<td>pullulanase (debranching enzyme)</td>
<td>starch degradation</td>
</tr>
<tr>
<td>BSU30940</td>
<td>glgP</td>
<td>glycogen phosphorylase</td>
<td>glycogen phosphorylase</td>
<td>glycogen biosynthesis</td>
</tr>
</tbody>
</table>

Table S. 3: Common DPPs between LB and M9

<table>
<thead>
<tr>
<th>Locus Tag</th>
<th>Gene</th>
<th>Product</th>
<th>Description</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>BSU20390</td>
<td>yorG</td>
<td>unknown</td>
<td>unknown</td>
<td>unknown</td>
</tr>
<tr>
<td>BSU00860</td>
<td>clpC</td>
<td>AAA unfoldase, ATPase subunit of the ClpC-ClpP protease</td>
<td>AAA unfoldase, ATPase subunit of the ClpC-ClpP protease, directs proteins phosphorylated on arginine residues to ClpP</td>
<td>protein degradation, positive regulator of autolysin (LytC and LytD) synthesis</td>
</tr>
<tr>
<td>BSU24870</td>
<td>yqqP</td>
<td>intramembrane protease</td>
<td>intramembrane protease</td>
<td>cell division or sugar metabolism</td>
</tr>
<tr>
<td>Locus Tag</td>
<td>Gene</td>
<td>Product</td>
<td>Description</td>
<td>Function</td>
</tr>
<tr>
<td>--------------</td>
<td>--------</td>
<td>------------------------------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>BSU18340</td>
<td>ppsA</td>
<td>plipastatin synthetase</td>
<td>production of the antibacterial compound plipastatin</td>
<td></td>
</tr>
<tr>
<td>BSU27260</td>
<td>mccA</td>
<td>O-acetylserine-thiol-lyase</td>
<td>methionine-to-cysteine conversion</td>
<td></td>
</tr>
<tr>
<td>BSU32770</td>
<td>yusE</td>
<td>unknown</td>
<td>similar to thioredoxin</td>
<td>unknown</td>
</tr>
</tbody>
</table>

Table S. 4: Common DPPs between LB and SH2

<table>
<thead>
<tr>
<th>Locus Tag</th>
<th>Gene</th>
<th>Product</th>
<th>Description</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>BSU15400</td>
<td>ylmG</td>
<td>unknown</td>
<td>transcription-repair coupling factor, eliminates genetic damage from transcriptionally active genes during sporulation, required for increased mutagenesis of lagging strand genes</td>
<td>promotes strand-specific DNA repair by displacing</td>
</tr>
<tr>
<td>BSU00550</td>
<td>mfd</td>
<td>transcription-repair coupling factor</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BSU21990</td>
<td>ypdQ</td>
<td>unknown</td>
<td>similar to RNase HI</td>
<td></td>
</tr>
<tr>
<td>BSU04100</td>
<td>kipR</td>
<td>transcriptional regulator</td>
<td>transcriptional repressor of the kip operon</td>
<td>regulation of sporulation initiation</td>
</tr>
<tr>
<td>BSU28280</td>
<td>leuA</td>
<td>2-isopropylmalate synthase</td>
<td>2-isopropylmalate synthase</td>
<td>biosynthesis of leucine</td>
</tr>
<tr>
<td>BSU29570</td>
<td>sspA</td>
<td>small acid-soluble spore protein (major alpha-type SASP)</td>
<td>small acid-soluble spore protein (major alpha-type SASP)</td>
<td>protection of spore DNA</td>
</tr>
</tbody>
</table>
List of different media used in this study

Lysogeny Broth (LB)

For 1 L of media add following to 800 mL H₂O

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 dH₂O 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. KH₂PO₄ : 1.36 g
6. NH₄Cl : 1.34 g
7. Na₂SO₄ : 0.11 g
8. NH₄NO₃ : 0.10 g
9. FeCl₃.6H₂O : 1 mg

- Add 800 mL H₂O, adjust the pH to 7.0 with 10 N NaOH, make up to 940 mL H₂O, 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$_2$O
Solution A (100 mL):

1. FeCl₃.6H₂O : 0.089 g
2. MgCl₂.6H₂O : 0.830 g
3. MnCl₂.4H₂O : 1.979 g
✓ Make up to 100 mL, sterilise by autoclaving

Solution B (1L):

NH₄Cl : 53.5 g
Na₂SO₄ : 10.6 g
KH₂PO₄ : 6.8 g
NH₄NO₃ : 9.7 g
✓ Add 800mL H₂O; adjust pH to 7.0 with 1 N NaOH; make up to 1 L; filter sterilise; store at 4°C.
Paris Media

**Stocks solutions**

<table>
<thead>
<tr>
<th>Solution Name</th>
<th>Compounds</th>
<th>Concentration</th>
<th>Amount</th>
<th>Sterilization</th>
</tr>
</thead>
<tbody>
<tr>
<td>PC 10X</td>
<td>$K_2HPO_4$</td>
<td>107 g/L</td>
<td>2X 250 mL</td>
<td>Steam sterilize (100˚C for 30 min)</td>
</tr>
<tr>
<td></td>
<td>$K_2HPO_4$</td>
<td>60 g/L</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Trisodium citrate</td>
<td>10 g/L</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Glucose</td>
<td>N/A</td>
<td>100 mL</td>
<td>0.45 µ filter</td>
</tr>
<tr>
<td></td>
<td>K-Glutamate.H_2O</td>
<td>N/A</td>
<td>50 mL</td>
<td>0.20 µ filter</td>
</tr>
<tr>
<td></td>
<td>Fe-ammonium citrate (FAC)</td>
<td>N/A</td>
<td>50 mL</td>
<td>0.20 µ filter</td>
</tr>
<tr>
<td></td>
<td>Casamino acids</td>
<td>N/A</td>
<td>100 mL</td>
<td>Steam</td>
</tr>
<tr>
<td></td>
<td>Tryptophan</td>
<td>N/A</td>
<td>50 mL</td>
<td>Steam</td>
</tr>
<tr>
<td></td>
<td>MgSO_4</td>
<td>N/A</td>
<td>50 mL</td>
<td>Autoclave</td>
</tr>
</tbody>
</table>

**Working solutions**

<table>
<thead>
<tr>
<th>Stock Solutions</th>
<th>Diluted for 100 mL</th>
<th>Diluted for 50 mL</th>
<th>Diluted for 10 mL</th>
</tr>
</thead>
<tbody>
<tr>
<td>PC 10X</td>
<td>10 mL</td>
<td>5 mL</td>
<td>1 mL</td>
</tr>
<tr>
<td>Glucose</td>
<td>5 mL</td>
<td>2.5 mL</td>
<td>500 µL</td>
</tr>
<tr>
<td>K-Glutamate.H_2O</td>
<td>2 mL</td>
<td>1 mL</td>
<td>200 µL</td>
</tr>
<tr>
<td>Fe-ammonium citrate (FAC)</td>
<td>200 µL</td>
<td>100 µL</td>
<td>20 µL</td>
</tr>
<tr>
<td>Casamino acids</td>
<td>5 mL</td>
<td>2.5 mL</td>
<td>500 µL</td>
</tr>
<tr>
<td>Tryptophan</td>
<td>2 mL</td>
<td>1 mL</td>
<td>200 µL</td>
</tr>
<tr>
<td>MgSO_4</td>
<td>300 µL</td>
<td>150 µL</td>
<td>30 µL</td>
</tr>
</tbody>
</table>