The sporulation-specific small regulatory RNAs of *Bacillus subtilis*

Holly Hall

Thesis submitted for the degree

of

Doctor of Philosophy in Medical Sciences

University of Warwick, Warwick Medical School

September 2017

Table of Contents

List of tables

List of Figures

Acknowledgements

I firstly would like to thank my supervisor Emma who gave me the opportunity to do this. Thank you for all your encouragement and support throughout the years, which includes learning how to translate the “Holly speak”. I couldn’t have done it without you!

Alexia, god of everything, thank you! Your constant supply of advice was invaluable. Magic story teller skills, cheerleading in the bluntest way, sanity checks and a lot of laughing, you are amazing!

Gemma, Nick, Richard, Joe, Paul, Josie, Branko, Tom, Ross and everyone else in our unit, you all made for the most amazing group of friends to work with. All our coffee breaks, drinks at the copper pot, late nights in the lab and lunch time madness, you guys completed the whole experience.

A special credit must go to Joe for the limitless debates and never ending technical support. Emily, from your first confession of stalking me to our plans for Vegas, I am so unbelievably glad to have done this with you from start to bitter end! And most of all, I am forever grateful to Paul, Josie and Branko for proofreading, you guys are absolute saints.

Finally, to my mother – for never pushing, but always believing.

Authors Declaration

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. It has been composed by myself and has not been submitted in any previous application for any degree in this or any other university.

The work presented (including data generated and data analysis) was carried out by the author except in the cases outlined below:

- Western blot in Figure 45 was performed by Setlow laboratory
- EM microscopy performed by Warwick Life Sciences Electron Microscopy services in Figures 43 and 44
- Operation of MS/MS machinery performed by the Proteomics Research Technology Platform at the University of Warwick

Signed……

Date……

Abstract

Constantly changing environments in nature have led to bacteria evolving regulatory strategies that result in differential gene expression. A novel and understudied aspect of these networks are regulatory RNAs. The Gram-positive model organism *Bacillus subtilis* not only modulates gene expression to survive a variety of stresses, but also can form endospores to ensure its survival. Sporulation is an essential survival mechanism for many species, allowing them to enter a state of dormancy with resistance to various harsh conditions. This, in turn, ensures survival of not only the population, but also the species.

The process of sporulation requires the controlled expression of approximately a quarter of the genes encoded by *B. subtilis*. Previous large-scale studies have identified that many transcripts do not encode proteins, but exhibited expression profiles similar to genes already known to be part of the sporulation network. Many of these transcripts were selected to likely function as small regulatory RNAs (sRNAs).

This study has shown that many putative sRNAs are active during sporulation, three of which show specific phenotypes that alter germination capabilities in the presence of specific germinants. Cells lacking the necessary components to reverse this process are at a strong disadvantage. Detection of favorable growth conditions is key, but how is this conveyed during metabolic inactivity?

Initial selection of putative sRNAs was done by *in silico* characterization. Prediction of transcriptional control and regulatory regions combined with tiling array profiling was used to select putative sRNAs for confirmation *in vivo.* Transcriptional fusion constructs were generated to confirm compartmental specific expression during sporulation. Spore specific sRNAs were further characterized with phenotypic studies, which suggested a role in endospore formation. This study explored some of the global analysis methods to identify sRNA targets. Whilst no targets for the four chosen sRNAs could be identified, this study produced the most comprehensive data set of proteins to be identified from a *B. subtilis* endospore.

List of abbreviations

AMT: 4’-Aminomethyltrioxsalen hypochloride

BAM: binary sequence alignment map

BCAA: branched-chain amino acid

BLAST: Basic local alignment tool

bp: base-pairs

caDPA: calcium dipicolinic acid

cDNA: complementary DNA

CopraRNA: Comparative Prediction Algorithm for sRNA targets

DBTBS: The Database of Transcriptional Regulation in Bacillus subtilis

DNA: deoxyribonucleic acid

dRNAseq: differential RNA sequencing

ECF: extracytoplasmic function

GEO: Gene Expression Omnibus

gff: general feature format

GFP: green fluorescent protein

GTP: guanosine-5’-triphosphate

LB: Lysogeny broth

LFQ: label free quantification

LIGR-seq: Ligation of interacting RNA and high-throughput sequencing

MFE: minimum free energy

mRNA: messenger RNA

MS/MS: tandem mass spectrometry

NCBI: National Centre for Biotechnology Information

OD: optical density

ORF: open reading frame

PCA: principle component analysis

PCI: Phenol/Chloroform/IAA

PM: Paris medium

RIL-seq: RNA interaction by ligation and sequencing

RM: resuspension medium

RNA: ribonucleic acid

RNase: ribonuclease

SAM: Sequence alignment map

SASP: small acid-soluble proteins

SG: Schaeffer's with glucose

SiLAC: stable isotope labelling by amino acid labelling

SRA: short reads archive

sRNA: small regulatory RNA

SRP: signal recognition particle

TA system: toxin-antitoxin system

TARs: transcriptionally active regions

TAU: arbitrary transcriptional activity units

Tn-seq: transposon sequencing

tRNA: transfer RNA

UMP: uracil monophosphate

1 Chapter one: Introduction

1.1 Bacterial genetic organisation and regulation of gene expression

Bacterial genetic information is generally held on a circular chromosome in the cytoplasm, with additional information potentially held on smaller extrachromosomal elements such as plasmids, and is made up of deoxyribonucleic acid (DNA). The genes encoded by DNA can, under the right conditions, be transcribed into ribonucleic acid (RNA) and then translated into proteins (Figure 1A) (Crick, F., 1970).

The bacterial genome is a highly organised structure, coiled into a circular nucleoid, tightly packing information in to smaller transcriptional units. Transcriptional units, or genes, are comprised of: regulatory information encoded by a promoter, followed by the coding region and, typically, proceeded by a transcriptional termination site (Figure 1B).

The conversion of DNA to RNA is coordinated by transcriptional regulators which recognise specific promoter sites, consisting of two boxes, termed -10 and -35 boxes. RNA polymerase transcribes the DNA to RNA, but different sigma factors are used to help control the specificity as to when transcription is carried out. The position and composition of these boxes can alter the amount or the temporal pattern of expression of a gene. Each box has a specific sequence which is recognised by a different sigma factor, therefore driving the differential production of messenger RNA (mRNA).

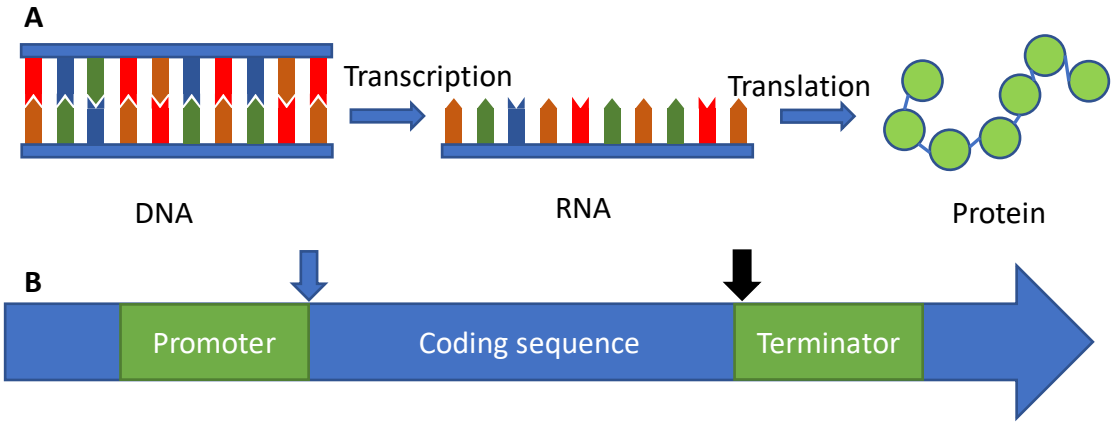


Figure 1 – A) The central dogma of molecular biology – flow from DNA to RNA to protein. B) Simplistic diagram of a gene. A gene consists of a transcriptional start site which is indicated by the promoter, followed by the coding sequence and a terminator. The start codon is indicated by the blue vertical arrow and the black arrow indicates the stop codon; the region between these two sites, including the start codon corresponds to the entire protein encoded by the gene.

In bacteria, transcriptional termination occurs in two different forms, rho-independent and rho-dependent (Ray-Soni, A. *et al.*, 2016). The protein Rho is an RNA helicase, which binds its recognition sequence on an mRNA being synthesised then proceeds to unwind the DNA-RNA complex until it finally dissociates the RNA polymerase from the DNA (Koslover, D. J. *et al.*, 2012). Rho-independent termination relies on palindromic sequences which, when transcribed, form a hairpin loop that results in the dissociation of the RNA polymerase. This is then followed by a stretch of adenine residues which bind weakly to the uracil residues, and therefore releases the mRNA (Wilson, K. S. & von Hippel, P. H., 1995).

In prokaryotes translation is initiated by the ribosome binding at the ribosomal binding site, which starts the conversion of the mRNA sequence into an amino acid chain or polypeptide with the assistance of transfer RNA (tRNA). tRNAs are short RNA sequences made up of a recognition sequence and the corresponding amino acid, with each recognition sequence being a unique triplicate of nucleotides; a codon (Saks, M. E. *et al.*, 1994). As such, RNA not only carries essential biological information into the next stage, in the form of mRNA, but also has a critical role in carrying amino acids during the process of translation via tRNA to ribosomes made up of rRNA. The ribosome binds to the Shine-Dalgarno sequence, which consists of

a six-base consensus sequence (Yang, C. *et al.*, 2016). The ribosome recruits the appropriate tRNA to deposit their amino acid, then moves forward to the next codon until it reaches the stop codon.

Each species of bacteria have their own set of genes that provides it with its capability to survive in its ecological niche (Romero, D., 2013). This set of genes can give an indication as to its growth and survival capabilities. Control over both the transcription and translation of each gene is essential for the cell, as each bacterium has several thousand genes and relatively few are essential for survival at any one time. The expression of different genes at specific times is further facilitated by transcriptional regulators (Roberfroid, S. *et al.*, 2016). Generally, small bacterial genomes are correlated with fewer transcriptional regulators and thus possess fewer mechanisms of survival in response to fluctuations in the environment (Santos, C. L. *et al.*, 2009). Pathogens and symbionts often have smaller sized genomes; such as *Nasuia deltocephalinicola*, an insect symbiont, with the smallest known bacterial genome (Bennett, G. M. & Moran, N. A., 2013) and *Mycoplasma genitalium*, a human pathogen consisting of only 475 genes (Fraser, C. M. *et al.*, 1995). In contrast, soil-dwelling organisms typically possess larger genomes, a solution to coping with a constantly changing environment. For example, *Bacillus* possesses the ability to produce an endospore (discussed in chapter 1 section 2.2), facilitated by this larger genome coding for over 4100 genes.

RNA has also been implicated in the control of gene expression through regulatory RNAs. These groups of RNAs have been identified in all cells from prokaryotes to eukaryotes. Traditionally, the central dogma of molecular biology stated that DNA is transcribed into RNA and then into proteins (Crick, F., 1970) but now know this is not always the case, in that many regulatory RNAs do not code for a translated protein product.

1.1.1 Regulatory RNAs

A DNA sequence that gives rise to the transcription of a regulatory RNA is structured like a typical protein coding gene; in that they are also transcribed from the DNA and therefore transcriptional start and stop signals are present. However, the RNA produced generally lacks the signals for translation and therefore no protein product is made, thus the RNA is non-coding. Despite this, regulatory RNAs act as post-transcriptional regulators of gene expression. As previously mentioned, regulatory RNAs have been found in diverse species, in all domains of life and the majority are likely to be functionally important.

The first regulatory RNA to be discovered was a small regulatory RNA (sRNA) in the Gram-negative bacterium *Escherichia coli*. The 6S RNA was discovered in 1967, although its true function was not identified for several decades. It was eventually shown to interact with RNA polymerase to both activate and inhibit transcription (Hindley, J., 1967; Wassarman, K. M., 2007). This is achieved by direct binding of the 6S RNA near the active site of RNA polymerase, competitively inhibiting or promoting the binding of DNA (Wassarman, K. M. & Saecker, R. M., 2006). The first regulatory RNAs were characterised in the 1980s and were found within plasmids and transposons (Stougaard, P. *et al.*, 1981; Simons, R. W. & Kleckner, N., 1983; Tomizawa, J. *et al.*, 1981). In 1981, two labs independently observed that plasmid replication was inhibited by non-coding elements in *E. coli (Stougaard, P. et al., 1981)*. Tomizawa, J. *et al.* (1981). Tomizawa, J. *et al.* (1981) was the first to present a kissing interaction of the two RNA species, whereby a stable structure is promoted transiently by a RNA-RNA interaction as opposed to stable binding of the two RNAs. Chromosomally located sRNAs, such as *micF* expressed by *E. coli* were also identified in the 1980s (Aiba, H. *et al.*, 1987; Mizuno, T. *et al.*, 1984).

Since the first discoveries of sRNAs, they have been shown to be a diverse group of molecules. sRNAs can be separated into two groups, depending on the location they are encoded with respect to their target RNA, this can either be in *cis* or *trans*. sRNAs have a diverse repertoire of mechanisms of action that include enabling or hindering

translation of an mRNA, stabilisation of transcripts, targeting transcripts for degradation by RNases and some even affect the activity of specific proteins (Hartzog, G. A. & Martens, J. A., 2009). Figure 2 displays a simplified map of the regulatory mechanisms that have been characterised previously; examples and explanations of the various types follow. Not only do sRNAs have diverse mechanisms of actions independently, but some also act in association with proteins, such as Hfq.

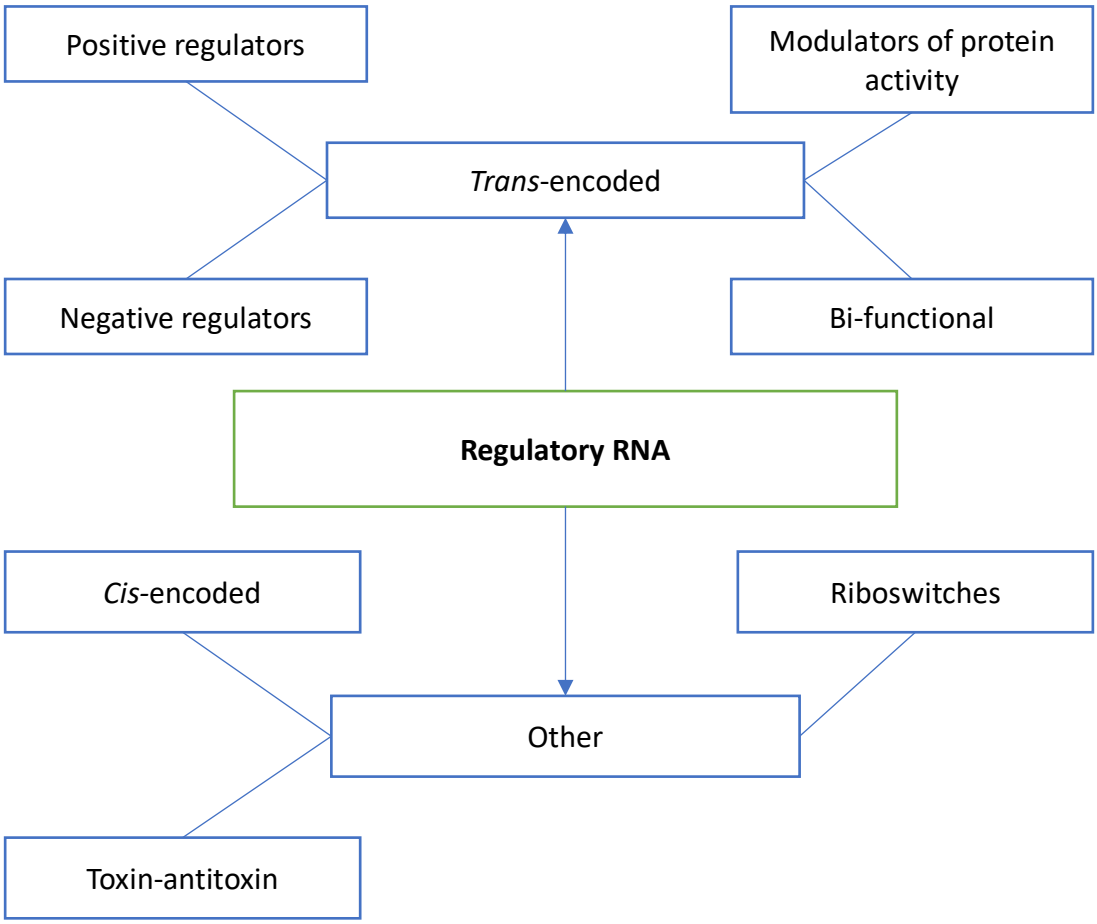


Figure 2 – Different classes of regulatory RNA. Dependent on their location in regards to their target, regulatory RNA can be split into two groups *cis* and *trans*. Further to this, regulatory RNAs are then classified based on their mechanism of action.

1.1.1.1 Proteins can facilitate sRNA activity

Hfq is a chaperone protein that essentially acts as a molecular meeting point for some sRNAs and their mRNA targets. Hfq was first studied in *E. coli*, where it was found to be necessary for RNA phage replication (Franze de Fernandez, M. T. *et al.*, 1972). Hfq has since been shown to be essential in some aspects of post-transcriptional regulation in several of the Gram-negative species. Homologues of Hfq have been identified in a number of different bacteria, both Gram negative and positive (*Listeria* (Christiansen, J. K. et al., 2006)*, E. coli* (Moller, T. et al., 2002), *Vibrio* (Ding, Y. et al., 2004) *Salmonella* (Sittka, A. et al., 2007) and *Bacillus* (Hammerle, H. *et al.*, 2014)). Hfq has a catalogue of sRNAs that it can bind to, with knockout mutants of the gene encoding Hfq resulting in a pleiotropic effect (Tsui, H. C. *et al.*, 1994). It has been shown to play an important role in stabilizing sRNAs and protecting against degradation by ribonucleases (Sledjeski, D. D. *et al.*, 2001; Zhang, A. *et al.*, 2002). As such, the phenotypes associated with Hfq knock-out mutants are likely due to the downstream effects of removing sRNA-mediated regulation, rather than direct effects.

Hfq is a member of the Sm and Sm-like family of RNA binding proteins that contain a common folding domain (Sauter, C. *et al.*, 2003). Hfq is a homohexamer ring and RNA interaction is reported to occur in three locations. The first is on the proximal face, which binds preferentially to the 3' uridine rich locations of some sRNAs (Link, T. M. *et al.*, 2009). Secondly, the distal face, in which adenine rich sequences are favoured (Mikulecky, P. J. *et al.*, 2004). The third location is the lateral face of Hfq, in which the entire sRNA is bound and the seed region is released upon mRNA recognition (Sauer, E. *et al.*, 2012). The exact mechanism behind its activity is not known, however it is recognised that Hfq improves both RNA annealing and strand displacement (Rajkowitsch, L. & Schroeder, R., 2007). Two classes of sRNAs have been proposed for those that bind with Hfq, the first relying on the proximal face of Hfq and the second depending on both the proximal and distal faces (Schu, D. J. *et al.*, 2015). There is no consensus sequence identified as the binding site of Hfq on sRNAs, however it is notable that AU rich regions often bind and as such the Rho-

independent terminators of sRNAs provide a likely location for binding (Otaka, H. *et al.*, 2011).

Hfq is highly conserved within bacteria however, at an amino acid level Hfq is divergent. For example, in *Bacillus cereus* the C-terminus of Hfq is missing (Sun, X. *et al.*, 2002). However, its activity is not universally conserved. *Staphylococcus* Hfq has been shown to have no effect on stress responses (Bohn, C. *et al.*, 2007). Whilst functionally important for Gram-negative bacteria such as *E. coli*, no evidence has been found for its importance in Gram-positives such as *B. subtilis*, despite its conservation. The genome of *B. subtilis* contains an *hfq* orthologue. This encodes a protein of 73 amino acids, which, in comparison to Hfq encoded by *E. coli*, has a shorter C-terminus (Zhang, A. *et al.*, 2002). In *B. subtilis,* however, although it has been shown to bind RNA (Dambach, M. *et al.*, 2013), it ultimately has limited impact on RNA levels (Hammerle, H. *et al.*, 2014). Potentially due to the C-terminus being important for regulatory functionality (Vecerek, B. *et al.*, 2008).

sRNA mediated regulation can work via the actions of several ribonucleases (RNases). sRNAs are known to influence the turnover of their target(s) and this is facilitated by RNases, which are key modulators of RNA decay. For example, it is known that RNase E works in tandem with Hfq in *E. coli*. Hfq has been co-purified with RNase E, forming a ribonucleoprotein complex with an sRNA (Morita, T. *et al.*, 2005). This, in turn, provides a degradation mechanism that is specific to the RNA binding partner of the Hfq-sRNA complex. The sRNA also actively encourages degradation by RNase E, whereby the seed region is optimally placed for the presentation of a 5’ monophosphate group and therefore stimulating RNase activity (Bandyra, K. J. *et al.*, 2012). Indeed, RNase E degradation mediated via sRNA-mRNA interactions has been implicated in a broad range of interactions and as such is an important part of sRNA mediated negative regulation (Waters, S. A. *et al.*, 2017).

1.1.1.2 *Trans*-encoded sRNAs

A major class of sRNAs are the *trans*–encoded sRNAs. This type of sRNA is transcribed distant from the region of the genome which encodes its mRNA target(s). *Trans*-encoded sRNA regulation is often mediated via imperfect base pairing, with an sRNA having a limited complementary sequence to its target(s) (Richards, G. R. & Vanderpool, C. K., 2011; Waters, L. S. & Storz, G., 2009). Previously reported as having primarily negative regulatory effects, binding was found to be at the 5' UTR of an mRNA, blocking the ribosomal binding site and as such, leading to repression of translation (Waters, L. S. & Storz, G., 2009). However, more recent studies have shown the wide variety of mechanisms that can come into play and examples of such are given below.

1.1.1.2.1 Negative regulators

Blocking a single ribosomal binding site of an mRNA is one way in which a *trans*-encoded sRNA can stimulate a negative response, but is one of the best characterised methods. Due to the short, imperfect base pairing a *trans*-encoding sRNA requires to implement its effects, most have the capacity to regulate more than one mRNA and potentially a particular physiological response (Waters, L. S. & Storz, G., 2009).

Post-transcriptional regulation allows the differential translation of genes and, as such, most *trans*-encoded sRNAs are selectively transcribed as a response to a specific condition. For example, regulatory changes in *E. coli* can be induced as a response to low iron stress via the sRNA RyhB. RyhB represses genes that are involved in the management of cellular iron, but in normal growth conditions this is prevented by the transcriptional repressor Fur, which is active in the presence of iron (Masse, E. & Gottesman, S., 2002). Fur binds its recognition sites, otherwise known as Fur boxes, in the presence of Fe^{2+}. As the concentration of Fe^{2+} in the cell drops, Fur is released from promoter sites and repression is lifted (Vassinova, N. & Kozyrev, D., 2000). Under these conditions, the levels of the RhyB transcript increases, which subsequently results in the down regulation of the targets of RhyB by recruitment of the RNA degradosome (Waters, L. S. & Storz, G., 2009). Specifically, Hfq is known to

form a complex with RhyB and RNase E, which in turn initiates degradation (Morita, T. *et al.*, 2005).

MicA is another sRNA from *E. coli* that negatively regulates its targets. MicA primarily works by acting to sequester the *ompA* Shine-Dalgarno sequence to prevent translation and only secondarily promotes degradation via RNase E cleavage. The production of the outer membrane protein OmpA is growth rate dependent (Lugtenberg, B. *et al.*, 1976) and is important for stability of the outer membrane (Wang, Y., 2002). In addition, *ompA* is particularly stable (Arnold, T. E. *et al.*, 1998) and it was previously hypothesized that this stability was directly due to Hfq binding (Vytvytska, O. *et al.*, 2000). However, it was later shown that the sRNA MicA interferes with ribosomal binding and hence inhibits the translation of *ompA* during stationary phase (Udekwu, K. I. *et al.*, 2005).

1.1.1.2.2 RNA sponges act to regulate regulatory RNAs

Sponge RNAs are a type of *trans*-encoded sRNA, which act to regulate non-coding RNAs. Similarly working via binding and altering expression, RNA sponges are RNA molecules with complementary sites to sRNAs, or in eukaryotes miRNAs. These sponges are known to act as mimics, binding sRNAs to prevent binding of an sRNA to its original target (Ebert, M. S. & Sharp, P. A., 2010; Overgaard, M. *et al.*, 2009). Synthetic RNA sponges have been found to be a useful tool for loss-of-function studies, but later it was revealed that RNA sponge mediated control was more naturally ubiquitous than previously thought (Ebert, M. S. *et al.*, 2007; Meng, Y. *et al.*, 2012).

Several examples of RNA sponges have been identified in Gram-negative species. These include SroC, an RNA sponge found in both *Salmonella* and *E. coli*, which targets the sRNA GcvB (Miyakoshi, M. *et al.*, 2015). GcvB is an sRNA that is widely conserved and has one of the largest regulons for post-transcriptional activity (Vanderpool, C. K., 2011; Sharma, C. M. *et al.*, 2011). GcvB is important for the optimisation of amino acid metabolism, with many members of its regulon encoding transporters of amino acids and amino acid biosynthesis factors (Sharma, C. M. *et*

al., 2011). Whilst tightly controlled at a transcriptional level, GcvB also has an extremely short half-life (Vogel, J. *et al.*, 2003). This has been shown to be the result of interaction with the RNA sponge SroC, as SroC facilitates RNase E degradation of GcvB (Miyakoshi, M. *et al.*, 2015).

1.1.1.2.3 Bi-functional sRNAs

Bi-functional sRNAs are a type of regulatory RNA that not only function as an sRNA, targeting mRNA(s) via imperfect base-pairing to alter their target gene expression, but also encode for small proteins or peptides.

The sRNA SgrS encoded by *E. coli* is 227 nucleotides in length and regulates a set of genes involved in the glucose-phosphate starvation response (Papenfort, K. *et al.*, 2013; Wadler, C. S. & Vanderpool, C. K., 2007). The same gene also encodes a short peptide, SgrT that also acts within the glucose-phosphate stress response (Lloyd, C. R. *et al.*, 2017). The sRNA SgrS was found to be responsible for binding and subsequent silencing of *ptsG* (Vanderpool, C. K. & Gottesman, S., 2004). PtsG is part of a phosphotransferase system that mediates the transfer of sugar across the membrane (Zhuang, J. *et al.*, 1999). Binding of SgrS with the *ptsG* mRNA was shown to be in the Shine-Dalgarno region with a total of six nucleotides implicated in the interaction (Kawamoto, H. *et al.*, 2006). Using an *in vitro* translation system, an oligo corresponding to the 14 nucleotides of the binding region of SgrS was shown to be sufficient to block the translation of *ptsG*, even though 31 nucleotides can be shown to have partial complementary using pair-wise alignment (Maki, K. *et al.*, 2010). Preventing the translation of the open reading frame producing SgrT has the same phenotype as deletion of *SgrS,* but it was thought that SgrT does not affect the same target (Wadler, C. S. & Vanderpool, C. K., 2007). It was later discovered that the protein is interacts with PtsG, not via transcriptional control but acts to block the activity of the C-terminal membrane domain, thus inhibiting glucose transport (Lloyd, C. R. *et al.*, 2017).

The SR1 sRNA was the first bi-functional sRNA to be characterised in *Bacillus subtilis* (Gimpel, M. et al., 2010). SR1 encodes for both a 205-nucleotide sRNA and a 39-amino acid long peptide. The regulatory RNA SR1 is involved in arginine catabolism, specifically targeting *ahrC* (Heidrich, N. et al., 2006). Translation of the open reading frame (ORF) of SR1 was initially not seen (Licht, A. *et al.*, 2005). However, the peptide SR1P, was later shown to affect the stability of the *gapA* operon, with various mutants indicating that the phenotype is specifically attributed to the peptide rather than the sRNA (Gimpel, M. *et al.*, 2010). The *gapA* operon is key in glycolysis and encodes the enzyme glyceraldehyde-3-phosphate dehydrogenase, with the 5' gene of the operon repressing induction of the latter part (Ludwig, H. *et al.*, 2001). SR1P acts to stabilise *gapA* and subsequently inhibits its degradation, but the precise mechanism of action is unknown (Gimpel, M. *et al.*, 2010). Together, SR1 and SR1P seem to be acting for the rapid response of the bacterium to glucose. Homologs of SR1 have been found in 23 species, including *Bacillus licheniformis, Bacillus anthracis, Bacillus cereus* and *Geobacillus kaustophilus* (Licht, A. et al., 2005) and homologs from other species have been shown to be capable of complementing the phenotypic variances of deletion mutants (Gimpel, M. *et al.*, 2012).

1.1.1.2.4 Positive regulators

sRNAs have also been shown to have a positive effect on their mRNA transcripts. A frequent mechanism of sRNAs is to interact with their mRNA targets to alter the secondary structure of the mRNA, to remove repression of a translational start site and therefore acting anti-antisense (Waters, L. S. & Storz, G., 2009; Urban, J. H. & Vogel, J., 2007).

The *E. coli* sRNAs DsrA, RprA and ArcZ are three such positive regulators. These sRNAs all target the mRNA encoding the sigma factor RpoS, a sigma factor of stationary phase genes and regulator of general stress genes (Papenfort, K. & Vogel, J., 2009). Each sRNA pairs with the 5' leader sequence of the *rpoS* mRNA, to open an inhibitory hairpin structure, which normally causes self-repression of translation (Brown, L. & Elliott, T., 1997; Soper, T. *et al.*, 2010). Each sRNA is active during different stresses, allowing RpoS to be translated in a variety of different stress

conditions (Lease, R. A. *et al.*, 1998; Majdalani, N. *et al.*, 2002). In addition to this, DsrA is not limited to the regulation of one target as it also functions to repress the translation of *hns* amongst other genes and therefore has been deemed a riboregulator (Lease, R. A. *et al.*, 1998).

1.1.1.2.5 Regulation of protein activity

Whilst many *trans*-encoded sRNAs interact with their mRNA target via base-pairing, some sRNAs have been identified to bind with RNA-binding regulatory proteins. The carbon storage regulator CsrA is an RNA binding protein found in many species of bacteria (Timmermans, J. & Van Melderen, L., 2010). In *E. coli,* CsrB and CsrC are two sRNAs that act to regulate the levels of functionally active CsrA within the cell by sequestering the protein. As cultures enter stationary phase CsrA works as a negative regulator, suppressing carbon metabolism by binding mRNA and preventing ribosome access and generally altering the stability of its RNA targets (Baker, C. S. *et al.*, 2002; Liu, M. Y. *et al.*, 1995). CsrB and CsrC have a crucial sequence and structural relationship with CsrA, where multiple stem-loop structures are capable of binding numerous CsrA proteins and therefore sequestering it to antagonise CsrA function (Dubey, A. K. *et al.*, 2005). This system is not just limited to *E. coli*, with similar systems being found in species such as *Pseudomonas* (Kay, E. et al., 2006) and *Vibrio* (Lenz, D. H. *et al.*, 2005). CsrA is present in *B. subtilis* and was found to control flagellin assembly, but via an alternative mechanism. The FliW-Hag complex is subject to translational repression mediated by CsrA, but the release of FliW from the complex allows FliW to bind CsrA and relieve the repression (Mukherjee, S. *et al.*, 2011).

1.1.1.3 Regulatory RNAs acting in *cis*

Cis acting sRNAs are transcribed with, or opposite their targets. These regulatory elements have similarly diverse characteristics as the *trans*-encoded sRNAs and examples of which follow.

1.1.1.3.1 Riboswitches

A riboswitch is a type of regulatory RNA present in the 5' UTR of the mRNA of which it controls the expression. This is typically achieved by the riboswitch folding into different structures that result in inactivation or activation of expression of the downstream gene/s. To activate gene expression binding of a small molecule induces a conformational change (Waters, L. S. & Storz, G., 2009). This allows the direct sensing of the environment and leads to controlled gene expression (Edwards, T. E. *et al.*, 2007).

There are two main features to riboswitches that alter depending on its structural state (Edwards, T. E. *et al.*, 2007). The aptamer region, which has the ability to bind a ligand and secondly, the expression platform, which implements the change in transcription or translation (Montange, R. K. & Batey, R. T., 2008). The determining factor on the class of a riboswitch is the aptamer region, where each class has a unique secondary structure recognizing a specific molecule offering high specificity (Mandal, M. *et al.*, 2003). The key benefit to riboswitches is that a ligand is needed to introduce a conformational change; this ligand can be an environmental signal, such as lysine (Block, K. F. *et al.*, 2010), allowing the transcription to change depending on the bacterium's environment. The expression platform is directly downstream of the aptamer region and the two are connected by a switching sequence. On binding of the aptamer region, the switching sequence transforms and incorporates itself into the structure of the expression platform, causing a conformational change (Figure 3) (Montange, R. K. & Batey, R. T., 2008).

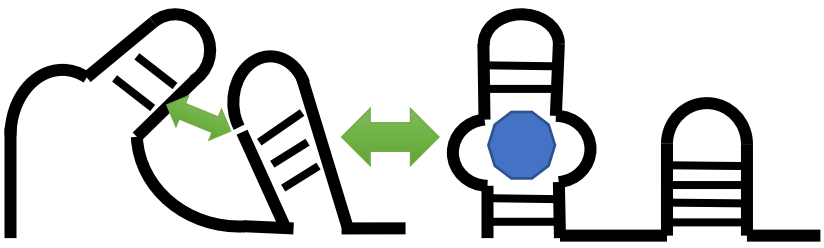


Figure 3 – Schematic of the mechanism of a riboswitch. Upon binding of a ligand (shown in blue), the switching sequence causes conformational change.

An induced fit mechanism binds the effector molecules to all riboswitches, which in turn induces the structural change in the RNA (Savinov, A. *et al.*, 2014). There are two main classes defined by the differences in the binding pocket architecture and the conformational changes on ligand binding. Type I contains a single binding pocket within a pre-organized tertiary structure, the conformational change caused by ligand binding restricted to the proximate binding pocket (Montange, R. K. & Batey, R. T., 2008). For type II, once bound to a ligand the conformational change causes the tertiary structure to alter globally, not just local to the binding pocket, and create the tertiary structure necessary for forming the effective regulatory structure (Trausch, J. J. *et al.*, 2011).

One example of a riboswitch in *B. subtilis* is *ydaO*, which acts via a termination anti-termination mechanism, whereby expression is switched off when c-di-AMP binds (Gundlach, J. *et al.*, 2017). This riboswitch also controls the expression of the *ktrAB* operon, which encodes an ATP-dependent potassium transporter (Block, K. F. *et al.*, 2010). An alternative class features the *glmS* riboswitch. Engaged in feedback regulation of metabolic genes, the *glmS* riboswitch self-cleaves upon binding glucosamine-6-phosphate in combination with other chemical signals (Watson, P. Y. & Fedor, M. J., 2011). This is contrary to *Entreobacteriaceae glmS* regulation, which is controlled by two *trans*-encoded sRNAs (Khan, M. A. *et al.*, 2016). This follows other observations, where riboswitches are favoured over *trans*-encoded sRNAs in Gram-positives as opposed to Gram-negatives.

1.1.1.3.2 *Cis*-encoded sRNAs

Cis-acting sRNAs, unlike the *trans*-encoded sRNAs, are transcribed from the antisense strand of the genes they act to regulate whilst existing as their own RNA species, unlike riboswitches. *Cis*-encoded sRNAs have perfect complementarity to their targets, and this complementarity can be to a part of, or to the full transcript on the opposite strand.

The *E. coli* sRNA GadY is transcribed on the opposite strand to the acid response regulator genes *gadX* and *gadW*. The *gadX-gadW* transcript is innately unstable.

Upon entering stationary phase, expression of GadY increases and this results in its interaction with the 3' UTR of the *gadX* mRNA. Upon binding of GadY, the *gadX* transcript is stabilised and accumulates within the cell (Opdyke, J. A. *et al.*, 2004). Additionally, this base-pairing with the 3' UTR of *gadX* also promotes cleavage of the *gadX-gadW* mRNA by ribonucleases, to produce separate transcripts of *gadW* alongside *gadX* (Opdyke, J. A. *et al.*, 2011). This in turn results in two stable transcripts and increased levels of translation (Opdyke, J. A. *et al.*, 2011).

In *B. subtilis* the general stress response mediated via the alternative sigma factor SigB controls the expression sRNA S1136 (Mars, R. A. *et al.*, 2015a). This sRNA is antisense to the S4 ribosomal binding protein *rpsD* and was reported to downregulate the transcription of *rpsD* in response to ethanol stress and thus reduces the level of the 30S ribosomal subunit in the cell. This presumably occurs by blocking of RNA polymerase (Mars, R. A. *et al.*, 2015a).

Another example in *B. subtilis* was found to function during the developmental process of sporulation, which leads to the formation of a dormant heat resistant endospore. During late sporulation, genes such as those involved in coat assembly are regulated by *cis*-encoded elements. CitB is an aconitase that functions as both an enzyme and, importantly, an RNA binding protein (Alen, C. & Sonenshein, A. L., 1999). The transcriptional regulator *gerE* is critical for coat assembly (Plomp, M. et al., 2014). Deletion of *citB* results in a lack of *gerE* production and *in vitro* gel mobility shift assays showed that *citB* interacts with the *gerE* mRNA. This led to the conclusion that *citB* binds the 3' UTR region of *gerE*. The 3' UTR is a cis-acting element and its stabilisation creates a stable *gerE* transcript that results in increased translation of GerE (Serio, A. W. *et al.*, 2006).

1.1.1.3.3 Toxin-antitoxin systems

Toxin-antitoxin (TA) systems generally include two genes, one of which is a gene encoding for a short protein toxin, whilst the product of the second "antitoxin" gene can either be another protein or a ncRNA molecule. The latter is a class of *cis*-encoded sRNAs that are complementary to their targets. These *cis*-encoded sRNAs

have been found to prevent the toxic build-up of their targets by repressing the translation of the mRNA which codes for the toxin (Kawano, M. *et al.*, 2005; Kawano, M. *et al.*, 2007). Left unchecked, the toxin alone would lead to abhorrent effects.

B. subtilis encodes several toxin-anti-toxin systems, one such TA system is the *ratA/txpA* system. Encoded on opposite strands, with an overlap of 75 nt, RatA is a non-coding RNA which binds *txpA* mRNA and causes its degradation. It has been hypothesised that it prevents the loss of the *skin* element, which is essential during sporulation (Silvaggi, J. M. *et al.*, 2005). Interestingly, this system is also the reason for the essentiality of RNase III, as it protects against the lethal toxins otherwise produced by the prophages *skin* and SPβ (Durand, S. *et al.*, 2012). Similarly, in *E. coli*, the toxin TisB is regulated by its corresponding RNA IstR-1, where SOS induction leads to TisB slowing normal growth (Vogel, J. *et al.*, 2004). IstR1 is thought to act via inhibition of translation, by binding *tisB* and promoting cleavage to produce a truncated mRNA (Darfeuille, F. *et al.*, 2007).

1.1.2 Approaches for the identification and characterisation of sRNAs

The first sRNAs were detected incidentally in the pursuit of other studies. Detection of sRNAs was initially difficult due to the lack of technology to aid their discovery, such as high-throughput unbiased sequencing. Moreover, once an sRNA was identified, characterisation of its mode of action and target genes proved challenging. Classical molecular biology techniques focused on the detection and evaluation of protein coding genes and as such many techniques were not applicable to sRNA characterisation. Since then, genetic screens and transcriptional mapping have accelerated discovery, with several new methods now being dedicated to the discovery and eventual characterisation of sRNAs.

1.1.2.1 Bio-computational screens for sRNAs

The first screens for sRNAs were computational and primarily looked for conservation and presence of promoters for detection of a putative sRNA (Argaman,

L. *et al.*, 2001). A global search for sRNAs was performed in *E. coli* and desirable characteristics were chosen based on the few characterised sRNA known at the time (Argaman, L. *et al.*, 2001). Intergenic regions were explored for potential promoter sequences and termination sites which were no more than 400 nt apart finding 24 putative sRNAs, experimentally validating 14 sRNAs via analysis of expression patterns on Northern blots (Argaman, L. *et al.*, 2001).

1.1.2.1.1 Comparative genomics searches for sRNAs

Early searches focussed on sequence conservation when identifying sRNAs (Argaman, L. *et al.*, 2001; Rivas, E. *et al.*, 2001; Rivas, E. & Eddy, S. R., 2001). Typically pipelines started with conservation, followed by clustering and multiple sequence alignments, which finally was complemented with structural information (Rivas, E. *et al.*, 2001). Alternatively, promoter and termination sites were mapped and accepted as putative sRNAs assuming that they were a certain length apart such as in the study by Chen, S. *et al.* (2002), who then verified the presence of several sRNAs via Northern blot.

Within the Gammaproteobacteria it is not unusual to find that sRNAs are well conserved (Peer, A. & Margalit, H., 2014). However, in the model Gram-positive species *B. subtilis* sRNAs appear to be more limited in their conservation, being primarily conserved within *Bacillus* (Mars, R. A. *et al.*, 2016). One of the more diverse sRNAs in Gram-positives is RsaE/RoxS, which although initially identified in *S. aureus*, is conserved across both Staphylococcal and Bacillus species (Durand, S. *et al.*, 2015). As such, phylogenetic distance is important to note in the comparison of putative sRNA sequences and should be more limited than typical approaches with protein coding genes (Lindgreen, S. *et al.*, 2014).

Many sRNAs are found to be involved in stress responses and therefore not present in typical lab growth conditions, such as in LB during exponential growth (Nicolas, P. *et al.*, 2012). With hundreds of potential conditions to be explored, the prediction of the targets of an sRNA was a logical next step, as the sRNA would likely be expressed at the same time as the target. From target predictions, it is hoped that

characterisation would be facilitated as it helps narrow down the conditions where an sRNA is active, and the process in which it controls.

1.1.2.1.2 sRNA Target prediction

It was previously thought that a sequence matching approach would be sufficient to predict targets of an sRNA (Rivas, E. & Eddy, S. R., 2001). However, as more sRNAs and their targets are being confirmed it has become clear that this is not the case. There are few sequence or structural features that are conserved between different sRNAs making it difficult to make rules needed to establish rules for pattern recognition programs. In addition, the interacting region can be located anywhere in the sRNA and binding is often imperfect, with the length of the hybridising regions often varying and potentially being non-consecutive (Papenfort, K. *et al.*, 2012). As such, whilst predictions can be made they are often plagued by also producing many false positives.

TargetRNA is an sRNA target prediction program which utilises binding free energy to predict interactions of a given nucleotide sequence and the coding genes of a given species (Tjaden, B., 2008). A recent update also advises structure and sequence homology of the sRNA sequence, but does not use this to inform predictions outside of using this to predict the most likely seed regions (Kery, M. B. *et al.*, 2014).

RNApredator is another webserver for the prediction of sRNA targets. (Tafer, H. & Hofacker, I. L., 2008; Eggenhofer, F. *et al.*, 2011). RNApredator searches the 3' and 5' UTR regions of an mRNA, +/- 200 nucleotides of a start or stop site. RNApredator utilises RNAplex for prediction of short RNA-RNA interactions between two RNA sequences, then utilises this information to consider accessibility when assessing potential interactions.

CopraRNA (for Comparative Prediction Algorithm for sRNA targets) utilises IntaRNA for its binding potential, evaluating minimum free energy (MFE) required for its interaction like the approaches utilised by TargetRNA and RNApredator (Busch, A. *et al.*, 2008). However, one of the core parameters for CopraRNA is the incorporation

of conservation, with at least three homologous sequences from different species needed to complete the analysis (Wright, P. R. *et al.*, 2013). CopraRNA then uses IntaRNA to evaluate all species, not just the species of specific interest (Wright, P. R. *et al.*, 2014).

sRNAs can target protein coding regions as discussed above, which effect stability or translation and are not accommodated in programs such as TargetRNA. In addition, whilst programs are improving, many still miss interactions confirmed experimentally or fail to identify an sRNA's multiple targets. Computational screens however, can be improved upon once some targets have been identified. Sharma, C. M. *et al.* (2011) could take advantage of the extensive validated targets sites to create a consensus sequence. This was then used to search for further targets within *Salmonella* 5' regions of all known ORFs which revealed 234 mRNAs. TargetRNA software was then utilised to reduce this down to 42 targets.

1.1.2.2 Wet lab based methods of identification and characterisation

After the initial prediction or identification of putative sRNAs it is required to confirm its expression and characterise its function. Confirmation of the presence of an sRNA and its resulting activity can be done in several ways, from basic molecular biology and classical Northern blotting to newer techniques involving next-generation sequencing and proteomics.

1.1.2.2.1 Direct monitoring of activity of a putative sRNA

Identification of the RNA transcript via Northern blotting was the method of choice to confirm a predicted sRNA. Northern blots facilitate the detection of expression, by monitoring the presence and absence of a transcript over time, and length of an sRNA. For example, Saito, S. *et al.* (2009) utilised Northern blots to identify new sRNAs directly by analysing intergenic regions of more than 500 nt in *B. subtilis*, identifying seven new putative sRNAs of which six were expressed from their own promoter. Further to this, the first sRNAs were mapped utilising primer extension and 3' RACE to map the exact start and stop sites of the sRNA transcripts, many

leading to the discovery of processing elements such as cleavage with RNase III or endonucleases (Argaman, L. *et al.*, 2001).

Detectable markers have long been used for the qualitative and quantitative analysis of diverse arrays for the study of biological processes. For example, Green fluorescent protein (GFP) and other fluorescent markers such as mCherry can be fused to a promoter of interest and the increase or decrease in expression can be used to measure the activity of the promoter (Chalfie, M. *et al.*, 1994). Not only can this system be used to monitor gene expression, but GFP translational fusions have been successfully used in the analysis of sRNA-mediated control of their target mRNAs, and the identification of other factors involved (Urban, J. H. & Vogel, J., 2007; Isaacs, F. J. *et al.*, 2004).

1.1.2.2.2 Hfq co-purification

In bacteria where Hfq is known to play a role in RNA-RNA interactions, co-immunoprecipitation followed by microarray or RNAseq analysis has proven to be successful in identifying sRNAs. One study by Zhang, A. *et al.* (2003) utilised this method to confirm the interaction of at least 15 known interactions between sRNAs and Hfq and further identified 20 potential Hfq-binding sRNAs. In addition to this, the compendium of Hfq-associated sRNAs was expanded by immunoprecipitation coupled with deep sequencing and revealed that almost 300 mRNAs bind to Hfq (Bilusic, I. *et al.*, 2014). This method of analysis has been carried out on *B. subtilis*, however, although many RNA species associate with Hfq, only a small subset of sRNAs identified binding with Hfq were affected by deletion of *hfq* (Dambach, M. *et al.*, 2013).

Despite being able to directly identify interacting sRNAs, the main limitation to this method is the narrow snapshot taken. This is because it will only detect those sRNAs interacting with Hfq, in the specific condition and at the time point chosen to carry out the study. A comprehensive study across many differing conditions would alleviate some of these issues.

1.1.2.2.3 RNomics studies are used for the identification of non-coding sequences

sRNA studies have advanced considerably since the advent of whole genome sequencing and hence enabling genome wide searches to be carried out. Specific studies for the search of previously unidentified RNA species have utilised the capability to look at the whole transcriptome without having to only search for known loci.

One study that resulted in the reannotation of a genome was carried out by Nicolas, P. *et al.* (2012) using tiling arrays. This study took the wild-type 168 *B. subtilis* strain and analysed the RNA expression profile over 104 different environmental conditions. This resulted in the identification of over 1500 new RNA segments. With the use of tiling arrays, the transcriptional landscape was documented, unbiased to the previous annotation of *B. subtilis*, and approximately 10% of these new RNA segments are now assumed to be new sRNAs of various types.

Gene expression studies have now moved into a new era, with RNAseq becoming the technique of choice over microarray use for transcriptomics, amongst other new techniques. Next generation sequencing has evolved to allow high-throughput sequencing of cDNA libraries, in which mRNA is reverse transcribed into complementary DNA (cDNA) followed by DNA sequencing to determine the transcriptional landscape (Wang, Z. *et al.*, 2009). RNAseq has been suggested to take over microarray use entirely, with benefits including single nucleotide resolution and the ability to detect low level transcripts.

Differential RNAseq (dRNAseq) is a method that can be used to globally map transcriptional start sites. In addition, this technique has been utilised in the identification of RNase cleavage sites (Gordon, G. C. *et al.*, 2017). Two libraries are prepared of a given sample, one for standard RNAseq and the other is treated to enable the enrichment of primary transcripts over processed RNA, such as rRNA and tRNA. Irnov, I. *et al.* (2010) utilised this method to enhance the understanding of sRNAs in *B. subtilis*, in which they expanded the number of potential sRNAs in *B. subtilis* to over 100 candidates. In addition, more than 1900 unique transcriptional

start sites have been identified in *Helicobacter pylori* with approximately 60 of these corresponding to sRNAs (Sharma, C. M. *et al.*, 2010). This method has also been successfully used for the identification of sRNAs and strain specific transcriptome analysis of four *Campylobacter jejuni* strains, with more than 1900 transcriptional start sites for each and a total of 39 putative sRNAs identified (Dugar, G. *et al.*, 2013).

Mapping of 3' ends is relatively easy in eukaryotes since mature mRNA has a poly(A) tail. This poly(A) tail can be used as a basis for reverse transcription followed by RNAseq. Reverse transcription is important so that the 3' end is read first and therefore producing a stronger signal than standard RNAseq, whereby typically reads are of a fixed length and generated from the 5' end (Gruber, A. R. *et al.*, 2014). Bacterial RNAs, however, lack this characteristic and as such mapping 3' ends is more difficult. The first high-throughput method of identifying *cis*-encoded elements in bacteria was Term-seq (Dar, D. *et al.*, 2016). Term-seq quantitatively maps 3' ends of available RNA similarly by reversing the reading of a transcript. Additional steps make this possible and by ligation of an adapter to the 3' end it is possible to create an identifiable characteristic like the poly(A) tail (Dar, D. *et al.*, 2016). Dar, D. *et al.* (2016) developed this method and successfully identified many new condition-dependant regulation events in *B. subtilis, Listeria monocytogenes,* and *Enterococcus faecalis* in response to antibiotics.

1.1.2.2.4 RNomics can also be utilised for discovery of targets

Whilst transcriptomics on its own can identify the expression of an sRNA in a certain condition, it does not facilitate elucidation of the target. Comparison of RNA expression profiles can be performed between wild-type and an sRNA null mutant, with those genes of altered expression indicating a disruption of function caused by the sRNA. Further to this, the identification of sRNA targets can be further facilitated by pulse expression of an sRNA. Under the control of an inducible promoter, an sRNA can be expressed for a defined period and the transcriptome can again be captured. Comparing the pulse expressed sRNA in a deletion background with typical expression patterns can therefore monitor the change of expression of any gene, with repression or activation indicating a primary role of the sRNA. From this

comparison, short bursts of expression can be used to confirm primary targets. This technique was pioneered by Sharma, C. M. *et al.* (2011) who successfully used it to identify new targets of the sRNA GcvB expressed in *Salmonella* using whole-genome microarrays, identifying 45 RNAs which could be related to GcvB activity.

Whilst pulse expression improves on standard comparisons between wild-type and deletion mutants, by lending confidence of changes being due to the effects on primary target(s) and not compensation for a deletion, there are limitations. Those sRNAs which control the translation of their targets will be missed, alongside those not active in the growth condition. Transcriptomics is also limited in statistical power, and weak changes in mRNA levels will be missed.

1.1.2.2.5 RNA-RNA interactions

The lack of tools to directly identify targets has resulted in new approaches being designed to address this. One of the first methods to directly identify RNA-RNA interactions involved crosslinking and RT-PCR (Lustig, Y. *et al.*, 2010). However, more recently crosslinking has been combined with high-throughput sequencing to identify sRNA-mRNA duplexes with the help of ligation. CLASH relies on the previous observations of interactions with proteins, such as the endoribonuclease RNase E or Hfq. Waters, S. A. *et al.* (2017) identified thousands of RNA-RNA duplexes, with 782 statistically significant sRNA-mRNA interactions in log phase in *E. coli*. This was comparable to a similar technique, RIL-seq, employed by Melamed, S. *et al.* (2016) which utilised the RNA chaperone Hfq to find 633 statistically significant sRNA-mRNA interactions in log phase in *E. coli*. Both studies highlight the complexity of sRNA mediated changes in the wider cellular network and our current lack of understanding of such.

1.2 *Bacillus subtilis*

B. subtilis is a rod shaped, Gram-positive bacterium within the phylum Firmicutes that is naturally found in the soil and human gut (Nakano, M. M. & Zuber, P., 1998). Much interest has been found in *B. subtilis* in both commercial and academic laboratories. *B. subtilis* is commonly utilised as a "microbial cell factory" for industrial scale production of enzymes and vitamins. In addition, its natural competence means that *B. subtilis* has and remains an excellent candidate for experimental investigations. As such, there is a large catalogue of information for this organism and it is widely considered as the Gram-positive model organism. The *B. subtilis* genome has been annotated to comprise of over 4100 protein coding genes (Kunst, F. *et al.*, 1997; Belda, E. *et al.*, 2013). Many of these genes have known functions due to years of research, but only 642 of which are found to be essential for growth in complex medium (Reuss, D. R. *et al.*, 2016).

1.2.1 Regulation

It is well known that bacteria alter their gene expression patterns to survive in stressful conditions. *Bacillus* species are found in many different environments such as the human gut and soil. Many of these environments often require the bacteria to undertake sophisticated developmental changes for them to survive. The bacteria take cues from the environment, which causes rearrangement of gene expression often resulting from complex signal transduction systems as explained below.

Genes that control normal homeostasis are controlled by sigma factor SigA (Yeh, H. Y. *et al.*, 2011), with alternative sigma factors being involved in response to different stresses. The general stress sigma factor SigB controls a regulon of around 150 genes; all being involved in returning the cell to normal homeostasis after exposure to stresses such as heat or osmotic changes (Hecker, M. *et al.*, 2007). Sigma factors M, W and X are three of seven sigma factors involved in extracellular functions. The SigW regulon includes approximately 89 genes for the protection against toxic proteins, altering transcription so that these toxins can be eliminated or inactivated (Cao, M. *et al.*, 2002a; Zweers, J. C. *et al.*, 2012). SigW combined with SigM is

essential for growth in high salt conditions. Additionally, SigW and SigX share partial overlap in activity, whilst maintaining indivual regulons, which contribute to resistance against cationic antimicrobial eptides (Turner, M. S. & Helmann, J. D. , 2000). SigM controls a regulon of genes that help to manage cell wall stress (Cao, M. *et al.*, 2002b).

A response to environmental stress is the formation of an endospore, a dormant cell type that is resistant to many types of external stresses. The production of an endospore is a controlled process, which includes the sigma factor SigH that is involved in regulating the initiation of the endospore formation process and the four compartment specific sigma factors SigE, F, G and K.

Table 1 – Summary of sigma factors in *B. subtilis*. (extracytoplasmic function = ECF) (Haldenwang, W. G., 1995; Michna, R. H. et al., 2016)

Sigma Factor	Function
SigA	Housekeeping
SigB	General stress
SigD	Chemotaxis and motility genes
SigE	Early mother cell-specific sporulation
SigF	Early forespore-specific sporulation
SigG	Late forespore-specific sporulation
SigH	Transition phase – post exponential and early sporulation
SigK	Late mother cell-specific sporulation
SigL	Degradative enzyme, enhancer-dependent (Sigma-54 family)
SigM	ECF & high salt concentrations
SigV	ECF required for resistance to lytic enzymes
SigW	ECF & cell wall stress
SigX	ECF resistance against cationic antimicrobial peptides
SigY	ECF
SigZ	ECF

1.2.2 Sporulation

The process of endospore formation appears to be the same in most endospore formers and as such, it is thought that most evolved from a single ancestor due to the crucial regulatory pathways for sporulation being highly conserved (Marchais, A. *et al.*, 2011). The term endospore relates to the specific way the endospore is formed, with the eventual resistant endospore being produced within the original cell. An endospore is a distinct morphological cell and metabolically dormant (Higgins, D. & Dworkin, J., 2012).

Briefly, the cell commits to sporulation with the formation of a septum being placed asymmetrically at a specific polar position. This results in splitting the cell into two distinct compartments. The smaller of the compartments is destined to become the endospore, named the forespore at this point, and the larger compartment, being the mother cell and is destined to perish at the end of sporulation. This mother cell engulfs the forespore, with the forespore becoming a cell within the mother cell, eventually becoming an endospore. The mother cell coats the forespore in a number of sporulation specific layers and the forespore is eventually liberated via lysis from the mother cell. The stages are designated by roman numerals and are explained fully below.

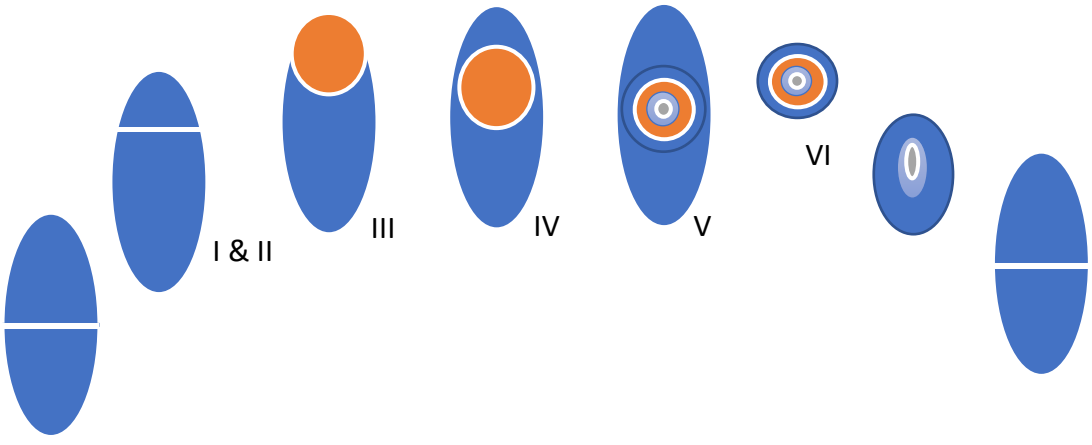


Figure 4 – Schematic representation of *B. subtilis* lifecycle, from vegetative growth, through endospore formation to germination. First observable morphological change is establishment of asymmetrically positioned septum, splitting the cell into the forespore (diplayed in orange) and the mother cell (displayed in blue) (I,II & III). Subsequently, the forespore is engulfed by the mothercell (IV). The Forespore is

coated in a number of sporulation specific layers (V). The forespore is released from the mothercell by lysis and is refered to as the endospore (VI). Further maturation events take place to produce the mature endospore.

1.2.2.1 Starvation

Determining when to trigger sporulation is essential for the cell's future survival. Sporulation is a high-energy process to complete, impossible without nutrients and with no reversibility after stage II (Eijlander, R. T. *et al.*, 2013). As such, there is no single prompt to enter sporulation, but rather a series of different signals aggregated into the activation of a key regulatory protein, Spo0A. Each cell makes its own individual decision, resulting in part of the population opting to enter the sporulation pathway and others not (Narula, J. *et al.*, 2012). It has been suggested that this variability is a "bet-hedging" mechanism, ensuring that part of the population is ready to survive if the environmental conditions change (de Jong, I. G. *et al.*, 2010).

To determine a cell's fate, the cell considers many different factors; processing signals from other cells in the population, such as peptide pheromones to ascertain a neighbouring cell's decision or quorum-sensing peptides for cell density (Schultz, D. *et al.*, 2009). Upon sensing it is nutrient limited or otherwise stressed, the cell might first employ several other tactics before finally deciding to proceed with endospore formation; it may become more motile to search for nutrients, or it may secrete a variety of weapons to destroy the competition for nutrients and scavenge available proteins (Schultz, D. *et al.*, 2009).

Whilst making the decision to commit to sporulation the cells have the option to go into a state termed competence. This is where the bacterium becomes permeable to genetic material in the surrounding media, using it for recombination or repair of DNA. This state is then returned into vegetative growth (Suel, G. M. *et al.*, 2006). This allows the cell the opportunity to survive, as in nutrient limited conditions it is possible to scavenge phosphates for DNA repair or the acquired DNA could be incorporated into the genome, providing a possible advantage in surviving its new

conditions. If still stressed, the cell may use other survival mechanisms (Hamoen, L. W. *et al.*, 2003).

The sporulation "master regulator", Spo0A, is activated via a phosphorelay process, allowing the phosphorylation of Spo0A (Spo0A-P) (Stragier, P. & Losick, R., 1996) where it leads to the differential regulation of -120 genes (Fujita, M. & Losick, R., 2005). Spo0A-P accumulation is measured to determine progression of the sporulation response amongst other signals (Narula, J. *et al.*, 2012) and this gradual increase in Spo0A-P has been shown to be essential for sporulation efficiency (Vishnoi, M. *et al.*, 2013). Because of this, it has been proposed that sporulation has a variable clock rate, the likelihood of sporulation responding to fluctuations of the concentrations of Spo0A-P as the outside environment becomes more or less favourable (Schultz, D. *et al.*, 2009). Many of the genes under the control of Spo0A-P have a varying range of activation thresholds; those that require low levels of phosphorylated Spo0A tend to be genes for survival tactics such as biofilm production or cannibalism. This low threshold means that only a small amount of Spo0A-P is required to activate them and thus the genes required for these phenotypes are activated before others in the regulon that result in more drastic phenotypes such as the process of sporulation (Chastanet, A. *et al.*, 2010).

For activation, Spo0A is unusual in the sense that there is no single sensor kinase that phosphorylates it. Instead several kinases are used indirectly to create this phosphorelay, with KinA and KinB being primarily utilized to phosphorylate Spo0F (Hilbert, D. W. & Piggot, P. J., 2004). This then progresses the pathway to transfer the phosphoryl group to Spo0B which in turn transfers the phosphoryl group to Spo0A, resulting in Spo0A-P and thus resulting in its activation (Fujita, M. & Losick, R., 2005). This phosphorelay is subject to negative regulation: with phosphatases preventing the activation of Spo0A by dephosphorylation (Hilbert, D. W. & Piggot, P. J., 2004). The Rap proteins are the main effectors upon low cell density, and kinase inhibition occurs by a small protein, Sda, by binding of KinA and blocking autophosphorylation of KinA (Higgins, D. & Dworkin, J., 2012; Whitten, A. E. *et al.*, 2007). Spo0A has two main domains, one where the phosphate group can be

accepted and the other being the DNA binding domain. The phosphate accepting domain has an aromatic switch mechanism where phosphorylation incurs a conformational change. From this conformational change, a phenylalanine residue changes position and activates the DNA binding domain (Higgins, D. & Dworkin, J., 2012). This DNA binding domain is essential for activity of Spo0A where it can then go on to bind 0A boxes to promote expression of members of the Spo0A regulon.

1.2.2.2 Stages I – III

Upon accumulation of the appropriate levels of Spo0A-P, the synthesis of factors needed for the next stages of endospore formation are activated. During normal development *B. subtilis* undergoes medial division, dividing down the middle in order to produce two identical cells (Stragier, P. & Losick, R., 1996). However, during endospore formation division is asymmetric. Spo0A-P activates the transcription of the gene *racA* which is an effector for axial filament formation (Hilbert, D. W. & Piggot, P. J., 2004). This allows alignment of the chromosome across the long axis of the cell, which may have formed two complete chromosomes or may be partially replicated (Hilbert, D. W. & Piggot, P. J., 2004), ready for the septum to form asymmetrically. DivIVA localises RacA, which binds specifically to GC-rich 14 bp repeats close to the origin of the chromosome but for axial filament creation appears non-specific, to the pole of the cell, connecting the two (Ben-Yehuda, S. *et al.*, 2003; Ben-Yehuda, S. *et al.*, 2005). With the chromosome stretched to form a filament, the forespore generally contains 30% of the chromosome upon septation and the remainder purported to being moved into the forespore by the DNA translocase SpoIIIE, forming two channels in the centre of the septum, in which the two arms of the chromosome are transported simultaneously (Higgins, D. & Dworkin, J., 2012). This results in the mother cell and forespore having the same genome (Xenopoulos, P. & Piggot, P. J., 2011).

During normal cell development, FtsZ forms a “Z-ring” mid-cell. However, during sporulation, Spo0A-P triggers this Z-ring to form near the poles of the cell (Levin, P. A. & Losick, R., 1996). From this, the cell membrane is used to create a septum asymmetrically at one pole. It has been suggested that the option to form

endospores at either end is a failsafe mechanism to ensure success in case of incorrect axial filament formation (Hilbert, D. W. & Piggot, P. J., 2004). This is created at a sub-polar location, by way of the protein FtsA localizing to one pole, which then folds in on itself to create a compartment that becomes engulfed by the mother cell as the septum pinches in on itself (Xenopoulos, P. & Piggot, P. J., 2011). Different SpoII proteins have been described as playing roles in the hydrolysis of the peptidoglycan to facilitate engulfment of the forespore, together with SpoIIIE (Higgins, D. & Dworkin, J., 2012). This creates a free cell within a cell, completion of this is designated stage III.

A major component of regulation between the distinctive gene expression patterns of the mother cell and the forespore is the different RNA polymerase sigma factors. These sigma factors tightly control gene expression and are expressed themselves in a pattern depending on the stage of sporulation and which compartment they are in (Steil, L. *et al.*, 2005). It is indicated that once engulfment has taken place and the different patterns of expression are initiated, the process of sporulation is committed and irreversible (Higgins, D. & Dworkin, J., 2012).

The different patterns of gene expression between the mother cell and forespore are linked. Compartmentalization of gene expression has been shown to be heavily influenced by the *spoIIA* operon. Its transcription is dependent on Spo0A and SigH prior to asymmetric division. Coded within this operon is SigF, which is regulated by two out of three other products of this operon, SpoIIAA and SpoIIAB. SpoIIAB is an anti-sigma factor, directly binding to SigF, and SpoIIAA is an antagonist of SpoIIAB, an anti-anti-sigma factor (Schmidt, R. *et al.*, 1990; Duncan, L. *et al.*, 1996). Despite the transcription of the gene being activated before polar septation, SigF becomes active afterwards.

SigF has a regulon of approximately 50 genes that control gene expression during early forespore development (Steil, L. *et al.*, 2005). Included in this regulon is the protein SpoIIR, which crosses the membrane between the forespore and mother cell and stimulates processing of SigE, which is the early regulator of gene expression

within the mother cell (Steil, L. *et al.*, 2005). This processing is essential for the activation of SigE, as it is encoded by the *spoIIG* locus and synthesized in an inactive state before proteolytic cleavage of 27 residues from the N terminus (Hilbert, D. W. & Piggot, P. J., 2004). The *spoIIG* locus codes for two products; firstly, *spoIIGA* which is responsible for processing SigE, and *spoIIGB* which codes for pro-SigE (Hilbert, D. W. & Piggot, P. J., 2004). The SigE regulon contains many genes essential for engulfment, scaffold proteins for endospore coat assembly and transcription of the *spoIIID* operon. At a later point in sporulation SpoIIID represses some of the genes controlled by SigE and activates transcription of others (Hilbert, D. W. & Piggot, P. J., 2004). SpoIIID is also important for the activation of the late sporulation sigma factor SigG in the forespore, which in turn activates SigK in the mother cell shortly after engulfment (Higgins, D. & Dworkin, J., 2012; Hilbert, D. W. & Piggot, P. J., 2004). SigG is transcribed from SigK, which is activated by all the products of the *spoIIIA* operon (Figure 5).

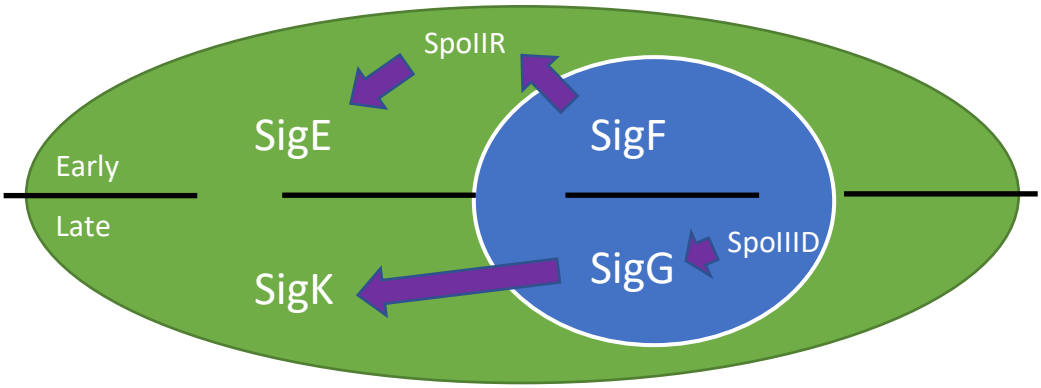


Figure 5 – Cascades of sigma factor activation and their compartment specificity. During early forespore development, SigF controls expression of SpoIIR, which crosses the membrane between the forespore and mother cell. SpoIIR stimulates proteolytic cleavage of pro-SigE and activates it to SigE. SigG is instrumental in late endospore development and is activated by SpoIIID in the forespore. Following this activation SigG is responsible for the transcription of SigK following engulfment.

Some pathways in which communication occurs between the mother cell and forespore have so far remained undetermined. One such pathway that remains

undetermined is how the transcription factor SigG, from the forespore, is activated following SigE directed gene expression. Interestingly, it has been suggested that this may be facilitated by a channel, a gap junction-like “feeding tube” in which small molecules from the mother cell are transported into the forespore (Camp, A. H. & Losick, R., 2009).

1.2.2.3 Stages IV – VII

Safeguarding the forespore from the mother cell are two membranes: the outer forespore membrane and the inner forespore membrane. The outer forespore membrane initially allows access of proteins from the mother cell (Higgins, D. & Dworkin, J., 2012). Three layers are laid down in the subsequent stages between the membranes surrounding the forespore; the first being the germ wall, which is composed of the membranes surrounding the forespore during its development, and the other the cortex (Hilbert, D. W. & Piggot, P. J., 2004) and finally a protein coat, double layered for protection against harsh chemicals. The germ wall will later become the cell wall upon germination (Steil, L. *et al.*, 2005). In stage IV, the layer of peptidoglycan forming the cortex is created between the two layers of the forespore membrane. Following this, during stage V, the endospore coat is created. Both the cortex and the coat contribute to the endospore’s protection from the outside environment (Hilbert, D. W. & Piggot, P. J., 2004).

The cortex plays an important role in protection against most of the severe environmental stresses; the dehydration of the cell lends protection against heat and is largely maintained by the cortex. This layer is made up of modified peptidoglycan, formed between the two layers initially found after engulfment. The mother cell-produced protein SpoVE plays an important role in early peptidoglycan polymerization, localizing at the outer forespore membrane (Higgins, D. & Dworkin, J., 2012).

The endospore cortex of *B. subtilis* has been shown to be *O*-acetylated. The result of this modification of the peptidoglycan is to reduce the sensitivity of the cell to the

anti-microbial lysozyme. This modification may be particularly useful if the lysozyme penetrates the endospore coat due to external damage (Laaberki, M. H. *et al.*, 2011).

The endospore coat is assembled by the mother cell, consisting of ~80 proteins arranged in inner and outer layers (Isticato, R. *et al.*, 2010). The protein SpoIVA targets the outer forespore membrane, hydrolysing ATP to form a base layer of structures for coat construction. Several Cot proteins have been found to be essential in the formation of the coat, with mutants leading to defective endospore coats. Not all defects in the endospore coat are obvious, as it is possible for the endospore to create thermo-resistance without having resistance to, for example, UV or chemical stress. For example, a protocol may discriminate between a normal growing cell and an endospore via prolonged exposure to heat however, the defect may impair the cell's resistance to ethanol. From this, a phenotype is difficult to determine and mutants may need to be tested for all types of resistance (Isticato, R. *et al.*, 2010; Zheng, L. B. *et al.*, 1988)

Coat assembly is followed by maturation in stage VI, where it obtains its resistance to harsh environments. Stage VII is the lysis of the mother cell and release of the endospore. Once the endospore is exposed to a more favourable environment, it can germinate and outgrowth into a standard cell can occur (Hilbert, D. W. & Piggot, P. J., 2004).

1.2.2.4 Germination

Germination is the process in which an endospore is triggered to revert to a vegetative cell. Despite being metabolically dormant, the endospore can sense its environment and germination is triggered, starting the path to eventually lead to the endospore shedding its protective layers. After this process is completed, the cell then proceeds to outgrowth and finally becomes a vegetative cell (Moir, A., 2006).

The process of germination is still under study and is not yet fully understood. However, the consensus is that there are two major factors that can trigger the germination process; nutrient and non-nutrient signals, also known as germinants

(Ghosh, S. *et al.*, 2012). Whilst the protective layers of an endospore are designed to keep molecules out, the coat and outer membrane allow small molecules to pass through to the inner membrane, where receptors are present (Hudson, K. D. *et al.*, 2001). The next steps are poorly understood, but it is thought that there is a signal transduction mechanism present which then prompts the dormant endospore to resume metabolic processes, therefore germinating to a vegetative cell (Sinai, L. *et al.*, 2015).

Nutrient germinants are amino acids and sugars such as alanine and asparagine (Moir, A., 2006), whereas non-nutrient germinants are Ca^{2+}-dipicolinic acid (CaDPA) and dodecylamine (Setlow, B. *et al.*, 2003). As previously mentioned, receptors are present in the inner membrane of an endospore and as such the metabolism of nutrients is not required to trigger a response. Non-nutrient germination appears to be receptor independent, but this is currently not fully understood (Perez-Valdespino, A. *et al.*, 2013).

The presence of germinants triggers certain populations of endospores to germinate in a heterogeneous manner. This is proposed to be a bet-hedging strategy, where some endospores are termed as superdormant and remain in this metabolically inactive state despite the majority of the population germinating around them in an environment dependent manner (Ghosh, S. & Setlow, P., 2009). There is some debate as to the cause of this, as initially this was attributed to the levels of germination receptors (Ghosh, S. *et al.*, 2012), but this was later shown not to be the case (Zhang, J. Q. *et al.*, 2013).

The germination receptors in *B. subtilis* are located within the endospore inner membrane and form trimeric protein complexes (Hudson, K. D. *et al.*, 2001). The three main germination proteins in *B. subtilis* are GerA, GerB and GerK. GerA is responsible for sensing alanine whilst GerB and GerK respond to a mixture of sugars (glucose and fructose), ions such as K+ and the amino acid asparagine. These receptors have been found to colocalise to create a germinosome with the help of GerD (Wang, G. *et al.*, 2011). It is thought that the binding of a ligand to a nutrient

receptor causes a conformational change, since the application of pressure on an endospore is also capable of inducing germination in the absence of other stimuli (Doona, C. J. *et al.*, 2014)

After activation, the endospore's core releases a large amount of monovalent cations resulting in a rise in core pH (Swerdlow, B. M. *et al.*, 1981). After this, it is the nutrient receptors again that facilitate the next major step in germination. This is the rapid release of the core's inner store of DPA, which in turn allows the endospore core to rehydrate following cortex lysis. The DPA present in endospores is one of the contributing factors to an endospore's resistance to wet heat and as such this is one of the first resistance properties that is lost (Setlow, P., 2006).

The next step in endospore germination is cortex hydrolysis (Atrih, A. & Foster, S. J., 1999). Both SleB and CwlJ are cortex lytic enzymes required in *B. subtilis* for this process but both are partially redundant (Chirakkal, H. *et al.*, 2002). CwlJ is localised to the outer layers and SleB in the inner membrane and it is suggested that DPA is required for CwlJ activation (Paidhungat, M. *et al.*, 2001). It is the degradation of the cortex that finally allows the core to rehydrate and allow the mobility of lipids and proteins within.

With metabolism active the germinating endospore can start outgrowth and transcription can now take place. Some of the first proteins to appear in this ripening stage are those involved in DNA repair, followed by lipid synthesis and metabolism proteins (Sinai, L. *et al.*, 2015). The production of these proteins is part of a complex program of transcription, in a temporal manner (Keijser, B. J. *et al.*, 2007). The ripening stage involves no morphological changes and time taken during this period is dependent on the initial resources available within the dormant endospore (Segev, E. *et al.*, 2013). This is followed by cell elongation resulting in a vegetative cell.

1.2.3 sRNAs expressed during sporulation

sRNAs have been identified to be transcribed specifically during the sporulation process previously, but little information has been found to enable characterisation of their functions. A combination of microarrays and comparative genomics has been utilised, where the intergenic regions were probed first for activity where protein genes have not been annotated and secondly predictions of RNA which have conserved secondary structures (Silvaggi, J. M. *et al.*, 2006). Numerous sRNAs were discovered, with sporulation-specific promoters being allocated to them, but attempts to assign functions was unsuccessful. In a similar theme, Marchais, A. *et al.* (2011) utilised phylogenetic clustering to find regions of co-occurrence in which sporulation related genes and putative sRNAs were hypothesized to both be active during sporulation. Indeed, one such sRNA (CsfG) was found to be active during sporulation utilising *lac* promoter fusions. Furthermore, it was shown to have a sporulation specific promoter, being primarily activated by SigF early in the forespore. Whilst a sporulation defect was found in a deletion mutant when in competition with the wild-type, again a specific function of the sRNA was not identified. Potential target prediction of CsfG was not useful, with a prominent and conserved pyrimidine rich stem-loop having the potential to bind to a long list of targets. Nicolas P. *et al.*, (2012) included the conditions of germination and sporulation within their study, and as such may have identified putative sRNAs which are differentially regulated during sporulation. However, this was not the focus of their study and therefore no comment was made.

1.3 Aims

Many published papers have demonstrated the importance of sRNAs in fine tuning gene expression in diverse conditions by a variety of different mechanisms. However, whilst Nicolas *et al.* (2012) successfully identified many new RNA species, more investigation is required to confirm if these are indeed sRNAs. As such, given the large bank of putative sRNAs identified in *B. subtilis*, and the diverse methods of characterising such sRNAs, this study aims to build upon previous findings to identify

the sRNAs most likely to play a role in sporulation in *B. subtilis.* The overall aims of this thesis are:

1. To investigate the newly identified RNA species with *in-silico* tools, curating available data and generation of new predictions to identify those most likely to play a role.
2. To experimentally confirm expression of such sRNAs during the process of endospore formation. More specifically, to characterise activity and explore potential phenotypes.
3. To explore “omics” approaches for the identification of targets. In particular, to develop current proteomic approaches to provide a comprehensive endospore protein profile.

2 Chapter two: Materials and methods

2.1 Bacterial Growth conditions

Bacillus subtilis and *Escherichia coli* were routinely cultured in Lysogeny Broth (LB) (10 g tryptone, 5 g yeast extract and 10 g NaCl 1 l H_20) or Lysogeny Broth Agar (LBA) (with the addition of 1.5% agar) plates at 37°C. Culture in liquid media was carried out in an orbital shaker, agitated at 250 rpm unless otherwise stated. Where appropriate, growth medium was supplemented with the correct antibiotics: ampicillin (100 µg/ml), erythromycin (2 µg/ml), phleomycin (4 µg/ml), spectinomycin (100 µg/ml) or tetracycline (10 µg/ml).

2.1.1 Growth in LB

A colony of *B. subtilis* was picked from a freshly prepared LBA plate with appropriate antibiotics in to LB and grown for 16 hours. The culture was diluted 1:50 into LB and grown until $OD_{600\ nm}$ 0.4, when the culture was diluted into LB to a starting $OD_{600\ nm}$ 0.05 and growth was monitored in a plate reader (BMG Labtech FLUOstar Omega) with linear agitation at 600 rpm with measurements taken every 5 minutes.

2.1.2 Growth in M9

M9 media was prepared by combining 1 ml 100 mM $CaCl_2$, 10 ml M9 trace element solution, 1 ml 1 M $MgSO_4$, 1 ml 50 mM $FeCl_3$/100 mM $C_6H_8O_7$ solution and 200 ml M9 5x stock solution with 781 distilled H_2O. A colony of *B. subtilis* was picked from a LBA plate with appropriate antibiotics in to LB and grown for 16 hours. The culture was diluted 1:50 into LB and grown until $OD_{600\ nm}$ 0.4. The LB culture was diluted 1:20 into pre-warmed M9 glucose and grown to $OD_{600\ nm}$ 0.4. This culture was diluted again 1:20 into M9 and the appropriate carbon source (Pyruvate 0.3 g, fructose 0.15 g, glucose 0.15 g per 50 ml) the $OD_{600\ nm}$ was measured every 10 minutes by the plate reader (500 rpm linear shaking at 37°C).

2.2 Bacterial strains

Table 2 describes the strains used in this study. All plasmids were first transformed into NEB 5-alpha Competent *E. coli* then into DH5α before finally being transformed into *B. subtilis*.

Table 2 – *B. subtilis* strains used in this study

Strain	Description	Resistances	Source
168 trp+	Wild-type	None	
168 trp- Δ*sigE*	Chromosomal deletion of *sigE* in 168 trp-	Erythromycin	BGSC
168 trp+ Δ*sigE*	Chromosomal deletion of *sigE* in 168 trp+	Erythromycin	This study
168 trp- Δ*sigF*	Chromosomal deletion of *sigF* in 168 trp-	Erythromycin	BGSC
168 trp+ Δ*sigF*	Chromosomal deletion of *sigF* in 168 trp+	Erythromycin	This study
168 trp- Δ*sigG*	Chromosomal deletion of *sigG* in 168 trp-	Erythromycin	BGSC
168 trp+ Δ*sigG*	Chromosomal deletion of *sigG* in 168 trp+	Erythromycin	This study
168 trp- Δ*sigK*	Chromosomal deletion of *sigK* in 168 trp-	Erythromycin	BGSC
168 trp+ Δ*sigK*	Chromosomal deletion of *sigK* in 168 trp+	Erythromycin	This study
168 trp+ ΔS357	Chromosomal deletion of S357 in 168 trp+	Phleomycin	This study
168 trp+ ΔS547	Chromosomal deletion of S547 in 168 trp+	Phleomycin	This study
168 trp+ ΔS612	Chromosomal deletion of S612 in 168 trp+	Phleomycin	RAT Mars
168 trp+ ΔS849	Chromosomal deletion of S849 in 168 trp+	Phleomycin	This study
168 trp+ Δ*gerA*	Chromosomal deletion of *gerA* in 168 trp+	Spectinomycin	Setlow
168 trp+ Δ*gerB*	Chromosomal deletion of *gerB* in 168 trp+	Chloramphenicol	Setlow
168 trp+ P*spoIIQ*-*gfp*	Integrated in-locus promoter fusion of *spoIIQ* to GFP utilising pBaSysBioII integrative plasmid	Spectinomycin	This study
168 trp+ P*spoIID*-*gfp*	Integrated in-locus promoter fusion of *spoIID*	Spectinomycin	This study

	to GFP utilising pBaSysBioII integrative plasmid		
168 trp+ P*gerE-gfp*	Integrated in-locus promoter fusion of *gerE* to GFP utilising pBaSysBioII integrative plasmid	Spectinomycin	This study
168 trp+ PS357-*gfp*	Integrated in-locus promoter fusion of S357 to GFP utilising pBaSysBioII integrative plasmid	Spectinomycin	This study
168 trp+ PS547-*gfp*	Integrated in-locus promoter fusion of S547 to GFP utilising pBaSysBioII integrative plasmid	Spectinomycin	This study
168 trp+ PS612-*gfp*	Integrated in-locus promoter fusion of S612 to GFP utilising pBaSysBioII integrative plasmid	Spectinomycin	This study
168 trp+ PS849-*gfp*	Integrated in-locus promoter fusion of S849 to GFP utilising pBaSysBioII integrative plasmid	Spectinomycin	This study
168 trp+ Δ*sigE* P*spoIIQ-gfp*	Chromosomal deletion of *sigE* in 168 trp+ and integrated in-locus promoter fusion of *spoIIQ* to GFP utilising pBaSysBioII integrative plasmid	Erythromycin and Spectinomycin	This study
168 trp+ Δ*sigE* P*spoIID-gfp*	Chromosomal deletion of *sigE* in 168 trp+ and integrated in-locus promoter fusion of *spoIID* to GFP utilising pBaSysBioII integrative plasmid	Erythromycin and Spectinomycin	This study
168 trp+ Δ*sigE* P*gerE-gfp*	Chromosomal deletion of *sigE* in 168 trp+ and integrated in-locus promoter fusion of *gerE* to GFP utilising pBaSysBioII integrative plasmid	Erythromycin and Spectinomycin	This study
168 trp+ Δ*sigF* P*spoIIQ-gfp*	Chromosomal deletion of *sigF* in 168 trp+ and integrated in-locus promoter fusion of *spoIIQ* to GFP utilising pBaSysBioII integrative plasmid	Erythromycin and Spectinomycin	This study

168 trp+ Δ*sigF* P*spoIID-gfp*	Chromosomal deletion of *sigF* in 168 trp+ and integrated in-locus promoter fusion of *spoIID* to GFP utilising pBaSysBioII integrative plasmid	Erythromycin and Spectinomycin	This study
168 trp+ Δ*sigF* P*gerE-gfp*	Chromosomal deletion of *sigF* in 168 trp+ and integrated in-locus promoter fusion of *gerE* to GFP utilising pBaSysBioII integrative plasmid	Erythromycin and Spectinomycin	This study
168 trp+ Δ*sigG* P*spoIIQ-gfp*	Chromosomal deletion of *sigG* in 168 trp+ and integrated in-locus promoter fusion of *spoIIQ* to GFP utilising pBaSysBioII integrative plasmid	Erythromycin and Spectinomycin	This study
168 trp+ Δ*sigG* P*spoIID-gfp*	Chromosomal deletion of *sigF* in 168 trp+ and integrated in-locus promoter fusion of *spoIID* to GFP utilising pBaSysBioII integrative plasmid	Erythromycin and Spectinomycin	This study
168 trp+ Δ*sigG* P*gerE-gfp*	Chromosomal deletion of *sigF* in 168 trp+ and integrated in-locus promoter fusion of *gerE* to GFP utilising pBaSysBioII integrative plasmid	Erythromycin and Spectinomycin	This study
168 trp+ Δ*sigK* P*spoIIQ-gfp*	Chromosomal deletion of *sigF* in 168 trp+ and integrated in-locus promoter fusion of *spoIIQ* to GFP utilising pBaSysBioII integrative plasmid	Erythromycin and Spectinomycin	This study
168 trp+ Δ*sigK* P*spoIID-gfp*	Chromosomal deletion of *sigF* in 168 trp+ and integrated in-locus promoter fusion of *spoIID* to GFP utilising pBaSysBioII integrative plasmid	Erythromycin and Spectinomycin	This study
168 trp+ Δ*sigK* P*gerE-gfp*	Chromosomal deletion of *sigF* in 168 trp+ and integrated in-locus promoter fusion of *gerE* to	Erythromycin and Spectinomycin	This study

	GFP utilising pBaSysBioII integrative plasmid		
168 trp+ ΔS357 P*spoIIQ-gfp*	Chromosomal deletion of S357 in 168 trp+ and integrated in-locus promoter fusion of *spoIIQ* to GFP utilising pBaSysBioII integrative plasmid	Phleomycin and Spectinomycin	This study
168 trp+ ΔS357 P*spoIID-gfp*	Chromosomal deletion of S357 in 168 trp+ and integrated in-locus promoter fusion of *spoIID* to GFP utilising pBaSysBioII integrative plasmid	Phleomycin and Spectinomycin	This study
168 trp+ ΔS357 P*gerE-gfp*	Chromosomal deletion of S357 in 168 trp+ and integrated in-locus promoter fusion of *gerE* to GFP utilising pBaSysBioII integrative plasmid	Phleomycin and Spectinomycin	This study
168 trp+ ΔS547 P*spoIIQ-gfp*	Chromosomal deletion of S547 in 168 trp+ and integrated in-locus promoter fusion of *spoIIQ* to GFP utilising pBaSysBioII integrative plasmid	Phleomycin and Spectinomycin	This study
168 trp+ ΔS547 P*spoIID-gfp*	Chromosomal deletion of S547 in 168 trp+ and integrated in-locus promoter fusion of *spoIID* to GFP utilising pBaSysBioII integrative plasmid	Phleomycin and Spectinomycin	This study
168 trp+ ΔS547 P*gerE-gfp*	Chromosomal deletion of S547 in 168 trp+ and integrated in-locus promoter fusion of *gerE* to GFP utilising pBaSysBioII integrative plasmid	Phleomycin and Spectinomycin	This study
168 trp+ ΔS612 P*spoIIQ-gfp*	Chromosomal deletion of S612 in 168 trp+ and integrated in-locus promoter fusion of *spoIIQ* to GFP utilising pBaSysBioII integrative plasmid	Phleomycin and Spectinomycin	This study
168 trp+ ΔS612 P*spoIID-gfp*	Chromosomal deletion of S612 in 168 trp+ and integrated in-locus	Phleomycin and Spectinomycin	This study

	promoter fusion of *spoIID* to GFP utilising pBaSysBioII integrative plasmid		
168 trp+ ΔS612 P*gerE-gfp*	Chromosomal deletion of S612 in 168 trp+ and integrated in-locus promoter fusion of *gerE* to GFP utilising pBaSysBioII integrative plasmid	Phleomycin and Spectinomycin	This study
168 trp+ Δ*sigE* PS357-*gfp*	Chromosomal deletion of *sigE* in 168 trp+ and integrated in-locus promoter fusion of S357 to GFP utilising pBaSysBioII integrative plasmid	Erythromycin and Spectinomycin	This study
168 trp+ Δ*sigE* PS547-*gfp*	Chromosomal deletion of *sigE* in 168 trp+ and integrated in-locus promoter fusion of S547 to GFP utilising pBaSysBioII integrative plasmid	Erythromycin and Spectinomycin	This study
168 trp+ Δ*sigE* PS612-*gfp*	Chromosomal deletion of *sigE* in 168 trp+ and integrated in-locus promoter fusion of S612 to GFP utilising pBaSysBioII integrative plasmid	Erythromycin and Spectinomycin	This study
168 trp+ Δ*sigE* PS849-*gfp*	Chromosomal deletion of *sigE* in 168 trp+ and integrated in-locus promoter fusion of S849 to GFP utilising pBaSysBioII integrative plasmid	Erythromycin and Spectinomycin	This study
168 trp+ Δ*sigF* PS357-*gfp*	Chromosomal deletion of *sigF* in 168 trp+ and integrated in-locus promoter fusion of S357 to GFP utilising pBaSysBioII integrative plasmid	Erythromycin and Spectinomycin	This study
168 trp+ Δ*sigF* PS547-*gfp*	Chromosomal deletion of *sigF* in 168 trp+ and integrated in-locus promoter fusion of S547 to GFP utilising pBaSysBioII integrative plasmid	Erythromycin and Spectinomycin	This study

168 trp+ Δ*sigF* PS612-*gfp*	Chromosomal deletion of *sigF* in 168 trp+ and integrated in-locus promoter fusion of S612 to GFP utilising pBaSysBioII integrative plasmid	Erythromycin and Spectinomycin	This study
168 trp+ Δ*sigF* PS849-*gfp*	Chromosomal deletion of *sigF* in 168 trp+ and integrated in-locus promoter fusion of S849 to GFP utilising pBaSysBioII integrative plasmid	Erythromycin and Spectinomycin	This study
168 trp+ Δ*sigG* PS357-*gfp*	Chromosomal deletion of *sigG* in 168 trp+ and integrated in-locus promoter fusion of S357 to GFP utilising pBaSysBioII integrative plasmid	Erythromycin and Spectinomycin	This study
168 trp+ Δ*sigG* PS547-*gfp*	Chromosomal deletion of *sigG* in 168 trp+ and integrated in-locus promoter fusion of S547 to GFP utilising pBaSysBioII integrative plasmid	Erythromycin and Spectinomycin	This study
168 trp+ Δ*sigG* PS612-*gfp*	Chromosomal deletion of *sigG* in 168 trp+ and integrated in-locus promoter fusion of S612 to GFP utilising pBaSysBioII integrative plasmid	Erythromycin and Spectinomycin	This study
168 trp+ Δ*sigG* PS849-*gfp*	Chromosomal deletion of *sigG* in 168 trp+ and integrated in-locus promoter fusion of S849 to GFP utilising pBaSysBioII integrative plasmid	Erythromycin and Spectinomycin	This study
168 trp+ ΔS357 cS357	Chromosomal deletion of S357 in 168 trp+ and reintroduced with *amyE::*pRMC-S357	Phleomycin and Tetracycline	This study
168 trp+ ΔS547 cS547	Chromosomal deletion of S547 in 168 trp+ and reintroduced with *amyE::*pRMC-S547	Phleomycin and Tetracycline	This study
168 trp+ ΔS612 cS612	Chromosomal deletion of S612 in 168 trp+ and	Phleomycin and Tetracycline	This study

	reintroduced with *amyE::*pRMC-S612		

Table 3 – Plasmids used in this study

Plasmid	Description	Resistances	Source
pRMC	Cloning vector for the complementation of deletion mutants. Integrates into the *amyE* locus of *B. subtilis*	Tetracyline in *B. subtilis* and Ampicillin in *E. coli*	(Mars, R. A. *et al.*, 2015a)
pBaSysBioII	Cloning vector for transcriptional promoter fusions to GFP	Spectinomycin in *B. subtilis* Ampicillin *E. coli*	(Botella, E. *et al.*, 2010)
pRMC-S357	pRMC containing the entire predicted S357 gene including native promoter and terminator	Tetracyline in *B. subtilis* and Ampicillin in *E. coli*	This study
pRMC-S547	pRMC containing the entire predicted S547 gene including native promoter and terminator	Tetracyline in *B. subtilis* and Ampicillin in *E. coli*	This study
pRMC-S612	pRMC containing the entire predicted S612 gene including native promoter and terminator	Tetracyline in *B. subtilis* and Ampicillin in *E. coli*	This study
pBaSysBioII-PS111	pBaSysBioII containing the promoter region of S111 and up to 18 nucleotides after ATG	Tetracyline in *B. subtilis* and Ampicillin in *E. coli*	This study
pBaSysBioII-PS249	pBaSysBioII containing the promoter region of S249	Tetracyline in *B. subtilis* and	This study

	and up to 18 nucleotides after ATG	Ampicillin in *E. coli*	
pBaSysBioII-PS357	pBaSysBioII containing the promoter region of S357 and up to 18 nucleotides after ATG	Tetracyline in *B. subtilis* and Ampicillin in *E. coli*	This study
pBaSysBioII-PS423	pBaSysBioII containing the promoter region of S423 and up to 18 nucleotides after ATG	Tetracyline in *B. subtilis* and Ampicillin in *E. coli*	This study
pBaSysBioII-PS458	pBaSysBioII containing the promoter region of S458 and up to 18 nucleotides after ATG	Tetracyline in *B. subtilis* and Ampicillin in *E. coli*	This study
pBaSysBioII-PS547	pBaSysBioII containing the promoter region of S547 and up to 18 nucleotides after ATG	Tetracyline in *B. subtilis* and Ampicillin in *E. coli*	This study
pBaSysBioII-PS612	pBaSysBioII containing the promoter region of S612 and up to 18 nucleotides after ATG	Tetracyline in *B. subtilis* and Ampicillin in *E. coli*	This study
pBaSysBioII-PS665	pBaSysBioII containing the promoter region of S665 and up to 18 nucleotides after ATG	Tetracyline in *B. subtilis* and Ampicillin in *E. coli*	This study
pBaSysBioII-PS849	pBaSysBioII containing the promoter region of S849 and up to 18 nucleotides after ATG	Tetracyline in *B. subtilis* and Ampicillin in *E. coli*	This study

pBaSysBioII-PS877	pBaSysBioII containing the promoter region of S877 and up to 18 nucleotides after ATG	Tetracyline in *B. subtilis* and Ampicillin in *E. coli*	This study
pBaSysBioII-PS912	pBaSysBioII containing the promoter region of S912 and up to 18 nucleotides after ATG	Tetracyline in *B. subtilis* and Ampicillin in *E. coli*	This study
pBaSysBioII-PS1009	pBaSysBioII containing the promoter region of S1009 and up to 18 nucleotides after ATG	Tetracyline in *B. subtilis* and Ampicillin in *E. coli*	This study
pBaSysBioII-P*spoIIQ*	pBaSysBioII containing the promoter region of *spoIIQ* and up to 18 nucleotides after ATG	Tetracyline in *B. subtilis* and Ampicillin in *E. coli*	This study
pBaSysBioII-P*spoIID*	pBaSysBioII containing the promoter region of *spoIID* and up to 18 nucleotides after ATG	Tetracyline in *B. subtilis* and Ampicillin in *E. coli*	This study
pBaSysBioII-P*gerE*	pBaSysBioII containing the promoter region of *gerE* and up to 18 nucleotides after ATG	Tetracyline in *B. subtilis* and Ampicillin in *E. coli*	This study

2.3 Oligonucleotide primers used in this study

Table 4 lists the oligonucleotide primers used in this study, all of which were synthesised by Sigma Aldrich (UK).

Table 4 – Oligonucleotide primers used in this study

Primer	Sequence (5' to 3')	Function
S357 DM P0	gctggctgagacatacataat	Primer used to confirm deletion
S357 DM P1	tggaaaacaaaataaagaacg	Primer used to amplify the flanking upstream region of S357
S357 DM P2	CGACCTGCAGGCATGCAAGCTatgatgtccagaaccctgt	Primer used to amplify the flanking upstream region of S357 including homologous sequence to phleomycin cassette
S357 DM P3	CGAGCTCGAATTCACTGGCCGTCGctattttctgcatttctgtgg	Primer used to amplify the flanking downstream region of S357 including homologous sequence to Phleomycin cassette
S357 DM P4	aatgttttcgttaattcgtca	Primer used to amplify the flanking downstream region of S357
S547 DM P0	atgcttttatggaaacctctt	Primer used to confirm deletion
S547 DM P1	ttgtctgaatttttgctgaag	Primer used to amplify the flanking upstream region of S547
S547 DM P2	CGACCTGCAGGCATGCAAGCTtagtatggacaaggatcatgc	Primer used to amplify the flanking upstream region of S547 including homologous sequence to Phleomycin cassette
S547 DM P3	CGAGCTCGAATTCACTGGCCGTCGatagaaggaccccgcagat	Primer used to amplify the flanking downstream region of S547 including homologous sequence to Phleomycin cassette

S547 DM P4	aagatggatgtcaatcaaatg	Primer used to amplify the flanking downstream region of S547
S111 PF P1	CCGCGGGCTTTCCCAGCactttgaattgtgcttgtca	Primer used to amplify the promoter region of S111 containing homologous end to pBaSysBioII LIC cloning site
S111 PF P2	GTTCCTCCTTCCCACCtacaacaaagcgcactataag	Primer used to amplify the promoter region containing homologous end to pBaSysBioII LIC cloning site
S249 PF P1	CCGCGGGCTTTCCCAGCtcatttgtcacaatctcctga	Primer used to amplify the promoter region containing homologous end to pBaSysBioII LIC cloning site
S249 PF P2	GTTCCTCCTTCCCACCacatgtgataacgtcctttct	Primer used to amplify the promoter region containing homologous end to pBaSysBioII LIC cloning site
S357 PF P1	CCGCGGGCTTTCCCAGCtgttagttcaagatgctccat	Primer used to amplify the promoter region containing homologous end to pBaSysBioII LIC cloning site
S357 PF P2	GTTCCTCCTTCCCACCtttgacaaacacctcattttt	Primer used to amplify the promoter region containing homologous end to pBaSysBioII LIC cloning site
S423 PF P1	CCGCGGGCTTTCCCAGCagggttttggcatatatgtga	Primer used to amplify the promoter region containing homologous end to pBaSysBioII LIC cloning site
S423 PF P2	GTTCCTCCTTCCCACCcccttcctttccttatgtcat	Primer used to amplify the promoter region containing homologous end to pBaSysBioII LIC cloning site
S458 PF P1	CCGCGGGCTTTCCCAGCtgctgaaaacctatgctgata	Primer used to amplify the promoter region containing homologous end to pBaSysBioII LIC cloning site

S458 PF P2	GTTCCTCCTTCCCACCcacctg acattatcctctgtt	Primer used to amplify the promoter region containing homologous end to pBaSysBioII LIC cloning site
S547 PF P1	CCGCGGGCTTTCCCAGCaccg tttgaatcgctttaaaa	Primer used to amplify the promoter region containing homologous end to pBaSysBioII LIC cloning site
S547 PF P2	GTTCCTCCTTCCCACCggcaac aacatcctctaga	Primer used to amplify the promoter region containing homologous end to pBaSysBioII LIC cloning site
S612 PF P1	CCGCGGGCTTTCCCAGCga ttatgatgagctgcaaagg	Primer used to amplify the promoter region containing homologous end to pBaSysBioII LIC cloning site
S612 PF P2	GTTCCTCCTTCCCACCgataat gatcttcggaccaca	Primer used to amplify the promoter region containing homologous end to pBaSysBioII LIC cloning site
S665 PF P1	CCGCGGGCTTTCCCAGCtcttt taggagttgcgagaat	Primer used to amplify the promoter region containing homologous end to pBaSysBioII LIC cloning site
S665 PF P2	GTTCCTCCTTCCCACCggccaa atagcagattcacta	Primer used to amplify the promoter region containing homologous end to pBaSysBioII LIC cloning site
S849 PF P1	CCGCGGGCTTTCCCAGCagg aacccagacatataatgc	Primer used to amplify the promoter region containing homologous end to pBaSysBioII LIC cloning site
S849 PF P2	GTTCCTCCTTCCCACCatttccc gctctgtgatacta	Primer used to amplify the promoter region containing homologous end to pBaSysBioII LIC cloning site
S877 PF P1	CCGCGGGCTTTCCCAGCgctc catatgtttccgaagta	Primer used to amplify the promoter region containing homologous end to pBaSysBioII LIC cloning site

S877 PF P2	GTTCCTCCTTCCCACCttccaaattccagttactcgt	Primer used to amplify the promoter region containing homologous end to pBaSysBioII LIC cloning site
S912 PF P1	CCGCGGGCTTTCCCAGCattccatttcctcgttgtttg	Primer used to amplify the promoter region containing homologous end to pBaSysBioII LIC cloning site
S912 PF P2	GTTCCTCCTTCCCACCttcattcgtacgtctccttttt	Primer used to amplify the promoter region containing homologous end to pBaSysBioII LIC cloning site
S1009 PF P1	CCGCGGGCTTTCCCAGCtttcagttagcttcggtacat	Primer used to amplify the promoter region containing homologous end to pBaSysBioII LIC cloning site
S1009 PF P2	GTTCCTCCTTCCCACCcctcatacatcaccttcctt	Primer used to amplify the promoter region containing homologous end to pBaSysBioII LIC cloning site
SpoIIQ PF P1	acttttaccgcgggctttcccCCGCTGTTTTATTATAGGC	Primer used to amplify the promoter region containing homologous end to pBaSysBioII LIC cloning site
SpoIIQ PF P2	gcatagtagttcctccttcccTCTTTTCTTCCTCTCTCATTG	Primer used to amplify the promoter region containing homologous end to pBaSysBioII LIC cloning site
GerE PF P1	acttttaccgcgggctttcccGTGAAAACAAACAGACCATG	Primer used to amplify the promoter region containing homologous end to pBaSysBioII LIC cloning site
GerE PF P2	gcatagtagttcctccttcccTGAAATTCTTTCTCCTTCAAG	Primer used to amplify the promoter region containing homologous end to pBaSysBioII LIC cloning site
S357 CM P1	GGGTTCCTGGCGCGAGCcgattaagccgacaatttgat	Primer used to amplify the whole S357 gene, including promoter and terminator, and homologous end to pRMC LIC cloning site

S357 CM P2	TTGGGCTGGCGCGAGCgccggcaataattttacaaaa	Primer used to amplify the whole S357 gene, including promoter and terminator, and homologous end to pRMC LIC cloning site
S547 CM P1	GGGTTCCTGGCGCGAGCcgaatatacaatgctcaaacc	Primer used to amplify the whole S547 gene, including promoter and terminator, and homologous end to pRMC LIC cloning site
S547 CM P2	TTGGGCTGGCGCGAGCtaaaattggaattcgagtcaa	Primer used to amplify the whole S547 gene, including promoter and terminator, and homologous end to pRMC LIC cloning site
S612 CM P1	tcacagggttcctggGAGTGTCAGATTATAAGAAACGG	Primer used to amplify the whole S612 gene, including promoter and terminator, and homologous end to pRMC LIC cloning site
S612 CM P2	cactagattgggctggAGAAGCAGCTTAAAGCCG	Primer used to amplify the whole S612 gene, including promoter and terminator, and homologous end to pRMC LIC cloning site

2.4 *B. subtilis* genetic manipulation techniques

2.4.1 Preparation of chemically competent *E. coli*

The *E. coli* strain DH5α was routinely used for cloning and was made chemically competent using calcium chloride. A fresh colony of DH5α was cultured in 5 mL LB broth at 37˚C with 200 rpm shaking overnight. This was then diluted 1/100 in fresh LB until OD_{600} 0.4 at 37˚C at 200 rpm. Culture was collected after incubation on ice for 30 minutes and pelleted by centrifugation. The pellet was then rinsed in 1/5 volume 0.1 M $CaCl_2$ twice. Finally, the cells were gently re-suspended in 1/25 volume of 0.1 M CaCl2 and 15% glycerol solution before storage at -80 ˚C. For plasmid transformation of *B. subtilis*, plasmids were passaged through *E. coli* strain TG1. Competent TG1 *E. coli* were prepared using *Mix & Go E. coli* Transformation Kit (Zymo Research) as per the manufacturers protocol.

2.4.2 Genetic Transformation of *B. subtilis*

B. subtilis was routinely transformed using Paris Medium (PM), as previously described (Harwood, C. R., and S. M. Cutting. , 1990). *B. subtilis* was grown overnight in PM at 37℃ with shaking. The culture was diluted 50-fold into fresh, warm PM and incubated for three hours at 37℃ with shaking. Chromosomal DNA, plasmid DNA, or PCR products were added and incubated for a further 5 hours. To select for transformants the culture was plated on LBA plates containing the appropriate antibiotics.

2.4.3 DNA extraction

Whole genomic DNA extraction of *B. subtilis* was achieved using the phenol-chloroform method. Aliquots (2 ml) of overnight culture was pelleted by centrifugation, re-suspended in 500 µl of solution A with 0.5 mg/ml lysozyme, and incubated at 37℃ for 10 minutes. Aliquots (20 µl) of RNase A (20 mg/ml), 20 µl proteinase K (20 mg/ml) and 25 µl 10% SDS was added and incubated at 37℃ for 10 minutes and then 60℃ for a further 45 minutes. 1 volume of Phenol/Chloroform/IAA (PCI) was added, vortexed and the aqueous phase separated by centrifugation for 10

minutes at 16,000 xg and the upper aqueous phase transferred into a fresh tube. This was repeated two times in total. 0.1 volume of 3 M NaAC (pH 5.2) and 1 volume of 96% ethanol was added and inverted until the DNA condensed. The DNA was pelleted by centrifugation for 5 minutes. The supernatant was discarded and washed with 70% ethanol before further centrifugation for 2 minutes. All ethanol was removed, the pellet dried before re-suspending in Tris-EDTA buffer.

2.4.4 Isolation of *B. subtilis* deletion mutants

Deletion of regions of the *B. subtilis* chromosome was achieved using the method described previously (Fabret, C. *et al.*, 2002; Tanaka, K. *et al.*, 2013). Briefly, fragments of approximately 1500 bp were amplified from upstream and downstream of the region to be deleted by PCR using Ex Taq polymerase (Takara Biosciences) or Phusion (New England Biolabs) as per the manufacturer's instructions. The primers closest to the region to be deleted were designed with regions overlapping sequences at the ends of a Phleomycin drug resistance cassette: *phleo*^{R}. The PCR products were purified using a Illustra GFX PCR DNA and Gel Band Purification Kit (GE Healthcare Life Sciences). The upstream and downstream PCR products were mixed with the phleomycin cassette and the furthest upstream and downstream primers in an overlapping PCR. The resulting PCR product was used to transform *B. subtilis* as described in section 2.4.2 and the resulting transformants were selected on LBA containing phleomycin. To determine for correct insertion of the phleomycin cassette, DNA was extracted from the transformants and analysed by PCR using primers either side of the gene used to amplify fragments. A size difference between the gene and the cassette indicated the insertion of the phleomycin cassette in the region homologous to the upstream and downstream fragments.

2.4.5 Complementation of *B. subtilis* deletion mutants

Reintroduction of sRNAs into the chromosome under the control of their own promoters was achieved using the plasmid pRMC (Mars, R. A. *et al.*, 2015a) that integrates in to *amyE* locus. The entire sequence of the sRNA, including the promoter

region and terminator were PCR amplified using Phusion (NEB) and introduced into the plasmid by Gibson Assembly (NEB) at the *Asc*I restriction site. Plasmids were isolated using GenElute™ Plasmid Miniprep Kit (Sigma-Aldrich) per the manufacturer's instructions including optional wash step. Restriction digest and Sanger sequencing confirmed the correct insertion of the insert in to the vector. The plasmid was transformed into *B. subtilis*. Correct integration at the *amyE* locus was determined by loss of α-amylase activity as shown using 1 % starch LBA plates and iodine staining.

2.4.6 Construction of promoter fusions to genes and sRNAs

Promoter fusions were constructed using the integrative pBaSysBioII plasmid as previously described (Botella, E. *et al.*, 2010). Promoter regions for genes included regions 400 bp to 18 bp upstream of the start codon. Promoter regions for sRNAs were chosen based on the transcription start site defined by Nicolas, P. *et al.* (2012) and the reverse primer was designed to start approximately 25 bp after the start of the sRNA. Promoter regions were PCR amplified using Phusion (NEB) as per the manufacturer's instructions. The pBaSysBioII plasmid was digested using the restriction enzyme *Sma*I. The insert and plasmid were joined via Gibson assembly. Plasmids were isolated using GenElute™ Plasmid Miniprep Kit (Sigma-Aldrich) per the manufacturer's instructions including optional wash step. Restriction digest and Sanger sequencing confirmed the correct insertion of the insert in to the vector.

2.5 *B. subtilis* sporulation experiments

2.5.1 Media for sporulation experiments

2.5.1.1 2x SG

Richer sporulation medium. 2x SG was prepared by combining 16 g of Difco Nutrient Broth, 2 g KCl, 0.5 g $MgSO_4 \cdot 7H_2O$ per litre of distilled water, pH was adjusted to 7.0 and sterilised by autoclaving. Once cooled to 50 ˚C, 1 ml each of 1 M $Ca(NO_3)_2$, 0.1 M $MnCl_2 \cdot 4H_2O$, 1 mM $FeSO_4$ and 50 % w/v glucose is added.

2.5.1.2 CH Medium

10 g Casein hydrolysate, 3.68 g L-glutamic acid, 1.25 g L-alanine, 1.39 g L-asparagine, 1.36 g KH_2PO_4, 1.34 g NH_4Cl, 0.11 g Na_2SO_4, 0.10 g NH_4NO_3 and 1 mg $FeCl_3$ added to 940 ml ddH_2O water, pH was adjusted to 7 with 10 M NaOH. This was split into 94 ml aliquots and autoclaved. On the day of culture, 40 µl of 1 M $MgSO_4$, 1 ml of 100 mM $CaCl_2$ and 200 µl of 50 mM $MnSO_4$ was added to an aliquot and sterile dd H_2O was added to 100 ml.

2.5.1.3 Resuspension medium (RM)

Solution A – 0.096 g $FeCl_3.6H_2O$, 0.830 g $MgCl_2.6H_2O$ and 1.979 g of $MnCl_2.4H_2O$ and ddH_2O to 100 ml. Sterilized by autoclaving.

Solution B – 53.5 g NH_4Cl, 10.6 g Na_2SO_4, 6.8 g KH_2PO_4 and 9.7 g NH_4NO_3, 1000 ml dd H_2O (pH 7.0 adjusted with 1 M NaOH). Filter sterilise and store at 4˚C.

Sporulation salts – 1 ml Solution A and 10 ml solution B in 1000 ml of sterile dd H_2O.

RM – 90 ml sporulation salts, 4 ml glutamic acid, 1 ml 100 mM $CaCl_2$, 4 ml 1 M $MgSO_4$ and 1 ml dd H_2O.

2.5.2 Sporulation

2.5.2.1 Sporulation in nutrient rich media

Fresh colonies were picked from LBA plates and was used to make a LB overnight culture with appropriate antibiotics. This was then diluted 1:100 into freshly made 2x SG medium. This was then incubated at 37˚C with agitation until $OD_{600\ nm}$ 0.4, where it was further diluted to $OD_{600\ nm}$ 0.1 and this was then the start of the growth prior to sporulation. The end of exponential growth was designated t_0.

2.5.2.2 Sporulation in chemically defined media

A colony of *B. subtilis* was suspended in 200 µl of CH medium. 5 µl of this suspension was inoculated into 5 ml CH medium and was grown at 37˚C with

shaking. The culture was grown for between 6 and 8 hours to ensure the culture did not enter too far into stationary phase. The pre-pre-culture was diluted 1 in 50 and grown to an $OD_{600\ nm}$ of 0.2 and then further diluted to $OD_{600\ nm}$ of 0.1. The culture was then grown to an $OD_{600\ nm}$ of 0.6. The culture was pelleted by centrifugation and the resulting pellet was suspended in an equal volume of RM media. This designated hour 0 of the sporulation process.

2.5.3 Monitoring GFP expression of promoter fusions

Sporulation of strains containing GFP promoter fusions was conducted utilising the chemically defined method above. All GFP experiments were performed in a 24 or 48 well plates (Greiner CELLSTAR multiwall culture plates) with measurements taken every 10 minutes. Both the OD and GFP fluorescence (excitation 485nm and emission 509 nm) measurements were taken. Correction of path length was carried out by plate reader software. Data processing included removal of background fluorescence, judged by the wild-type with no GFP. Arbitrary transcriptional activity units (TAU) represent the increase of GFP from one time point to another, where the total level of GFP from the previous measurement (t-1) was subtracted from the next timepoint (t): $GFP^{t} - GFP^{t-1}$.

2.5.4 Purification of endospores

A fresh colony was suspended in 500 ml 2x SG liquid medium and incubated for 48 hours at 37°C shaking. The culture was pelleted by centrifugation, 10 000 g for 10 minutes. Cells were then suspended in 1/10 volume of 50 mM tris (pH 7.8) and 50 µg/ml lysozyme and incubated for 1 hour at 37°C. Cells were again pelleted by centrifugation and washed once with 0.05% SDS and a further three times with sterile dd H_2O. Purity of the endospore crop was determined via microscopy and deemed acceptable if free of debris and vegetative cells.

2.5.5 Endospore resistance testing

1 OD units of purified endospores were used for treatment. Treatment was administered and then serially diluted. 5 µl of the serial dilutions from 10^{-1} to 10^{-8}

was spotted onto LB agar and incubated overnight. Ethanol treatment was administered to 50% ethanol and incubated at 65˚C 1 hour. Heat treatment was performed as per Harwood, C. R., and S. M. Cutting. (1990). Chloroform was added to a total volume of 25% as per Harwood, C. R., and S. M. Cutting. (1990). Hydrogen peroxide was added and KPO_4 buffered as described by Melly, E. *et al.* (2002).

2.5.6 Germination capabilities of purified endospores

Germination experiments were performed as per Harwood, C. R., and S. M. Cutting. (1990) with the following modifications. Briefly, heat-activated endospores were diluted to a final OD of 1 in 10 mM Tris-HCl (pH 8.4) in a 96 well plate. For germination with AGFK, a mixture of 100 mM L-asparagine, 10 mM D-glucose, 10 mM D-fructose and 100 mM KCl was included. For germination with alanine 100 mM of L-alanine was used. Subsequent readings of $OD_{600\ nm}$ at 37°C was taken every 5 minutes until equilibrium was reached. Percentage loss of OD was calculated using the following equation.

$$\%\ loss\ of\ OD\ at\ time\ t\ (t_t)\ = \left[1 - \left(\frac{OD\ at\ t_t}{OD\ at\ t_1}\right)\right] \times\ 100$$

2.5.7 Electron Microscopy of purified endospores

EM of purified endospores was adapted from Laue, M. *et al.* (2007). Purified endospores were suspended in 0.05 M HEPES (pH 7.2), 10% formaldehyde, 0.05% glutaraldehyde and incubated for 2 hours at room temperature. The endospores were washed twice in HEPES and once in ddH_2O and incubated at room temperature for 10 minutes each. The endospores were then dehydrated in a series of ethanol at 4°C for 5 minutes each, utilising 70% ethanol first followed by twice with 100% ethanol. Endospores were incubated overnight in anhydrous ethanol and then overnight again in 50% epoxy resin and propylene oxide and 100% epoxy resin. Polymerisation was carried out at 60°C for 48 hours. EM was performed by the University of Warwick School of Life Sciences EM services.

2.6 *In-silico* methodology

2.6.1 DBTBS

The Database of Transcriptional Regulation in *Bacillus subtilis* (DBTBS) advanced search was utilised (http://dbtbs.hgc.jp) to predict transcriptional search sites via weight matrix search of sequences. 200 nucleotides upstream and 100 nucleotides downstream of the start of putative sRNAs as defined by (Nicolas, P. *et al.*, 2012). All sigma factors and transcription factors were selected to search for with a threshold p-value of 1%.

2.6.2 GLASSgo

Homologs of putative sRNAs were identified utilising GLASSgo version 1.4.2 which is part of part of Freiburg RNA online tools (http://rna.informatik.uni-freiburg.de/GLASSgo/Input.jsp). Parameters were left to automatic, which results as global taxon selection, a maximum allowed E-value of 1 and minimum allowed identity of 58%. Structure based filtering was left to the automatic setting

2.6.3 Heatmap

A binary of presence/absence was created and utilised to create a heatmap. Heatmap was created by using "heatmaply" (Galili, T., 2017) with seriation option "OLO" and plotly (Inc., Plotly Technologies, 2015) in R was used to export the image.

2.6.4 Folding

RNAfold from the Vienna RNA package (Lorenz, R. *et al.*, 2011) was utilised for secondary structure predications for putative sRNA regions. For consensus folding, first all sequences were aligned using ClustalW. The output of which was then used with RNAalifold, part of the Vienna RNA package, which is a program for the detection of conserved RNA secondary structures. The folding algorithm is essentially the same as RNAfold, apart from it creates a covariance score, in which it calculates the fraction of consensus pairs having a compensatory or consistent mutation.

2.6.5 Weblogo

Visual representation of multiple sequence alignments were made utilising WebLogo source code (Crooks, G. E. *et al.*, 2004). WebLogo displays conservation of a nucleotide as height, the greater the height the greater the conservation measured in bits.

2.6.6 Reanalysis of RNAseq data

RNAseq expression data was downloaded from the Gene Expression Omnibus (GEO) database (http://www.ncbi.nlm.nih.gov/geo/) under the accession number GSE81238. The reference index was created by indexing the reference NCBI genome NC_000964.3 by bowtie2-build command (Langmead, B. & Salzberg, S. L., 2012). This was then followed by mapped utilising bowtie2 software with very sensitive flag which produces a SAM (Sequence alignment/map) file of read alignment data. SAMtools was then utilised to first convert this into a binary SAM (BAM) file and subsequently SAMtools was then used to sort and index BAM files ready for visualisation with ARTEMIS.

BedTools coverage tool was then used to compute depth of coverage for each coding gene, with a modified general feature format (gff) file containing all features identified by Nicholas *et. al.* being the reference. The output files contained a tally of reads per gene.

For the CodY analysis, the R package DESeq2 was used which determines the differential gene expression in RNAseq data. Sorted BAM files and a GFF file featuring all the genes identified from Nicholas *et al.* was used as input files for BedTools coverage bed, which in turn was used as input for DESeq2. P values were adjusted for a FDR of 5% using the Benjamini-Hochberg method.

2.6.7 TNseq

The pipeline for the analysis of high-resolution transposon-insertion sequences technique (ARTIST) was utilised in Matlab. First, a genome map of *B. subtilis* insertion sites was created using 'genome_paser_TA' script and all the reads mapping to TA sites was counted utilising 'SAMreader_TA' function. This was repeated for three time points, prior to sporulation, hour 5 of sporulation and after 24 hours.

2.7 Proteomics analysis of *B. subtilis* endospores

2.7.1 Endospore protein extraction

A fresh colony was suspended in 2x SG liquid medium and incubated for 48 hours at 37°C with shaking. The cells were pelleted by centrifugation at 10,000 g for 10 minutes. Cells were then suspended in 50 mM Tris (pH 7.8) and 50 µg/ml lysozyme for 1 hour at 37°C. The cells were pelleted by centrifugation and washed first in 0.05% SDS and then in water three times. The purity of the endospore crop was determined via microscopy and deemed acceptable if free of debris.

15 OD units of endospores were re-suspended in 1 ml of de-coating solution (4% SDS, 0.1 M DTT, 0.1 M NaOH and 0.1 M NaCL) and incubated for 1 at 70 °C. The cells were pelleted by centrifugation for 2 minutes at max speed. The supernatant was transfer into fresh tube and retained for MetOH precipitation. The remaining pellet was then suspended in 10 mM tris (pH 7.8) and 1 mg/ml lysozyme, and incubated for 1.5 hours at 37 °C (200 µl total volume). Phase-contrast microscopy was then used to ensure the presence of phase dark endospores to confirm the treatment had removed all coat layers.

2.7.2 SDS-Polyacrylamide gel electrophoresis for the visulation of protein

Laemmli buffer was added to the protein extraction and incubated for 10 minutes at 100˚C. Protein was then loaded on to a 4-15% Mini-PROTEAN TGX gel and ran for 30 minutes at 145 V. The gel was then stained for 2 hours using InstantBlue protein

stain. The gel was then washed with water before proceeding with in-gel protein digestion

2.7.3 Concentration of protein by MetOH precipitation for proteomics

Protein sample was mixed with 4 volumes of MetOH and vortexed thoroughly for 1 min. 1 volume of chloroform was added and again mixed thoroughly via vortex for one minute. 3 volumes of water were then added and vortexed for a further minute. After this, the mixture was separated by centrifugation for 1 minute at max speed. The aqueous phase was discarded and 4 volumes of MetOH was added. This was then vortexed for 1 minute and further separated by centrifugation for 5 minutes at max speed. All MetOH was then removed, the pellet was air dried and the resulting proteins were then suspended in loading buffer.

2.7.4 Protein In-gel trypsin digest

Protein gels were sliced into cubes of 2 – 4 mm and transferred into 1.5 ml Eppendorf tubes. ABC buffer (50 % v/v ethanol and 50 mM ammonium bicarbonate) was added to cover the gel and incubated at 55˚C for 20 minutes shaking at 650 rpm, after which the solution was discarded. This was repeated at least three times until all blue staining was removed. 100 % ethanol was then added and incubated for 5 minutes at room temperature. Gel pieces were then dehydrated with 100% ethanol for 5 minutes at 55 °C after which the liquid was discarded. The gel pieces were then reduced in 10 mM DTT in 50 mM ammonium bicarbonate and incubated for 45 minutes at 56 °C. Gel pieces were allowed to cool to room temperature to avoid protein modification by IAA. Alkylation of cysteine residues was achieved by addition of 55 mM IAA in 50 mM ammonium bicarbonate for 30 minutes at room temperature in the dark. Next, a wash with 50% ethanol in 50 mM ammonium bicarbonate was performed and incubated for 20 minutes at 55 °C with shaking. Waste was discarded and the wash with 50% ethanol in 50 mM ammonium bicarbonate was repeated. Gel pieces were then dehydrated again with 100% ethanol for 5 minutes at 55 °C with shaking. Liquid was removed and 40 µl of 2.5 ng/µL trypsin was added. Gel pieces then were allowed to rehydrate for 10 minutes

on ice and before an additional 50 mM ammonium bicarbonate/trypsin solution was added to ensure gel pieces were sufficiently covered. This was incubated at 37 °C with shaking overnight.

The following day, an equal volume to the digestion volume was added of 5% formic acid in 25% acetonitrile to stop reaction. The gel pieces were then sonicated for 10 minutes and the supernatant was transferred to a fresh tube. This process was repeated a further 2 times by adding new formic acid/acetonitrile solution and adding the supernatant to the fresh tube. The resulting volume was then reduced using a speed-vac at 40 °C for 3 hours. Once dry, the samples were suspended in TFA.

A stage tip was prepped by centrifugation at 2000 rpm for 2 minutes with each solution; first 50 µl MetOH, followed by 50 µl Acetonitrile, then 50 µl of 2.5% acetonitrile and 0.05% trifluoroacetic acid. After preparing the stage tips, the samples were centrifuged through and captured. Flow-through was discarded and the stage tip was further washed with 50 µl of 2.5% AcN/TFA solution and flow through was discarded. In a fresh tube, the peptides were eluted with 50 µl of 50% AcN/TFA.

An aliquot of 15 µL of extracted peptides from each sample was analysed by means of nanoLC-ESI-MS/MS using the Ultimate 3000/Orbitrap Fusion instrumentation (Thermo Scientific) using a 120 minute LC separation on a 25-cm column.

2.7.5 Proteomics data *in silico* analysis

The data were used to interrogate the *Bacillus subtilis* database (http://www.uniprot.org/proteomes/UP000001570) MaxQuant software was used for protein identification and quantification. Scaffold software was used to visualise and statistically analyse the data generated by MaxQuant.

Perseus software (Tyanova, S. *et al.*, 2016) was then used from the output of MaxQuant software for differential protein analysis.

2.8 Sample preparation and library preparation for RNA-RNA interactome analysis

Sporulation was initiated utilising the minimal medium resuspension protocol. Samples were taken at 4 hours after resuspension. Crosslinking was performed by rinsing cells in the same volume of ice-cold PBS twice by pelleting by centrifugation and resuspension. The cells were then suspended in 2 ml PBS with a final concentration of 0.7 mM AMT. This was incubated for 10 minutes at 37°C and transferred to a 6 well plate for irradiation on ice for 10 minutes with 365 nm UV (0.120 Jcm^{-2}). Cells were then pelleted by centrifugation and suspended in 1ml of ice-cold killing buffer for cell lysis.

Cell lysis was performed via mechanical disruption with FastPrep-24 5G homogeniser (MP Biomedicals) in FastPrep R matrix tubes containing 1 mg of lysing matrix B beads. Cells were split into two matrix tubes and 3 cycles of homogenisation was performed at 6.0 m/s for 40 seconds each, with 5 minutes pause time between cycles on dry ice. 1 ml lysis buffer was then added and the samples was then RNA extracted. RNA was extracted by RNA-chloroform extraction as described previously (Nicolas, P. *et al.*, 2012). Turbo DNase was used as per manufacturers 'rigorous' protocol. Depletion of rRNA by Ribo-Zero rRNA Removal kit for bacteria (Illumina) as per the manufacturer's instructions, with ethanol precipitation for purification of RNA. Subsequent processing was performed by Adam Callan-Sidat.

Efficiency of rRNA removal was checked utilising Bioanalyzer RNA 600 Pico kit. S1 digestion was performed by incubation of 400 ng RNA with 2 µl of S1 nuclease in 20 µl total volume and incubated for 30 minutes at room temperature. Sample was RNA extracted by addition of 80 µl water and 100 µl PCI. This was then separated by centrifugation at 12,000 rpm at room temperature. Aqueous phase was then transferred to a fresh tube for precipitation with ethanol. 200 µl ethanol, 10 µl 3 M sodium acetate (pH 5.2) and 2.5 µl of 5 mg/ml linear acrylamide was added and precipitation was performed overnight at -20°C. Centrifugation was performed at

maximum speed, 4°C and supernatant was discarded. RNA pellet was washed with 75% ethanol and further pelleted via centrifugation before suspension in ddH_2O. RNA fragments were then ligated utilising CircLigase ssDNA ligase (Epicentre). 50 ng of RNA was incubated for 2 minutes at 85°C in 1x CircLigase ssDNA ligase buffer without the addition of $MnCl_2$. Then incubated for 1 hour at 60°C with the addition of 1 mM ATP and 1 µl CircLigase. RNase R treatment was performed with half a unit of RNase R (Epicentre) and incubated for 10 minutes at 37°C, then placed on ice. RNA was purified with RNA extraction and ethanol precipitation as before, before performing reverse crosslinking. RNA was suspended in Tris-buffered saline (pH 7.4) and irradiated at UV 254 nm for 5 minutes on ice. TruSeq Stranded mRNA sample preparation low-sample protocol was used to prepare the resulting RNA for sequencing which was carried out on an Illumina MiSeq.

2.8.1 Bioinformatic analysis for RNA-RNA interactome

Raw reads were processed utilising the RIL-seq scripts freely available on github (https://github.com/asafpr/RILseq) by first using simple mapping, map_single_fragments.py script. Subsequently, chimera mapping using map_chimeric_fragments.py. Artemis was used to view mapping and excel was used to view chimeric fragments.

3 Chapter three: Selection and *in-silico* characterisation of putative *trans*-encoded small regulatory RNAs in *Bacillus subtilis*

3.1 Introduction

3.1.1 sRNAs: their role and mechanisms of action

Many factors affecting translation such as post-transcriptional regulation have been identified in the past decade (Waters, L. S. & Storz, G., 2009). Different RNA molecules, such as small RNAs (sRNAs), have been shown to have roles in post-transcriptional regulation. As highlighted in chapter one, these regulatory elements can control various activities within the cell, including stabilisation of mRNA transcripts, the translation process, and even the activity of some proteins (Hartzog, G. A. & Martens, J. A., 2009).

The most in-depth studies of the role sRNAs play in bacterial cell biology have been carried out in Gram-negative bacteria, mainly from within the Enterobacteriaceae family. However, many similarities are present between Gram-positive and Gram-negative sRNAs. sRNAs are typically short: approximately 50 – 400 nucleotides in length. sRNAs are generally not translated, though there are exceptions; with some sRNAs encoding peptides, such as the bifunctional sRNA SrgS and its peptide counterpart SgrT (Lloyd, C. R. *et al.*, 2017). *Trans*–encoded sRNAs are transcribed distant from their mRNA target(s), many regulating via imperfect base pairing (Richards, G. R. & Vanderpool, C. K., 2011; Waters, L. S. & Storz, G., 2009). As the field of bacterial sRNAs has been gaining momentum, many sRNA prediction methods and large-scale studies searching for sRNAs have emerged (Rivas, E. & Eddy, S. R., 2001; Zhang, A. *et al.*, 2003; Silvaggi, J. M. *et al.*, 2006; Vogel, J. & Sharma, C. M., 2005; Waters, S. A. *et al.*, 2017; Melamed, S. *et al.*, 2016).

With the advances being made in large-scale RNomics studies and algorithms for differential expression analysis, the small number of sRNA candidates to look at is now an issue of the past and no longer the result of serendipity. The development of effective statistical methods for characterisation is primarily hindered when studying sRNAs as they lack known sequence information or bias to facilitate a reliable identification. Since there are no definite characteristics, sRNA prediction methods are constantly evolving as new sRNAs and their modes of action are being

discovered (Tjaden, B. *et al.*, 2006; Backofen, R. & Hess, W. R., 2010). As such, with so many potential new regulatory elements, searching for those that are likely to be the most fruitful is advantageous.

3.1.2 The many putative sRNAs of *B. subtilis*

B. subtilis has been shown to produce a compendium of RNA species that did not appear in the original annotation of the genome (Nicolas, P. *et al.*, 2012; Kunst, F. *et al.*, 1997). The most in-depth study on sRNAs in *B. subtilis* was carried out by Nicolas *et al.* in 2012. This study documented the transcriptional landscape of the wild-type laboratory strain, *B. subtilis* 168 trp+, using tiling arrays. A total of 104 different conditions were analysed, which identified 1583 previously unannotated RNA species (Nicolas, P. *et al.*, 2012). The identified RNA features were subdivided into eight classes and of interest for the search of *trans*-acting sRNAs are the classes Indep and Indep-NT. Each new RNA feature was given a sequential number and called S segments. Indep indicates the segments displayed their own upshift and downshift in expression. Indep-NT classification indicated that an identified upshift was seen but no downshift, or obvious termination occurred. This can be used as a good indicator for a *trans*-acting sRNA as *trans*-encoded sRNAs are independent RNA species, under the control of their own promoter.

The incorporation of many different conditions in the Nicolas *et al.* study showed that many of the putative sRNAs are differentially regulated. One of the conditions featured in the study was the process of endospore formation. Sporulation is a major part of the life cycle of *B. subtilis* and involves careful regulatory cascades to overhaul normal cell morphology (Steil, L. *et al.*, 2005). Previous studies have already identified several sRNAs being expressed during sporulation in *B. subtilis* (Silvaggi, J. M. *et al.*, 2006; Schmalisch, M. *et al.*, 2010) although no role for these sRNAs in the process of sporulation has yet been identified. Nicolas *et al.* showed that 36% of the variability of gene expression in *B. subtilis* was attributed to sporulation when comparing all the transcriptome profiles. Such large contribution to variability in

gene expression means that studying sporulation is a clear step towards trying to better understand the flexibility of gene expression.

3.1.3 Aims and objectives

The aim of this chapter was to identify potentially functional sRNAs active in sporulation in *B. subtilis* via bioinformatics analysis to determine which sRNAs warranted further experimental study. Strategies were determined as described in the following sections. Evidence of the expression of an RNA is not sufficient to infer the likelihood of functional relevance since transcription could be spurious or 'noise' from other RNAs (Wade, J. T. & Grainger, D. C., 2014; Lybecker, M. *et al.*, 2014).

RNA is single-stranded but can base pair with itself to create secondary and tertiary structures. This folding gives rise to loops and stems, therefore sequestering parts of an RNA. It is thought that these structures yield regions of an RNA molecule which are more accessible and, therefore, more available for binding (Bohn, C. *et al.*, 2010). The ease by which genomes can be sequenced has increased the number available for analysis. As such, conserved structural predictions can become more informative and confident, although reliable *de novo* prediction of sRNAs is still not possible (Rivas, E. *et al.*, 2001; Lorenz, R. *et al.*, 2011). Combinational approaches can then be employed to reduce the likely false positive sRNAs. For example, by combining secondary structure prediction with comparative genomics of regions with known expression, sequence homologies can be supported with structural biases. For this study criteria were selected to rank the putative sRNAs, with points awarded for desirable characteristics. These are displayed in Figure 6 and discussed in the following sections.

In addition, sporulation-related next generation sequencing data will be identified and reanalysed to see if any putative sRNAs are involved in the process of sporulation, as many studies to date have only focused on protein-coding genes.

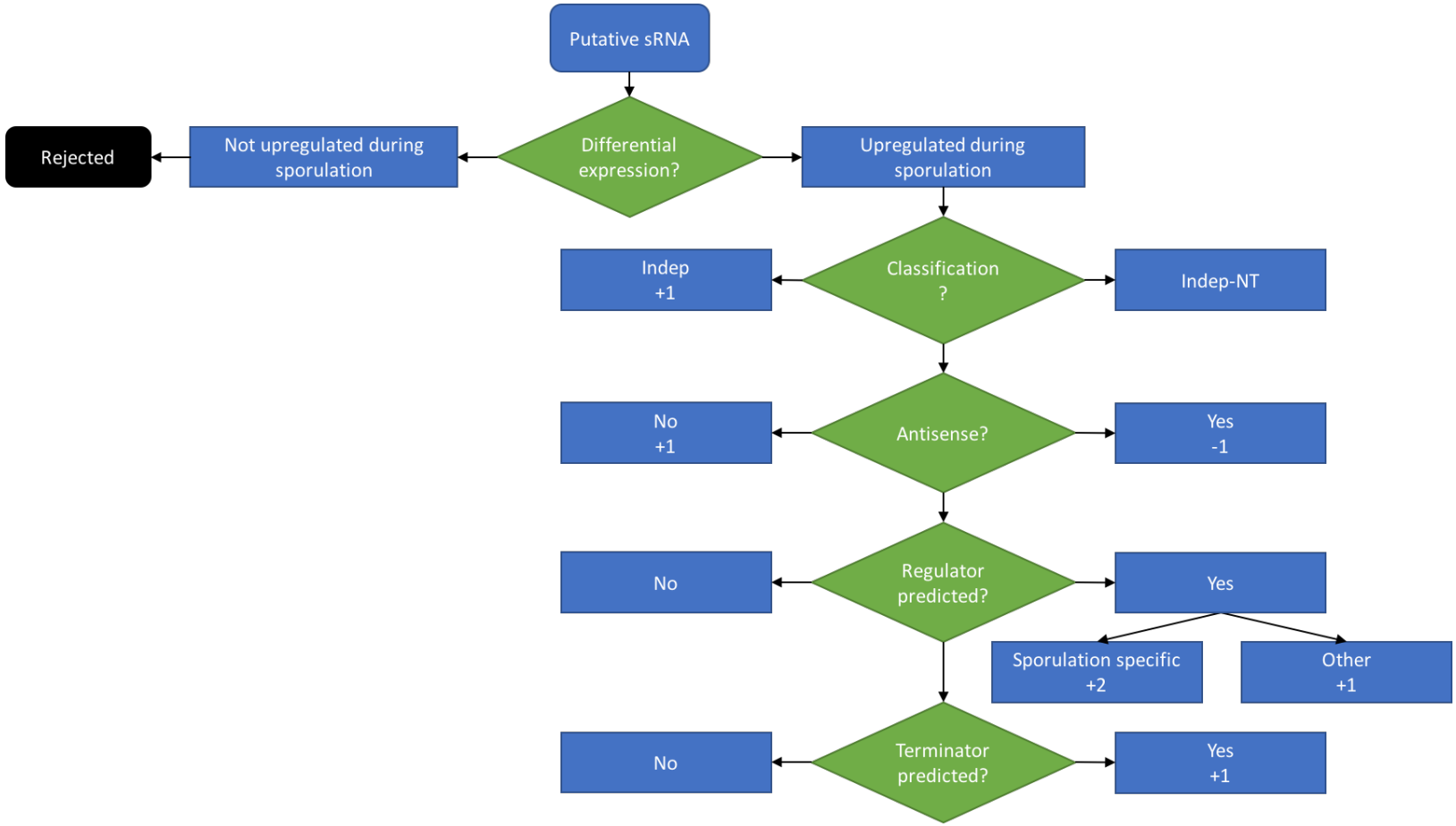


Figure 6 –Workflow to select putative sRNAs involved in sporulation. Putatitve sRNAs were screened for differential expression. sRNAs that were upregulated were selected and fed through the analysis pipeline. This yielded a list of sRNAs which were then ranked for desired characteristics.

3.2 Strategies for *in-silico* identification and characterisation of sRNAs

3.2.1 Identification of putative sRNAs upregulated during endospore formation

Using transcriptional profile cluster analysis, Nicolas *et al*. (2012) determined similarities between the expression profiles of the transcriptionally active regions (TARs) over 104 different conditions. During the analysis, cluster analysis was carried out between the different TARs to identify expression correlations. Clustering of the active regions was carried out to form three subgroups, that have increasing degrees of correlation. Cluster A have the highest expression correlation, followed by B and then C. These clusters can be validated by the distribution of classified promoter recognition sequences throughout the group, as genes with the same expression patterns are likely to be controlled by the same sigma factor (Nicolas, P. *et al.*, 2012).

Sporulation has several distinct stages defined by morphological characteristics. These characteristics are carefully regulated by a series of sporulation-specific sigma factors. SigF and SigG direct gene expression specifically in the forespore and SigE and SigK do the same in the mother cell (Stragier, P. & Losick, R., 1990). Nicolas, P. *et al.* (2012) took samples at hourly intervals for eight hours in cells going through sporulation (Sterlini, J. M. & Mandelstam, J., 1969). The data available from Nicolas *et al.* will be extracted and the defined C clusters will be used to select putative sRNAs, where clusters with increased expression during sporulation and independent segments will be extracted and termed putative sporulation-specific sRNAs. Those that are nominated to be likely sporulation-related will then be used to predict promoters and terminators.

3.2.1.1 Transcriptional control predictions

Sporulation has four dedicated sigma factors and therefore the presence of a binding site of one of these in the promoter region of a gene or sRNA is a good indicator that

transcription of these elements may be activated during the sporulation process. As such, sporulation-specific sRNAs would presumably have a binding site for sigma factors SigE, SigF, SigG or SigK. Nicolas *et al.* performed promoter analysis to validate their clustering, using an unsupervised promoter classification tool TREEMM (Nicolas, P. *et al.*, 2012).

In addition to the prediction made previously, an independent study will be performed here utilising the Data Base of Transcriptional regulation in *B. subtilis* (DBTBS). DBTBS is a web-based database containing information about transcriptional binding sites and gene regulation in *B. subtilis* (Sierro, N. *et al.*, 2008), which allows the prediction of sigma factor and transcriptional regulator binding sites.

Both the predictions from Nicolas *et al.* and DBTBS will be used to determine if the predicted regulatory elements can be used to explain the upregulation of the sRNA during sporulation. A sporulation-specific regulator will be used as a positive indicator for sRNA expression and, therefore, activity during sporulation.

3.2.1.2 Termination considerations

Rho-dependent transcription termination relies on the RNA-specific helicase Rho, which is a termination factor. Rho in *E. coli* is an essential and abundant protein (Das, A. *et al.*, 1976; Li, G. W. *et al.*, 2014). Conversely, *B. subtilis* can survive in nutrient-rich media in the absence of *rho* (Quirk, P. G. *et al.*, 1993). Interestingly, a proteomics analysis revealed the levels of Rho to be of low cellular abundance in comparison to levels of RNA polymerase in *B. subtilis* (Muntel, J. *et al.*, 2014). Despite this, Rho is known to significantly influence the transcriptome (Nicolas, P. *et al.*, 2012) suppressing persuasive transcription in addition to influencing cellular differentiation programs, such as sporulation (Bidnenko, V. *et al.*, 2017). Rho-independent (or intrinsic) termination does not depend on termination factors, but relies on a hairpin loop and a poly-U tail.

Transcript termination is another good indication of an independent RNA species and it will be used in this study to predict a likely functional sRNA. Alongside the prediction of Rho-dependent termination made by Nicolas *et al.*, predictions will be made for Rho-independent termination sites utilising ARNold. ARNold is a web-based tool used for the identification of Rho-independent terminators and utilises two corresponding programs, ERPIN and RNAmotif, to predict the existence of a Rho-independent terminator in a given sequence (Naville, M. *et al.*, 2011).

3.2.2 Further characterisation of putative sRNAs

Further work in this study will surround the conservation of a putative sRNA sequence. In addition to this, previously published data on sporulation-related conditions will be utilised to further the information available of potential sRNAs, such as RNAseq data available during germination (Nagler, K. *et al.*, 2016).

3.2.2.1 Conservation

Through phylogeny, shared traits can be traced back to an evolutionary ancestor. Evolution is a consequence of random advantageous changes being selected for in inheritable traits, resulting in biological diversity despite a common ancestor. Therefore, distinct species can have shared characteristics despite being otherwise diverse. As such, species can diversify and specialise to the environment they are found in whilst maintaining shared advantages, a prime example of which is sporulation. Endospore formation provides the ultimate "bunker" to survive unfavourable environments. Most endospore formers are soil and gut dwellers, which are both subject to extreme changes. The vast array of soil ecology creates selective pressures for bacteria, as each location has its own challenges (temperature, pH, etc.). As such, many different species share the same ability to produce an endospore, with the end-product of sporulation not being identical (Henriques, A. O. & Moran, C. P., Jr., 2007).

Onyenwoke, R. U. *et al.* (2004) were one of the first to distinguish between endospore formers, non-endospore formers and asporogenic bacteria. Non-

endospore forming bacteria are those that are lacking most sporulation-specific genes, while those termed asporogenic contain many of the sporulation-specific genes but remain unable to produce an endospore. Endospore formation is shared between *Bacilli* and *Clostridia*, which include many species lacking this ability such as the non-endospore formers *Listeria, Staphylococcus, Streptococcus* and *Lactobacillus*.

If part of the essential sequential processes involved in endospore production is disrupted, then it is possible for a bacterium to still have most of the genes required but remain asporogenic (Onyenwoke, R. U. *et al.*, 2004). Thus, it is important to note that when analysing the evolutionary relationships of sporulation genes. Selective evolutionary pressure is also considered to be the reason behind gene regulation being more tightly conserved than the presence of genes required for endospore formation (de Hoon, M. J. *et al.*, 2010). Narrow conservation may therefore not indicate a lack of essentiality, but rather an adaptation to a particular niche. For example, many of the genes encoding endospore coat proteins are not widely conserved due to their exposure to the different environments bacteria reside in (Galperin, M. Y. *et al.*, 2012).

3.2.2.1.1 Conservation pattern of sporulation-specific sRNAs

sRNAs have previously been shown to be conserved between or within species (Durand, S. *et al.*, 2015; Sharma, C. M. *et al.*, 2011; Marchais, A. *et al.*, 2011). The process of sporulation is more conserved than other biological processes, such as the general stress response (de Hoon, M. J. *et al.*, 2010). This has enabled phylogenetic profiling to be successfully used to identify new sporulation genes (Traag, B. A. *et al.*, 2013). Therefore, this study asked the question if sporulation-specific sRNAs were also more conserved. Sporulation has known tiered conservation; the most conserved functions are the primary form of control of sporulation via alternative sigma factors. If a putative sRNA is also conserved in the relevant species, could the sRNA play a vital role? Equally, if an sRNA is only specific to *B. subtilis*, could it indicate the involvement of this sRNA in the species-specific elements of endospore formation in *B. subtilis*?

Specific areas of conservation within the sequence of an sRNA, despite the diversity of species, could show areas of interest to then help guide future target predictions. This is because *trans*-encoded sRNAs require 6-8 bases of complementarity to accomplish their effects, although much longer interacting regions have also been reported and as little as one nucleotide can be important for regulation (Gottesman, S. & Storz, G., 2011; Papenfort, K. *et al.*, 2012). As such, there could be an increased level of conservation for the interacting regions. MicroRNAs in eukaryotes are analogous to bacterial sRNAs, in that they base pair with their target through short seed pairing regions, and there has been a success in using computational methods for discovering their targets (Krek, A. et al., 2005; Lewis, B. P. et al., 2003). However, this is aided by the fact that miRNAs are much shorter in length (about 22 nt) in comparison to their bacterial counterparts. Since the section of interaction is so small for bacterial sRNAs, it has been proposed that these regions are made more accessible via the use of stem-loop structures (Papenfort, K. et al., 2008). Thus, combining structure and regions of conservation could be an important part of searching for bacterial sRNAs that would then be attractive for further study.

The Basic Local Alignment Search Tool (BLAST) is a program dedicated to searching for regions of similarity between sequences by comparing a given sequence to a large database (Altschul, S. F. *et al.*, 1990). Whilst BLAST is useful for sequence similarity, it does not take into consideration the structural relationship, which is an essential feature of RNA biology. Combining structural information in searching for similarity allows relaxation of the parameters for sequence based matching thereby reducing the number of false results. GLASSgo is a package from Freiburg RNA Tools which uses iterative BLAST searches and structure-based clustering to search for sRNA homologs (Freiburg Bioinformatics, Group, 2017). This tool will be utilised to search for homologs of putative sRNAs in this study.

3.2.2.1.2 Relationship between sequence and structural conservation

Traditionally, identification of protein-coding genes is followed by a prediction of their function. This is done by comparing the predicted gene products with

experimentally validated proteins. This is followed by automation utilising large databases of protein domains and families clustered based on taxonomic organisation, such as the PFAM database (Finn, R. D. *et al.*, 2008). Another approach involves the introduction of phylogenetic profiling and gene expression analysis alongside sequence-matching (Galperin, M. Y. & Koonin, E. V., 2000). Predictions of sRNA functionality are hampered by the lack of functionally characterised sRNAs, with no comprehensive databases to compare against.

As mentioned previously, another characteristic shared among most sRNAs is the importance of secondary structure (Rivas, E. & Eddy, S. R., 2001). Initial studies explored the possibility to look for stable secondary structures, but RNA secondary structure alone is not enough to confer a putative sRNA over spurious folding (Rivas, E. & Eddy, S. R., 2000). From there, conservation was incorporated into predictions assuming that the sRNA, and hence structures of importance, would have been safeguarded (Rivas, E. & Eddy, S. R., 2001; Gorodkin, J. *et al.*, 2010). However, looking for conserved sequences alone can identify many false positive results. It is important to note that the nucleotide sequence can evolve and still maintain the crucial secondary structure (Gorodkin, J. *et al.*, 2010).

Determination of the Minimum Free Energy (MFE) structures is a popular tool for analysis of RNA folding. The MFE structure is comprised of a loop-based energy model, where the secondary structure contributes a minimum of free energy (Zuker, M. & Stiegler, P., 1981). However, it was found that folding of intergenic regions was not so useful for *de novo* prediction of functionally relevant sRNAs (Rivas, E. & Eddy, S. R., 2000). This, however, can be circumvented when additional information is considered. For example, consensus folding and MFE analysis of aligned sequences has been shown previously to be successful in predicting functional RNAs (Gorodkin, J. *et al.*, 2010). Utilising RNAalifold (Lorenz, R. *et al.*, 2011), it is possible to resolve uncertainties in alignment where there are conserved secondary structures, despite sequence variation (Hofacker, I. L. *et al.*, 2002).

3.3 Results

3.3.1 Identification of putative sRNAs upregulated during endospore formation

The expression values during sporulation were extracted for all genes and grouped per the clusters defined by Nicolas *et al.* Using a combination of visual inspection and promoter clustering, it was apparent that many of the sporulation-specific genes are present in Cluster C2. Using classification of genes as defined by SubtiWiki (Michna, R. H. *et al.*, 2016), over half (55%) of the functionally classified genes in Cluster C2 are recognised as part of the sporulation sub-category. Cluster C17 contained genes associated with the initiation of sporulation, amongst other genes involved in the transition phase to sporulation such as those controlled by the transcription factor AbrB (Chumsakul, O. *et al.*, 2011).

Clusters C2 and C17 also included newly identified RNA segments. It was hypothesised that the new RNA features present in these two clusters were likely to also be sporulation-specific. Within these two clusters are 75 potential independent RNAs with patterns of increased expression during sporulation. Considering there are a total of 154 Indep segments identified, this is a surprisingly high number. Of these 75, 41 are classified as independently expressed with their own upshift and downshift in expression.

All independent RNA features from the study by Nicolas *et al.* were extracted and the expression values during sporulation were averaged where there was more than one feature (Figure 7). Clusters were identified that exhibited a large change in expression as sporulation progressed over time, clusters C2, C17, C25 and C74 were identified to be the most relevant. C2 again is a strong candidate, with average expression strongly increasing at the fourth hour during sporulation. C17, whilst having high expression at all time points, also increases as sporulation progresses. The two new clusters, C25 and C74, appear to dramatically increase in expression from hours 5 and 6 respectively. This could indicate a role in the late stages of

sporulation. The inclusion of these two new clusters resulted in a total of 85 independent putative RNAs, with 45 having their own promoter and terminator.

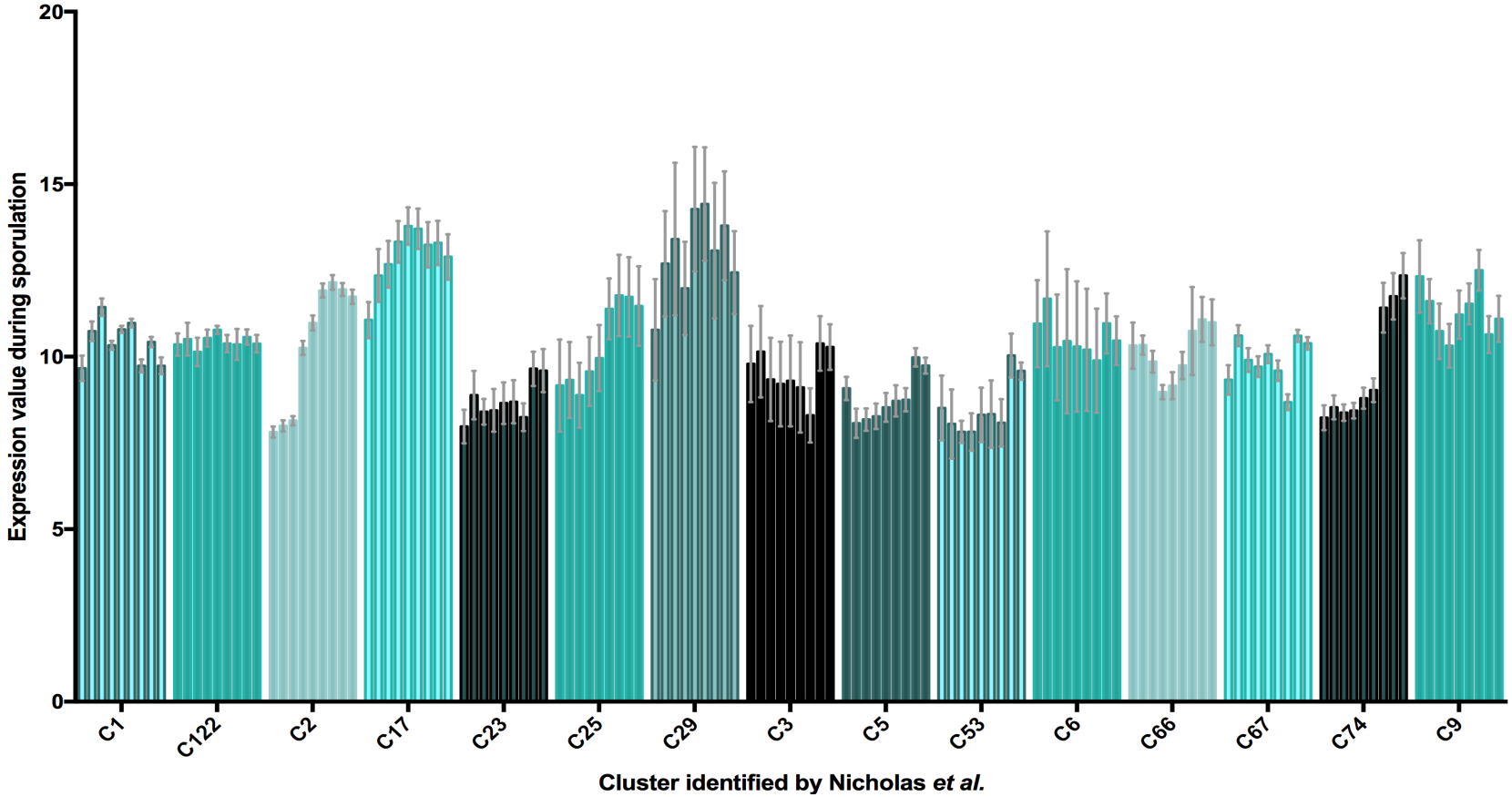


Figure 7 – Independent RNA features expression during sporulation. Average expression values of all independent RNA features at each time point during sporulation. Each column represents one time point, from hour 0 to 8. Only clusters with more than one putative RNA are represented. Error bars represent standard error mean.

3.3.1.1 Selected putative sRNAs

Since there are no definitive attributes to an sRNA, and there are exceptions to many of the rules, it is hard to dismiss many candidates outright. However, to determine which putative sRNAs to study further, this study utilised a set of criteria and hence making it possible to pick those which would be most suitable for future studies. Point were assigned for desired characteristics as per the workflow set out in Figure 6 and the results of which are found within Table 5. The cut-off score for the number of desired characteristics was set at 3 and those above this threshold were chosen to be subject to further study in chapter four. This resulted in the initial number of sRNA candidates being reduced from 85 to a total of 24.

Table 5 – Summary table of independent RNA features ordered by score. Sigma Nicolas = Sigma factor predicted by Nicolas *et al.* Sigma DBTBS and TF DBTBS = Sigma factors and transcription factors predicted by DBTBS respectively. Antisense overlap is the number of nucleotides the putative sRNA and any antisense gene overlap by.

Name	Classification	Sigma Nicolas	Sigma DBTBS	TF DBTBS	Antisense gene	Antisense overlap	Arnold terminator prediction	Rho terminator prediction	Score	Length (bp)
S111	Indep-NT	SigEF	NA	MntR, DegU	NA	0	Yes	0	5	148
S1024	indep	SigK, SigA	SigX, SigW	NA	NA	0	Yes	0	4	216
S1227	Indep-NT	SigK, SigA	NA	ComK	NA	0	0	0	4	106
S2	Indep-NT	SigK, SigEF	NA	NA	NA	0	0	0	4	273
S357	indep	SigGF	NA	NA	NA	0	Yes	0	4	296
S423	indep	SigK, SigEF	NA	NA	NA	0	Yes	0	4	186
S547	indep	SigGF	SigG, SigF	ComA	NA	0.76	Yes	0	4	123
S645	indep	SigK	NA	SpoIIID, DegU, AhrC	NA	0	Yes	0	4	160
S849	indep	0	SigG	NA	NA	0	Yes	0	4	74
S968	indep	SigK, SigA, SigB	SigW	NA	NA	0	Yes	0	4	147
S1009	indep	SigK, SigA, SigA	NA	NA	NA	0	0	0	3	151
S1027	indep	SigA, SigA	NA	NA	NA	0	Yes	0	3	219
S1202	Indep-NT	SigB	SigF, SigE, SigB	NA	*yugH*	0.91	Yes	Yes	3	984

S145	Indep-NT	SigEF, SigB	NA	ComA	*ydzK*	0.36	Yes	0	3	215
S198	indep	SigA	NA	NA	NA	0	Yes	0	3	197
S249	indep	SigK	NA	NA	NA	0	0	0	3	560
S275	indep	SigGF, SigA, SigB	NA	AbrB	NA	0	0	0	3	156
S612	indep	SigA	NA	DegU	NA	0	Yes	0	3	133
S665	indep	SigK	NA	NA	NA	0	0	0	3	109
S731	indep	SigEF	NA	LexA	NA	0	0	0	3	498
S863	indep	SigK, SigA, Sig-	NA	NA	NA	0	0	0	3	199
S877	indep	SigEF	NA	GltR	NA	0	0	0	3	107
S912	indep	SigWXY, SigK	NA	NA	NA	0	0	0	3	170
S978	indep	SigA, SigA	NA	NA	S789	0.00	Yes	0	2	219
S1038	Indep-NT	SigK	NA	NA	*alaS*	0.86	0	0	2	1189
S1042	Indep-NT	SigGF	NA	SpoIIID, BkdR	*yrrC*	1.00	0	0	2	349
S1083	Indep-NT	SigK	NA	ComA	*ysfB*	0.78	0	0	2	1412
S1105	Indep-NT	SigEF	NA	NA	*ytsJ*	0.70	0	0	2	1174
S1157	Indep-NT	SigEF	SigW	PerR	*ytvB*	0.75	0	0	2	440
S1279	indep	Sig-	NA	DegU, ComK	NA	0.57	0	0	2	967
S13	Indep-NT	SigEF	NA	NA	*pdxS*	1.00	0	0	2	729
S1359	Indep-NT	SigK	NA	TnrA	*tagO*	0.80	0	0	2	1233
S1376	Indep-NT	SigGF	NA	NA	*tagA*	1.00	0	0	2	348
S1445	Indep-NT	SigEF	NA	NA	*clsB*	1	0	0	2	878

S1455	indep	SigA	NA	NA	NA	0	0	0	2	262
S1520	Indep-NT	SigK	NA	NA	*yxeI*	1	0	0	2	633
S1534	indep	SigA	NA	CodY	NA	0	0	0	2	424
S254	indep	SigA	NA	NA	NA	0.00	0	0	2	175
S265	Indep-NT	SigEF	NA	CodY	*mapB*	0.37	0	0	2	1064
S283	Indep-NT	SigEF	NA	NA	*yfjC*	0.76	0	0	2	587
S286	Indep-NT	SigK	NA	ComA	*yfiQ*	1.00	0	0	2	388
S309	indep	SigA	NA	NA	NA	0.00	0	0	2	329
S31	Indep-NT	SigEF	NA	NA	*ispE*	0.37	0	0	2	2354
S313	indep	SigA	NA	NA	NA	0.00	0	0	2	175
S359	Indep-NT	SigK	NA	NA	*yhfC*	0.46	0	0	2	621
S372	indep	SigEF	NA	FNR	*glcP*	0.92	Yes	0	2	526
S41	Indep-NT	SigEF	NA	DegU	*dusB*	0.55	0	0	2	1516
S416	Indep-NT	SigEF	NA	NA	*yjbH*	0.33	0	0	2	1901
S58	Indep-NT	SigK	NA	NA	*truA*	1.00	0	0	2	468
S601	Indep-NT	SigEF	NA	NA	*dapA*	0.45	0	0	2	1184
S632	Indep-NT	SigK	NA	NA	*xynB*	1.00	0	0	2	305
S651	Indep-NT	SigK	NA	NA	*yndH*	0.56	0	0	2	1098
S653	indep	Sig-, SigA	NA	GltR, DegU	NA	0.00	0	0	2	131
S675	Indep-NT	SigK	NA	PucR	*eglS*	0.71	0	0	2	746
S733	Indep-NT	SigK	NA	Xre	*yocH*	0.98	0	0	2	767
S827	Indep-NT	SigK	NA	NA	*ugtP*	0.95	0	0	2	1029
S869	indep	SigEF	NA	NA	*ypdA*	0.60	Yes	0	2	193
S882	Indep-NT	SigEF	NA	AhrC	*sipS*	0.60	0	0	2	922

S951	Indep-NT	SigA, SigA	NA	CcpA	*sigA*	0.52	0	Yes	2	2149
S956	Indep-NT	SigEF	NA	NA	*recO*	0.54	0	0	2	1412
S962	indep	SigK, SigA	NA	DegU	*yqzM*	0.40	Yes	0	2	339
S975	Indep-NT	SigB, SigEF	SigG	NA	*yqxH*	0.50	0	0	2	1322
S1082	Indep-NT	SigA	NA	NA	*ysgA*	0.83	0	0	1	264
S1214	Indep-NT	SigA	NA	NA	*comX*	0.42	0	0	1	972
S1234	indep	SigK	NA	CodY	*yuiC*	0.70	0	0	1	446
S1236	indep	Sig-, SigK	SigF	NA	*yuiB*	0.68	0	0	1	109
S1388	indep	SigK	SigK	NA	*cotG*	0.59	0	0	1	197
S181	indep	0	NA	NA	NA	0.00	0	0	1	545
S326	indep	SigGF	NA	DegU, CodY	*ygxB*	0.13	0	0	1	110
S37	indep	SigEF	NA	DegU	*spoIIE*	1.00	0	0	1	609
S4	indep	SigEF	NA	NA	*dnaA*	0.86	0	0	1	768
S400	Indep-NT	Sig-	NA	BkdR	*yjaV*	0.46	0	0	1	602
S526	indep	SigEF	NA	LexA, FNR	*kinC*	0.79	0	0	1	1629
S732	indep	SigK	NA	DegU, CodY	*yocH*	0.58	0	0	1	194
S81	indep	SigK	NA	NA	*glpT*	0.53	0	0	1	626
S821	indep	SigEF	NA	NA	*ypjP*	0.65	0	0	1	946
S871	Indep-NT	0	NA	CodY	*recQ*	0.95	0	0	1	746
S885	indep	SigK	NA	NA	*ypuC*	0.27	0	0	1	1071
S1136	indep	SigB	NA	NA	*rpsD*	0.10	0	0	0	219
S1180	Indep-NT	0	NA	NA	*glgB*	0.83	0	0	0	1419
S1197	Indep-NT	0	NA	NA	*yugK*	0.91	0	0	0	921
S1335	indep	Sig-, SigA	NA	NA	*cypX*	1.00	0	0	0	285

S289	Indep-NT	0	NA	NA	*yfhH*	0.35	0	0	0	268
S562	indep	SigA	NA	TnrA, ComK, BkdR	*spoIIGA*	0.81	0	0	0	1142
S1574	indep	0	NA	NA	*exoAA*	0.93	0	0	-1	503

3.3.2 Identification of sRNAs with regulatory sites

3.3.2.1 Promoter predictions from Nicolas, P. *et al.* (2012)

The sigma factor binding motifs predicted by Nicolas *et al.* were extracted for all putative sRNAs (Table 6). Single sigma factor predictions are difficult for SigE, SigF and SigG as their recognition motifs are very similar (Eichenberger, P. *et al.*, 2004; Wang, S. T. *et al.*, 2006). A total of 61 sporulation-specific sigma factor binding sites were identified (Table 6), with the late mother cell specific sigma factor SigK attributed to almost half (30) of these. Clusters C25 and C74 were predicted to be associated with late sporulation, based on the average expression profile of the independent RNA features (Figure 2), and this was reflected in the predicted sigma factor binding sites (Table 6B):

Table 6 – A) Summary of extracted sigma factor binding sites from Nicolas *et al.* per RNA segment classification B) Summary of extracted sigma factor binding sites by Nicolas *et al.* per cluster

A	Sporulation				Other				
	SigK	SigEF	SigGF	Total	SigA	SigB	SigWXY	Undefined	Total
Indep-NT	14	16	2	**32**	5	3	0	5	**13**
Indep	16	9	4	**29**	20	1	1	7	**29**
Total	**30**	**25**	**6**	**61**	**25**	**4**	**1**	**12**	**42**
B									
C2	23	23	5	**51**	9	3	1	8	**21**
C17	0	1	1	**2**	1	1	0	2	**4**
C25	3	0	0	**3**	1	0	0	1	**2**
C74	4	1	0	**5**	3	0	0	1	**4**

3.3.2.2 Promoter predictions made utilising DBTBS

The primary benefit of utilising DBTBS software tool, despite not being automated, is its ability to identify transcriptional regulator binding sites. Using 200 nucleotides upstream and 100 nucleotides downstream of the predicted start site of an RNA, the 5' regions of putative sRNAs were examined for predicted binding sites of transcriptional regulators with a threshold of 1% p-value.

A total of 59 binding sites were predicted for different sigma factors and transcriptional regulators (Figure 8). All predicted sigma factors are involved in stress responses. The housekeeping sigma factor, SigA, was not predicted in this study contrary to the findings from Nicolas, P. *et al.* (2012)(Table 2). Sigma factors SigB (general stress response), SigW (membrane-active compound adaptation) and SigX (resistance to cationic antimicrobials) are predicted binding sites in addition to the sporulation-specific sigma factors, SigE and SigF, from early sporulation, and SigK and SigG from the later stages in sporulation. The transcriptional regulators predicted in addition to sigma factors mainly consist of those important in controlling the expression of metabolism genes. Of interest is the transcriptional regulator SpoIIID, an early regulator of mother cell genes (Eichenberger, P. *et al.*, 2004). Two putative sRNAs were revealed to have a potential SpoIIID binding site, S1042 and S645. Interestingly, the most predicted binding motif is for DegU, a two-component response regulator of multicellular behaviour (Mader, U. *et al.*, 2002a; Lopez, D. & Kolter, R., 2010).

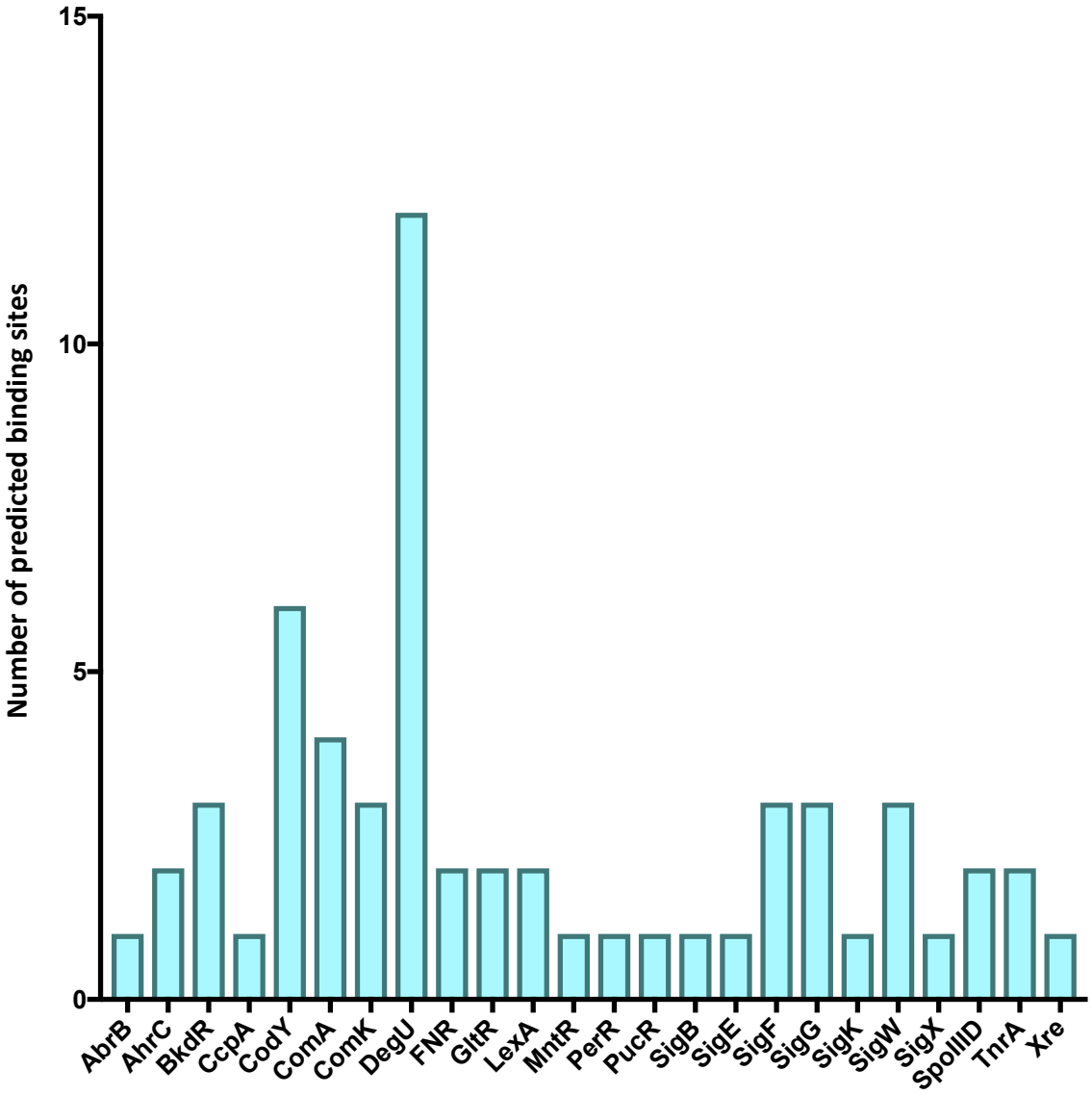


Figure 8 – Number of sigma factor and transcription factor binding sites. Binding sites were predicted utilising DBTBS, a web-based database containing information about transcriptional binding sites. The overall number of these binding sites was quantified. DegU was found to have the highest number of predicted binding sites in this analysis.

3.3.2.2.1 Validation of transcriptional regulator predicitions using RNAseq

CodY is a Guanosine-5'-triphosphate (GTP)-binding transcriptional regulator with a large regulon (Ratnayake-Lecamwasam, M. *et al.*, 2001). It was predicted to have six binding sites for the putative sRNAs in this study (Figure 8). CodY repression of gene expression is dependent on high cellular GTP and branched-chain amino acid (BCAA) levels, which increase the affinity of CodY to its DNA targets, resulting in repression of transcription (Handke, L. D. *et al.*, 2008). The CodY regulon is comprised of many genes important in the transition from exponential growth to stationary phase and sporulation. For example, a member of the CodY regulon is Spo0A, which is a major transcription factor itself, essential for the initiation of sporulation (Chastanet, A. *et al.*, 2010). CodY has also been implicated in the initiation of sporulation, where high glucose and therefore high GTP delays sporulation and deletion of the *codY* gene circumvents this (Ratnayake-Lecamwasam, M. *et al.*, 2001; Fujita, M. & Losick, R., 2005). In addition, CodY has been shown previously to activate an sRNA BsrF/S728 (Preis, H. *et al.*, 2009).

Brinsmade, S. R. *et al.* (2014) employed RNA sequencing to further understand the gene regulation of CodY, showing that over 200 genes are controlled by CodY in a hierarchical manner. RNA was extracted from the wild-type and a *codY* null strain during the exponential growth stage in chemically defined media. At this time point, CodY is known to be repressing transcriptions of its targets. Whilst their study was comprehensive for protein-coding genes, there was no mention of putative sRNAs during the analysis.

The data for the wild-type and Δ*codY* mutant strains was recovered from the Gene Expression Omnibus (GEO) and reanalysed to consider all transcripts, therefore including potential sRNAs. Differential gene expression analysis of this data showed that two of the predicted CodY binding sites for sporulation-related sRNAs were indeed under the control of CodY (Table 7), either directly or indirectly. Both S1534 and S1234 are upregulated in the absence of *codY* (5.35 and 4.04 log2 fold change respectively), confirming the prediction from DBTBS.

Differential gene expression analysis revealed there are 56 S segments with a statistically significant log2 fold change (Table 7). Within this, there are an additional two Indep segments (S1236 and S254) and four Indep-NT segments (S1070, S1290, S1498 and S117) affected by deletion of *codY*.

Table 7 – Log2 fold change between wild-type and Δ*codY* for RNA segments with predicted CodY binding motifs (Figure 3)

	log2 fold change	Adj p-value
S1534	5.35	0.00
S1234	4.04	0.00
S326	0.91	0.89
S732	0.35	1.00
S871	0.09	1.00
S265	0.01	1.00

Table 8 – Summary of the types of RNA segments with statistically significant changes (log2 fold change > 1.5 and an adjusted p-value < 0.05) upon deletion of *codY*

Feature type	count
Indep-NT	4
intra	1
indep	4
inter	16
5'	21
3'PT	7
3'UTR	3
Total	**56**

3.3.2.3 Terminator predictions

Nicholas *et al.* observed the transcripts of two of the putative sRNAs, S1202 and S951 to be elongated in a Δ*rho* background. Therefore, lacking transcription termination hence potentially possessing rho-dependent termination. However, the elongation was less than 50 nts for each. The expression of S1202 during sporulation is strong after 3 hours, though it should be noted that strong transcription is shown for a small portion of the gene before it tails off (Figure 9). S951 is an extremely long RNA transcript of 2148 nt, opposite *sigA* and *dnaG* amongst other newly identified RNA segments. The transcription of S951 also appears to be strong at the start, but expression levels quickly drop.

Rho-independent terminator predictions were made by this study for 17 of the 85 putative sRNA sequences. Of the 17 predicted to possess a rho-independent terminator, three sRNAs were classified as indep-NT (S111, S1202 and S145) in the Nicolas *et al.* study. S111 is strongly upregulated from the second hour of the sporulation process. Analysis of the tiling array profile suggests some termination, with a clear partial downshift in expression (Figure 10). S1202 was identified as potentially having Rho-dependant termination and does appear to have a mild drop in expression at the end (Figure 9). Three segments of S1202 have distinct levels of expression. There is a clear peak of expression in the beginning, then a stretch of expression at a lower value and then followed by a final lower strength expression signal. Figure 11 displays the expression profile of S145, which is entirely unclear.

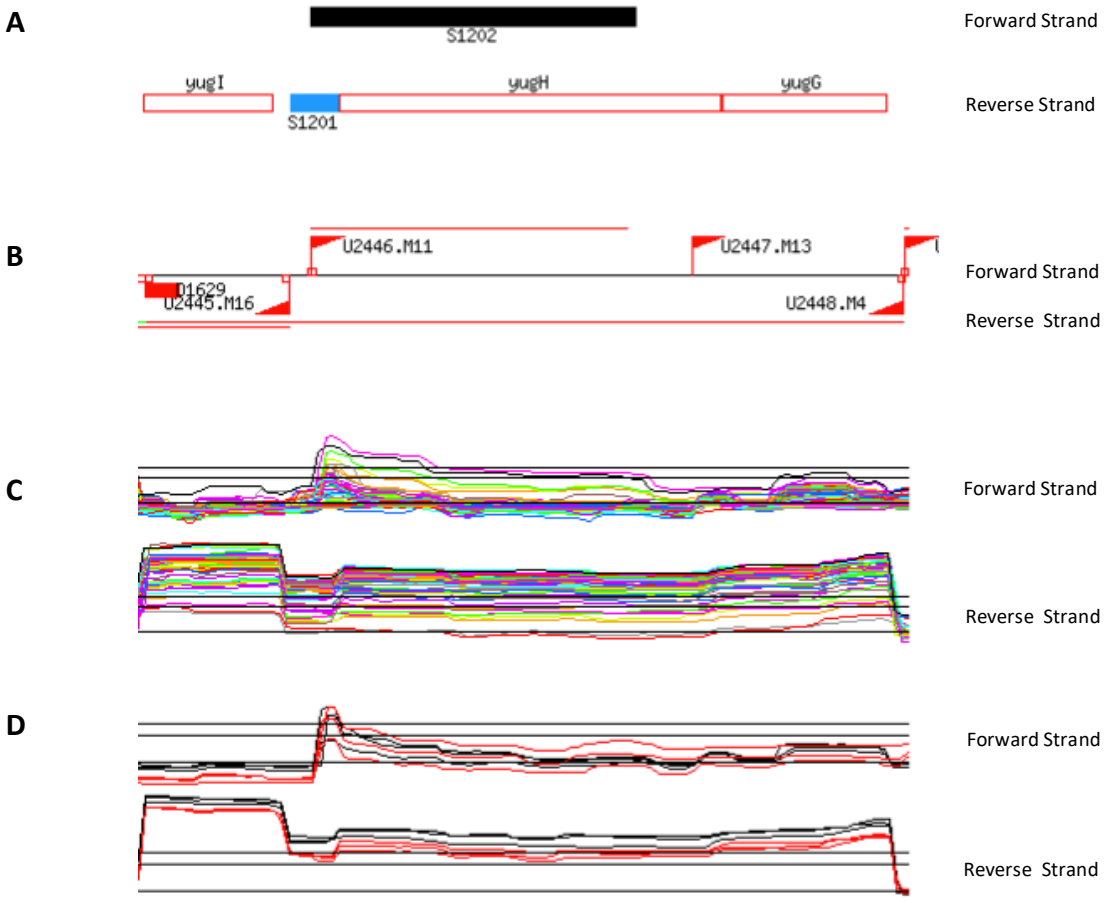


Figure 9 – Expression profile of S1202 from Nicholas *et al.* shows a minor termination deficiency in the absence of *rho*. A) Schematic representation of genomic location for S1202. B) Simplification of transcriptional subunits. Flags indicate an upshift in expression. Lines are used to indicate length of transcriptional unit. C) Transcription profiles across all 104 conditions, the S1202 transcription tapers off. D) *rho*-mutant (red) vs wild-type (black) transcription profile.

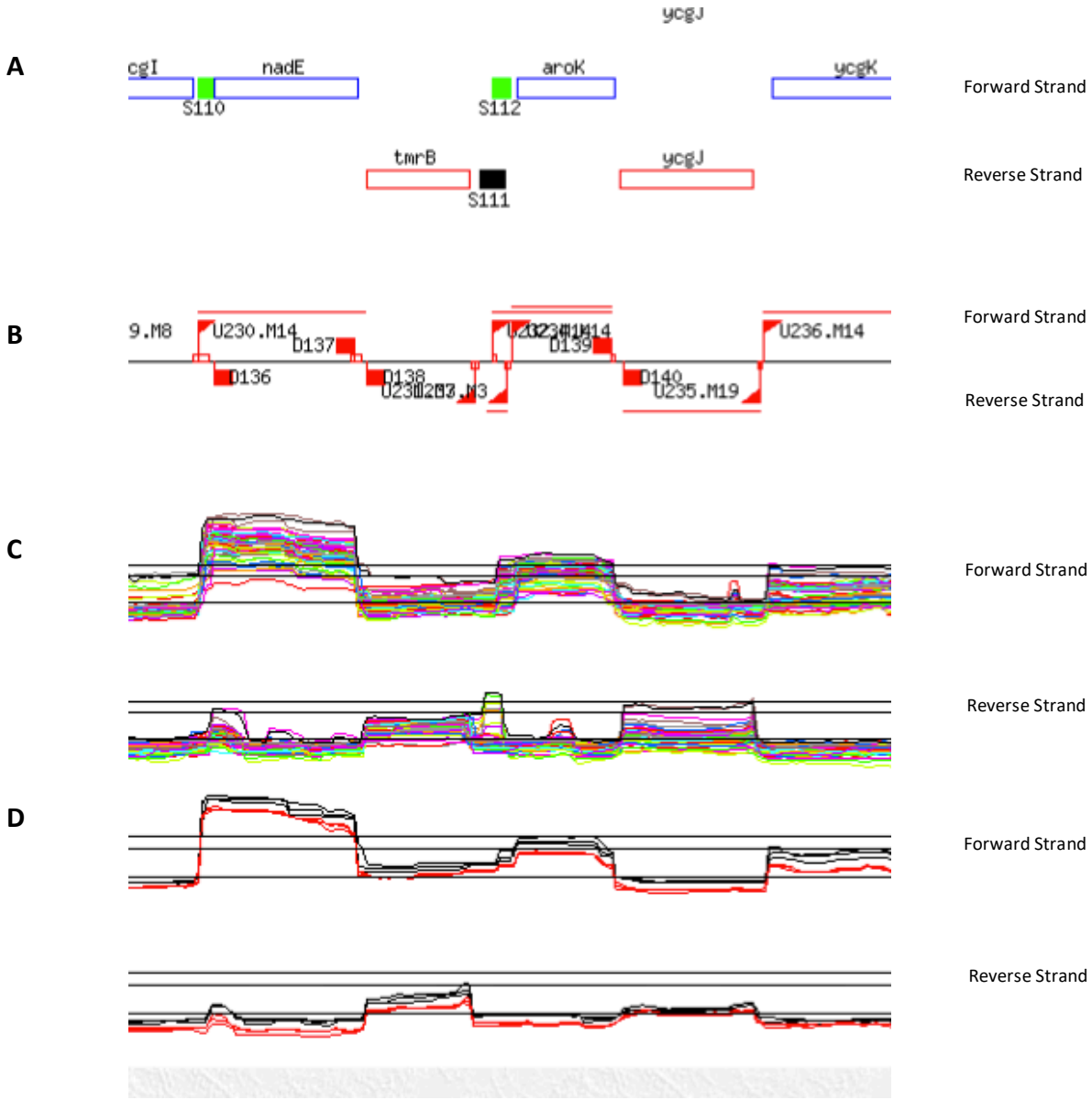


Figure 10 – Expression profile of S111 from Nicholas *et al.* shows conditional termination of S111. A) Schematic representation of genomic location for S111. B) Transcriptional simplification. Flags indicate an upshift in expression. Lines used to indicate length of activity and hence a transcriptional unit C) Transcription profiles across all 104 conditions – in a small number of conditions there is an upshift in S111 expression. D) *rho*-mutant (red) vs wild-type (black) transcription profile

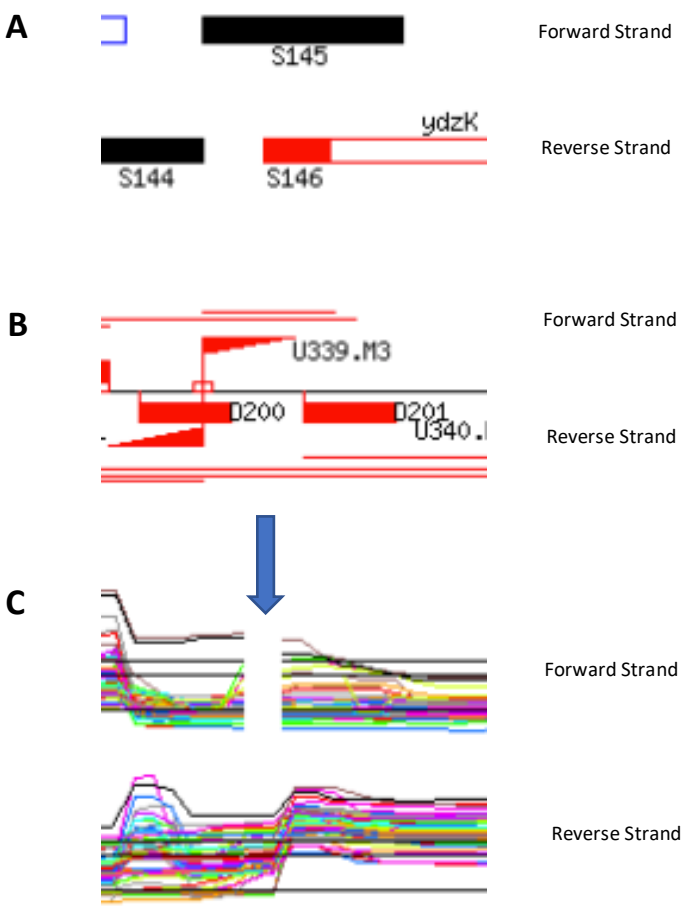


Figure 11 – Gap in the expression profile of S145 prevents further conclusions. (A) Schematic representation of S145 genomic location. (B) Transcriptional simplication, flags indicate an upshift in expression. Lines used to indicate length of activity and hence a transcriptional unit. (C) Transcriptonal profiles across all 104 conditions. A gap in the data is present at the position of S145, this prevents any conclusions from being drawn.

3.3.2.4 Phylogenetic analysis

Endospore formation is shared between three classes of phylogenetically Gram-positive organisms, including *Bacilli, Clostridia* and *Negativicutes* (stains Gram-negative). The variability of the essential sporulation genes between the three highlights the diversity of the sporulation process (de Hoon, M. J. *et al.*, 2010). However, conservation of sets of genes involved in the various sporulation stages as well as morphological variability during sporulation allows the construction of a phylogenetic profile. This could then give an indication of where an sRNA may be functioning as themes have emerged as to the function and phylogenetic distribution of some sporulation genes (Galperin, M. Y. *et al.*, 2012; de Hoon, M. J. *et al.*, 2010).

The program chosen to identify homologs of putative sRNAs from Table one was GLASSgo (Lott, S. *et al.*, 2017). However, GLASSgo is limited to the sequence length below 800 nucleotides and therefore not all sRNA candidates could enter this part of the pipeline. A total of 25 sRNAs were too long and were therefore excluded from this analysis. However, 60 sRNAs were entered into the GLASSgo web server (http://rna.informatik.uni-freiburg.de/GLASSgo/Input.jsp) and were reported to have homologs with varying species diversity.

The analysis identified 555 different species that had at least one homolog to the putative sRNAs identified. Many of the species identified belong to the family *Bacillaceae* (Figure 12), which are Gram-positive rod-shaped bacteria, most of which are capable of producing endospores and to which *B. subtilis* is a member. To explore the distribution of sRNAs identified to have homologs across different species (orthologs), a simple binary approach was produced from the output of GLASSgo, determining whether a putative homolog was present (1) or absent (0).

Using the presence or absence as an indicator of diversity, the phylogenetic profiling showed four clear clusters of putative sRNAs (Figure 12). Figure 13 shows exclusively a subset of endospore formers. The counts from the top three endospore formers, *Clostridia*, *Bacillus* and *Geobacillus*, were summarised at a strain level, highlighting previous observations from Figure 12 regarding the diversity of presence (or

absence) of species. All sRNAs involved have homologs in *Bacillus* species, 14 of which (S1574, S1042, S632, S4, S547, S732, S2, S1455, S198, S37, S283, S978, S1227, S1236) expand this to more distant relations, such as *Clostridia*, *Listeria* and *Staphylococcus*.

Cluster 1 is highlighted in green on Figure 12 and contains two sRNAs, S1042 and S1574, that are the most diverse. S1042 had a total of 400 homologs, 50.5% of which are found within *Bacillus*, ranging from the closely related to *B. subtilis, B. amyloliquefaciens,* to the more distant *B. anthracis*. Of the remaining homologs of S1042, 26% belong to the non-endospore forming *Staphylococcus*, primarily *S. aureus*, and 8.5% are counted among the *Listeria*, primarily *L. monocytogenes.* This is followed by the endospore forming *Paenibacillus, Geobacillus* and *Clostridium* (4.75%, 3.75% and 2.25% respectively). S1574 is the most diverse but has 51.5% of its 299 homologs belonging to *Bacillus*. In addition, S1574 is likely an antisense RNA and therefore unable to be differentiated between its conservation and that of the gene to which it is antisense. Like S1042, S1574 shares the highest numbers of remaining homologs with *Listeria, Clostridium* and *Paenibacillus*. S1574 has a small number of homologs for each of many endospore formers, as well as more distant bacteria such as the Gram-negative *Bacteroides*. With such diversity of organisms that are not closely related to *B. subtilis*, it would be considered unlikely that S1574 is specific to sporulation. Despite this, Nicholas *et al.* found high levels of expression exclusively during sporulation. As such, S1574 cannot be disregarded as having no role during sporulation.

Cluster 2, highlighted in blue in Figure 12, is the second most diverse group and is made up of the putative sRNAs S198, S1455, S4, S547, S2 and S732. S2 and S198 are largely made up of homologs within the *Bacillus* species (92.7% and 96.5%). S4 and S547 are again mostly made up of homologs within the *Bacillus* species (88.2% and 82.2%), with the addition of various other *Bacillaceae* and *Paenibacillaceae*. S732 and S1455 have homologs in additional, more diverse species of *Listeria* for S732 (7%) and *Staphylococcus* (<1%) for S1455.

Cluster 3, highlighted in yellow in Figure 12, is made of 23 putative sRNAs belonging almost exclusively to *Bacillus*, except for S1227, which unexpectedly has a homolog present in *Vitis vinifera,* the common grape vine. All 23 are present in numerous *B. subtilis* subspecies. Very few are more diverse, with the cluster only expanding to species such as *B. atrophaeus* amongst the most closely related *Bacillus* species.

Finally, cluster 4, highlighted in orange in Figure 12, contains 27 putative sRNAs. Homologs are primarily in *B. subtilis* and its closest ancestors, such as *B. amyloliquefaciens* and *B. atrophaeus*, and completely lacking in *B. cereus and B. anthracis.*

.

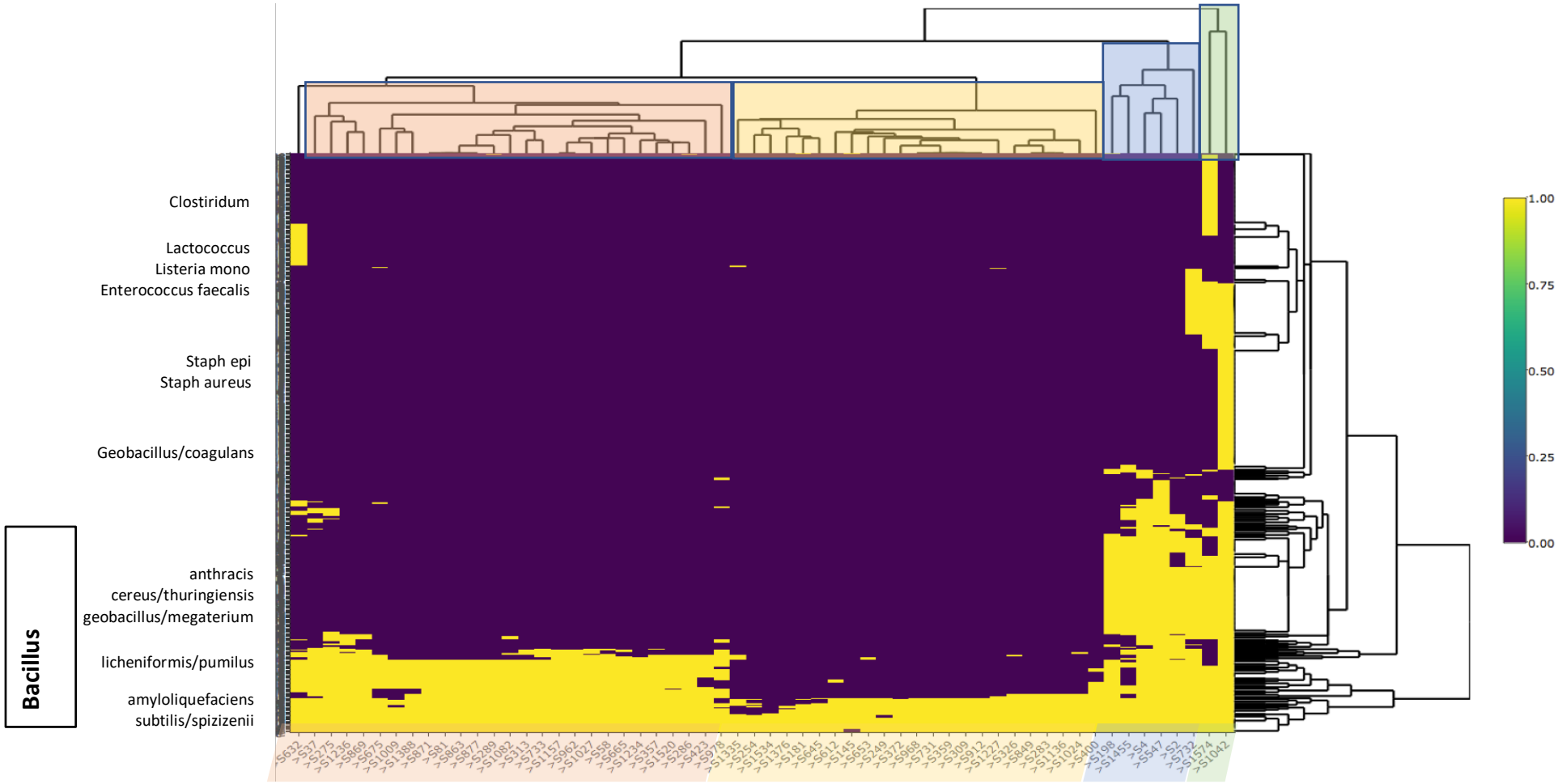


Figure 12 – Heat map of all selected sRNA conservation. The presence (yellow) or absence (purple) of an sRNA in a species was plotted for each sRNA based on homolog reporting by GLASSgo. sRNAs are on the bottom and clusters of similar patterns are highlighted in red, yellow, blue or green based on their species diversity. Purple indicates a lack of identification and yellow a positive identification. Species are summarised due to overpopulation of species names. Both sRNAs and species are reordered based on hierarchical clustering

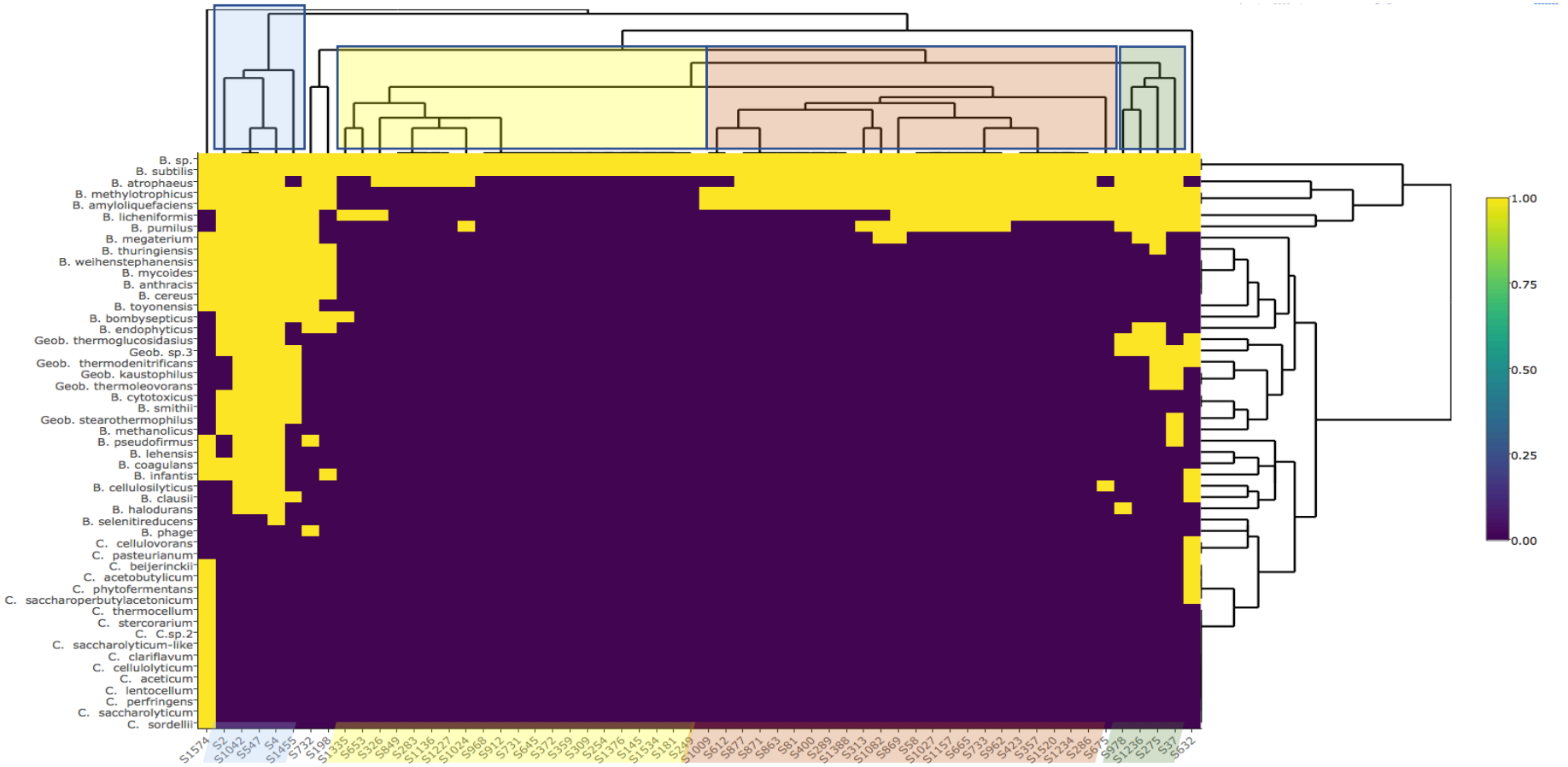


Figure 13 -Heat map of all selected sRNAs with three endospore forming species: *Bacillus, Geobacillus* and *Clostridium* for putative sRNA homologs. The presence (yellow) or absence (purple) of an sRNA in a species was plotted for each sRNA based on homolog reporting by GLASSgo. sRNAs are on the bottom and clusters of similar patterns are highlighted in red, yellow, blue or green based on their species diversity. Purple indicates a lack of identification and yellow a positive identification. Species are summarised due to overpopulation of species names. Both sRNAs and species are reordered based on hierarchical clustering

3.3.2.4.1 Evaluation of sRNA orthologue sequence alignments

Sequence similarity comparison approaches have been made possible due to many closely related bacterial species having been genome sequenced. It is well known that secondary structure is an important characteristic of functional RNA species and this can be directly predicted based on the sequence. Despite this, it is important to remember that whilst sequence similarity is important to a certain degree, particularly for the seed region of interaction, the sequence can "slip" surrounding the interaction region, whilst maintaining the crucial secondary structure (Savill, N. J. *et al.*, 2001).

The length of an sRNA had no correlation to overall alignment scoring of homologous sequences, despite the sequence having more bases and therefore more opportunity to deviate (Figure 14). As expected, the sRNAs containing homology with the most diverse species, identified from those within clusters 3 (blue) and 4 (green) (Figure 12), show a mild tendency to have lower alignment scores (Figure 14). As the constraints dictated that sequence similarity of 60% was needed, the alignment score is generally high as anticipated.

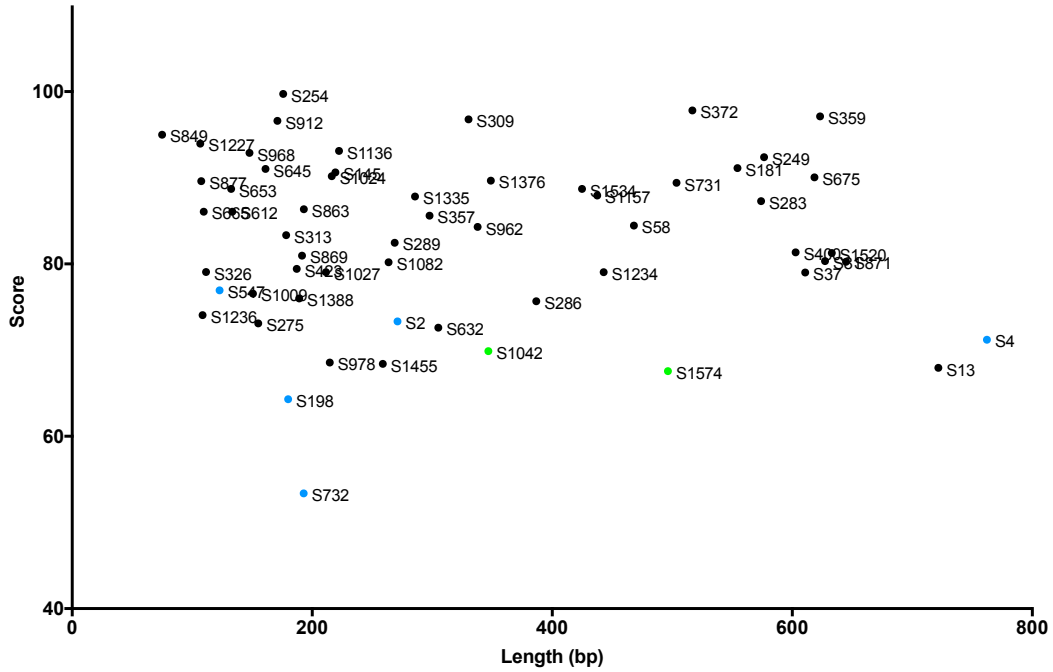


Figure 14 – Average length of sRNA homolog vs alignment score. sRNA homologs were aligned and scored based on homology. Alignment scores were compared to average length of homologous sRNAs to confirm no length bias.

S547 has been previously shown to be an sRNA expressed during sporulation with structural and sequence conservation (Marchais, A. *et al.*, 2011). Identified as being part of a cluster of conserved genes in endospore forming species, Marchais, A. *et al.* (2011) utilised Blastn to identify 24 orthologs of S547, which ranged from *Bacillaceae* to *Clostridia*, and alignment of these revealed a central conserved motif. To compare with this analysis, where homology was picked based on structural and sequence homology, of which the former was lacking from Marchais, A. *et al.* (2011), a total of 242 putative homologs from 50 species were found, 26 of which were *Bacillus*. In contrast to the previous predictions by Marchais e*t al.*, this approach did not find any orthologs in *Clostridia*.

MFE folding of the sequence from *B. subtilis* revealed a similar secondary structure as previously described by Marchais, A. *et al.* (2011), with their consensus motif present in a stem loop structure (Figure 15). Utilising RNAalifold (Bernhart, S. H. *et al.*, 2008) and the folding of a putative sRNA in the organism of interest it is possible to use covariance information from predicted homologs to create a consensus. Consensus folding of the identified homologs revealed a strikingly similar structure, again with the essential motif in a similar location (Figure 16). MFE scores for *B.*

subtilis (-41.50) and the consensus folding (-40.20) are similar. Comparing residue probabilities to see conservation per base, there is particularly strong conservation between nucleotide 22 to nucleotide 100 (Figure 17), which includes the putatively essential motif identified previously.

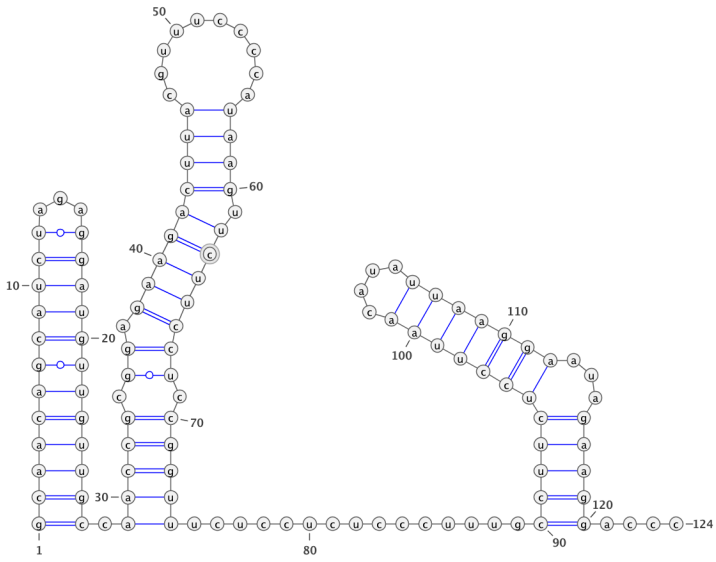


Figure 15 – S547 predicted secondary structure from *B. subtilis* 168 strain. Secondary structure predition was determined using RNAfold.

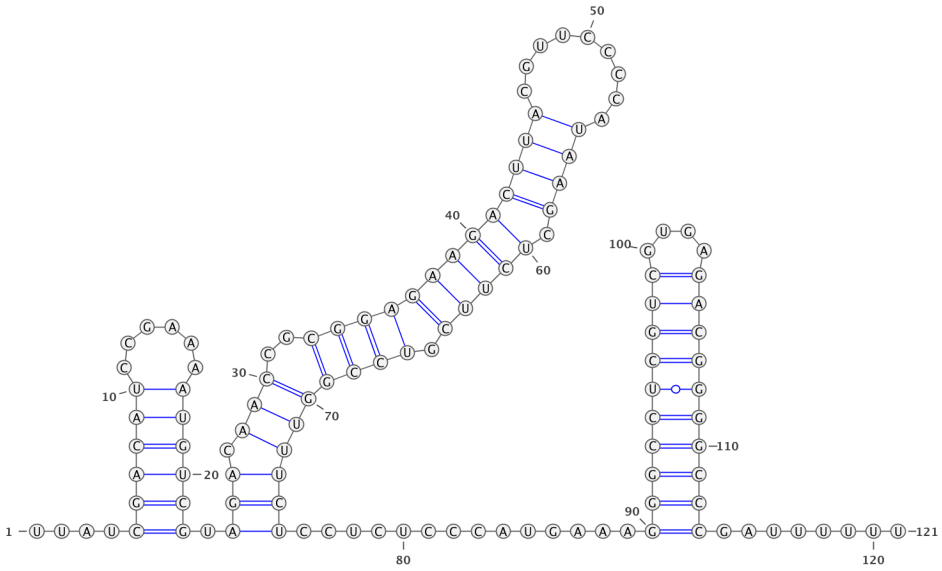


Figure 16 – Consensus predicted secondary structure for all homologs of S547. Multiple sequences were aligned by ClustalW. The secondary structure of the aligned sequences was determined by RNAfold analysis.

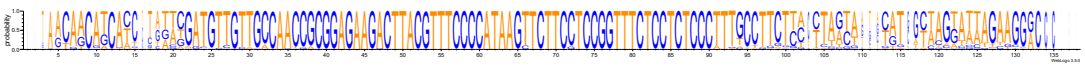


Figure 17 – Probability logo for S547 sequence alignments. Multiple sequences from homologs of S547 were aligned by ClustalW. Weblogo was used to show conservation of nucleotides at specific positions. The height of the letter indicates the conservation.

A total of ten putative sRNAs score 4 or above when the initial selection criteria was applied, S547 belongs to this group (Table 5). The ten putative sRNAs lie within different conservation clusters, with only the most diverse cluster not being represented (Figure 12, green). S547 and S2 are within the most diverse cluster. S2 homologs were identified in many *Bacillus* species such as in the *cereus sensu lato* group, in addition to others such as *Geobacillus* species. In terms of sequence diversity, S2 is very varied amongst the different homologs, however very brief sections of its sequence show complete consensus (Figure 18C). Despite this variation of sequence, a number of structural loops in its terminal region are predicted to be conserved. For example, both the single fold of *B. subtilis* and the consensus sequence folding have a terminal loop between nucleotides 120 and 150. In addition, there is a smaller loop between 170 and 180. A stem between two large loops is also particularly conserved (Figure 18A and B).

The next level of conservation includes the putative sRNAs S357 and S423 (Figure 12). Both have been shown by this study to have homologs entirely within *Bacillus*, primarily *B. subtilis* and *B. amyloliquefaciens* (Figure 12). These sRNAs have larger stretches of nucleotide conservation, which typically leave terminal loops that are predicted to be accessible. Those that are conserved primarily within *B. subtilis* subspecies exhibit greater sequence conservation, and hence greater structural conservation (Figure 19 A and B, Figure 20 A and B).

S645 is only reported to have 29 homologs, all entirely within *B. subtilis* and is very well conserved (Figure 21C). In particular, the 5' region of the sRNA and its putative homologs have both high sequence and structural conservation. Interestingly, for both the singled folded RNA from *B. subtilis* and the consensus folded structure show a loop structure with the same 4 AU rich nucleotides from 81 to 84 to be accessible (Figure 21 A and B). Similarly, S849 is a short RNA of 83 nucleotides with very narrow sequence conservation, specifically within *B. subtilis* and *B. atrophaeus* (Figure 12). The 5' and 3' regions of S849 are differing the most in sequence (Figure 22C). Despite this, much of S849 and its homologs are perfectly conserved. A stem-loop structure between bases 11 and 34 is entirely conserved, with only the terminal loop being

slightly less conserved where *B. atrophaeus* has an additional five nucleotides, but maintains the same stem structure (Figure 22 A and B).

The additional putative sRNAs, S968, S1024 and S1227, are similarly well conserved at a nucleotide sequence level (Figure 23, 24 and 25 C). In addition, both S968 and S1024 homologs are within the *Bacillus* species exclusively.

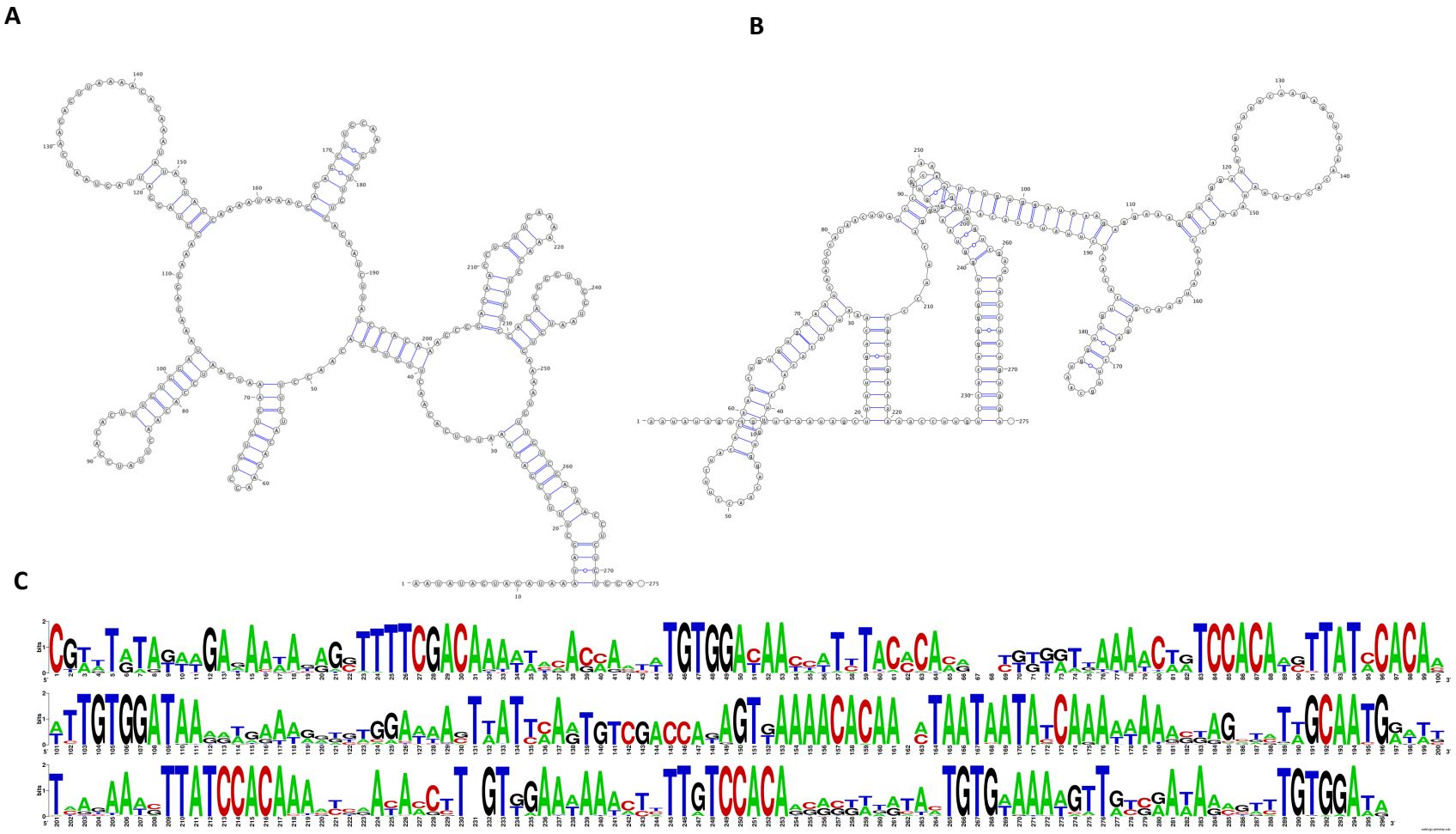


Figure 18 – S2 structure and conservation. (A) Multiple sequence alignment by ClustalW was used to compare homologous sequences before structural prediction by RNAfold. (B) Secondary structure of sRNA sequence from *B. subtilis* was conducted by RNAfold. (C) Probability logo shows per-base conservation of sequence. The conservation was analysed utilising Weblogo, with size of the letter indicating conservation. S2 conservation is varied, with brief sections of complete consensus.

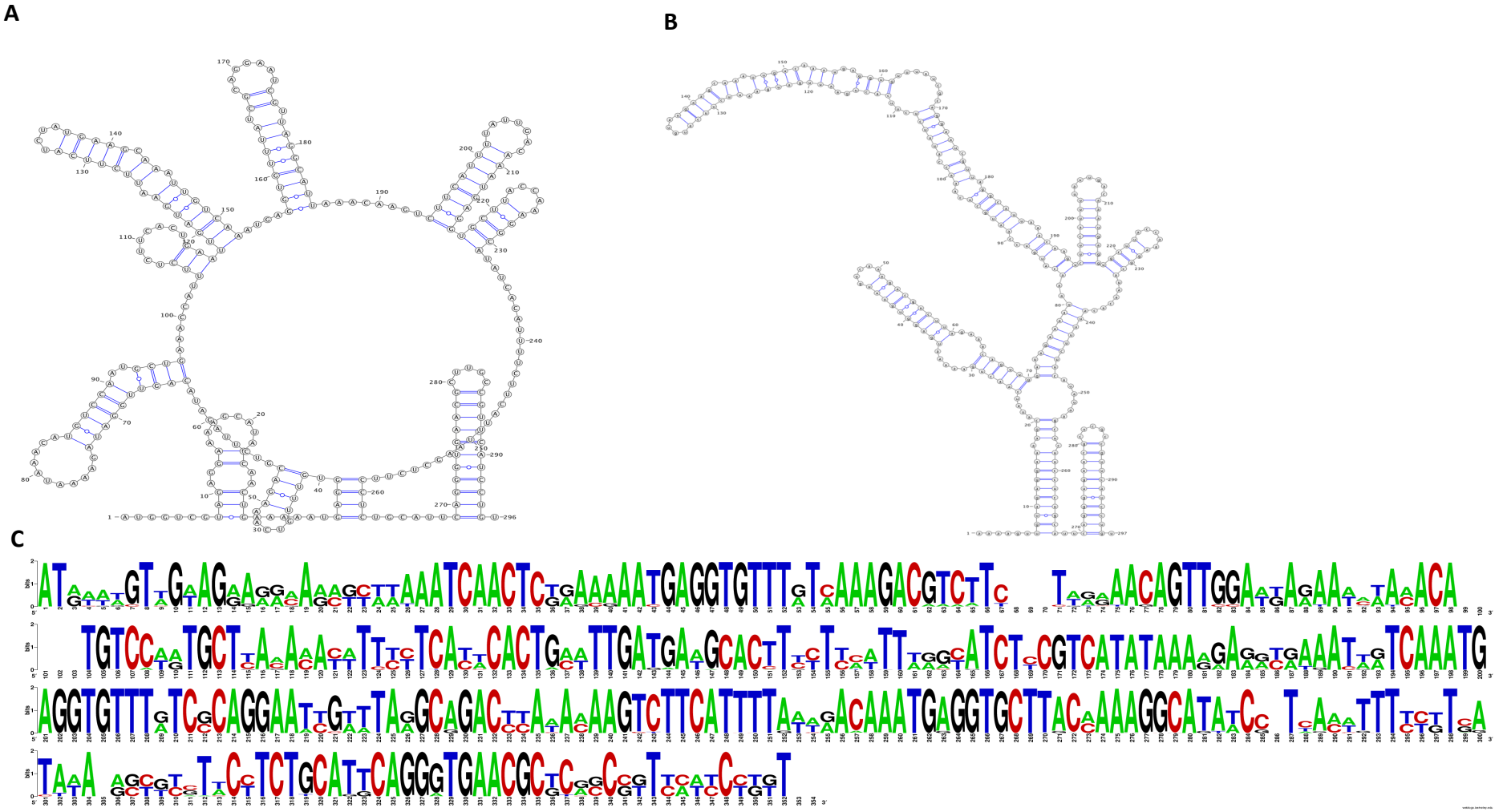


Figure 19 – S357 structure and conservation. (A) Multiple sequence alignment by ClustalW was used to compare homologous sequences before structural prediction by RNAfold. (B) Secondary structure of sRNA sequence from *B. subtilis* was conducted by RNAfold. (C) Probability logo shows per-base conservation of sequence. The conservation was analysed utilising Weblogo, with size of the letter indicating conservation.

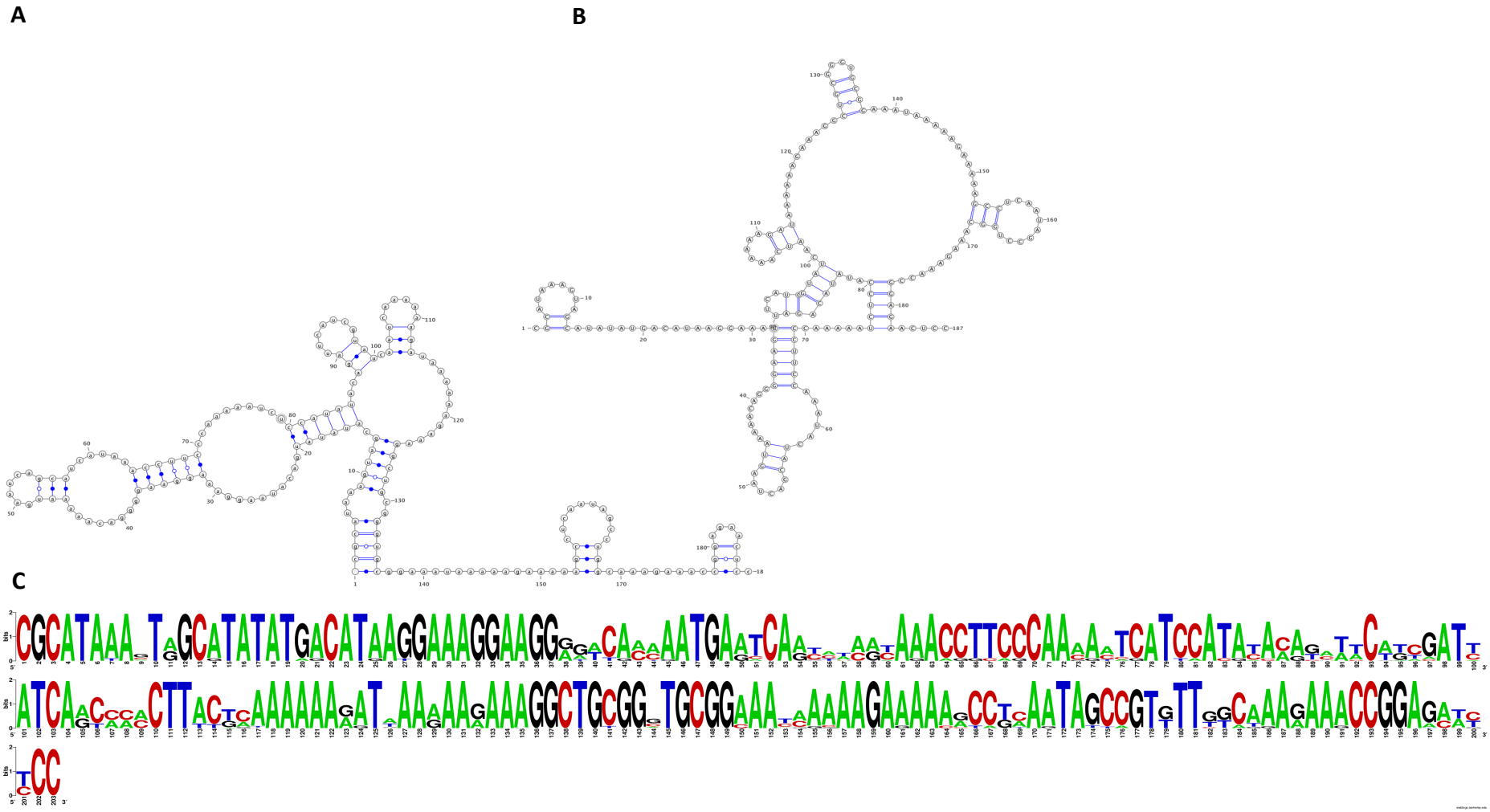


Figure 20 – S423 structure and conservation. (A) Multiple sequence alignment by ClustalW was used to compare homologous sequences before structural prediction by RNAfold. (B) Secondary structure of sRNA sequence from *B. subtilis* was conducted by RNAfold. (C) Probability logo shows per-base conservation of sequence. The conservation was analysed utilising Weblogo, with size of the letter indicating conservation.

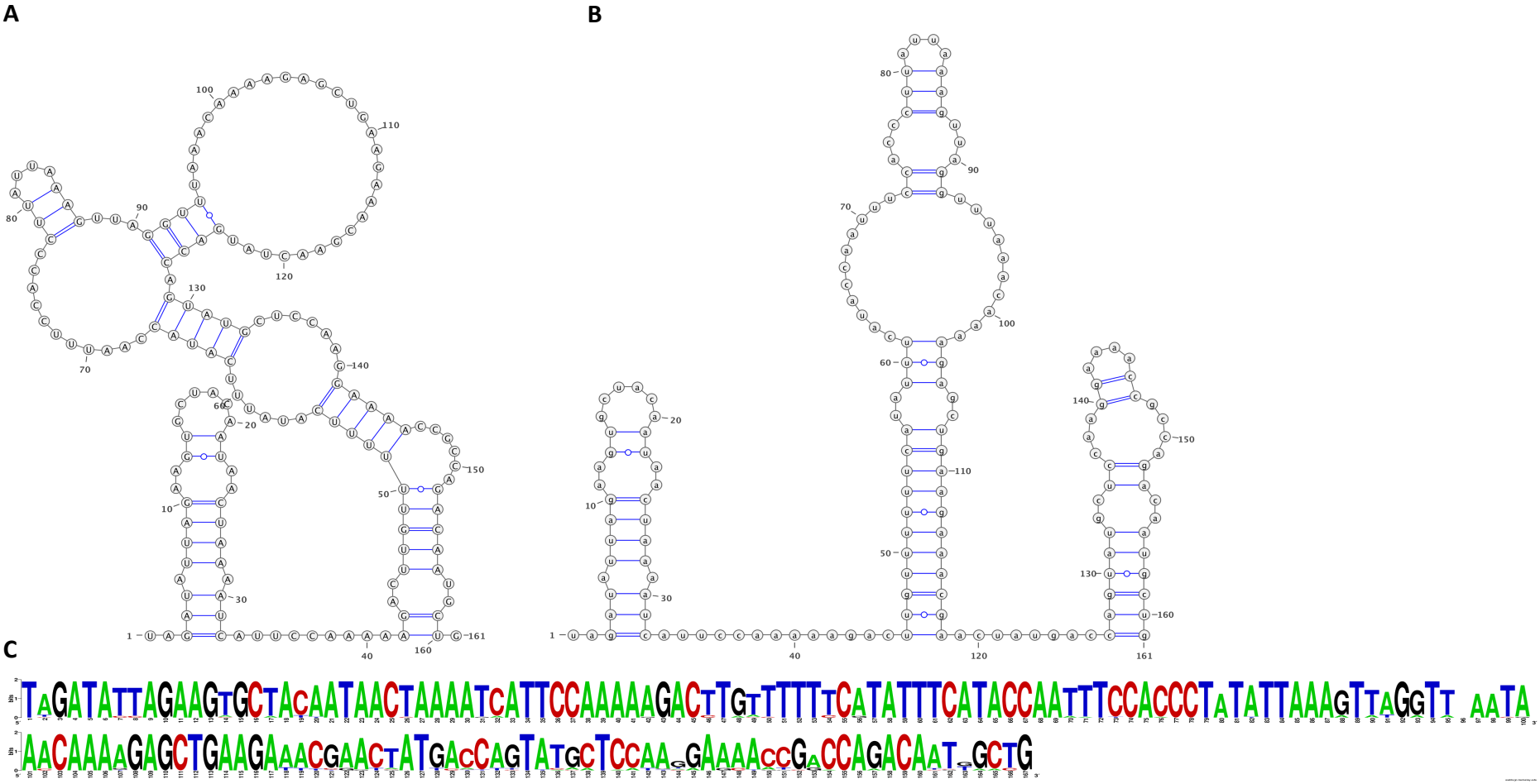


Figure 21 – S645 structure and conservation. (A) Multiple sequence alignment by ClustalW was used to compare homologous sequences before structural prediction by RNAfold. (B) Secondary structure of sRNA sequence from *B. subtilis* was conducted by RNAfold. (C) Probability logo shows per-base conservation of sequence. The conservation was analysed utilising Weblogo, with size of the letter indicating conservation. S645 is very well conserved within *B. subtilis.*

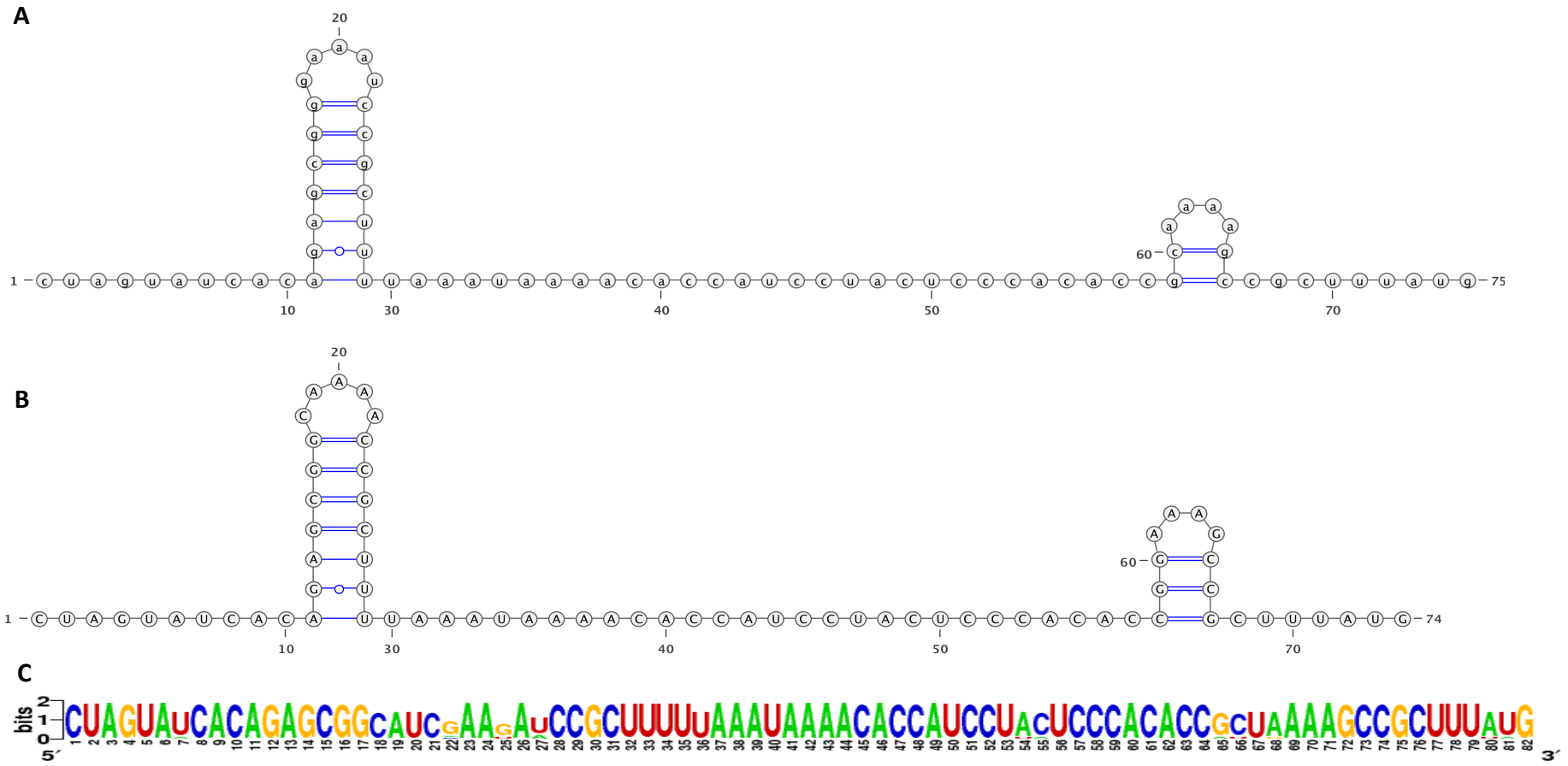


Figure 22 – S849 structure and conservation. (A) Multiple sequence alignment by ClustalW was used to compare homologous sequences before structural prediction by RNAfold. (B) Secondary structure of sRNA sequence from *B. subtilis* was conducted by RNAfold. (C) Probability logo shows per-base conservation of sequence. The conservation was analysed utilising Weblogo, with size of the letter indicating conservation. Many of the S849 homologs share hight sequence similarity and hence structural conservation

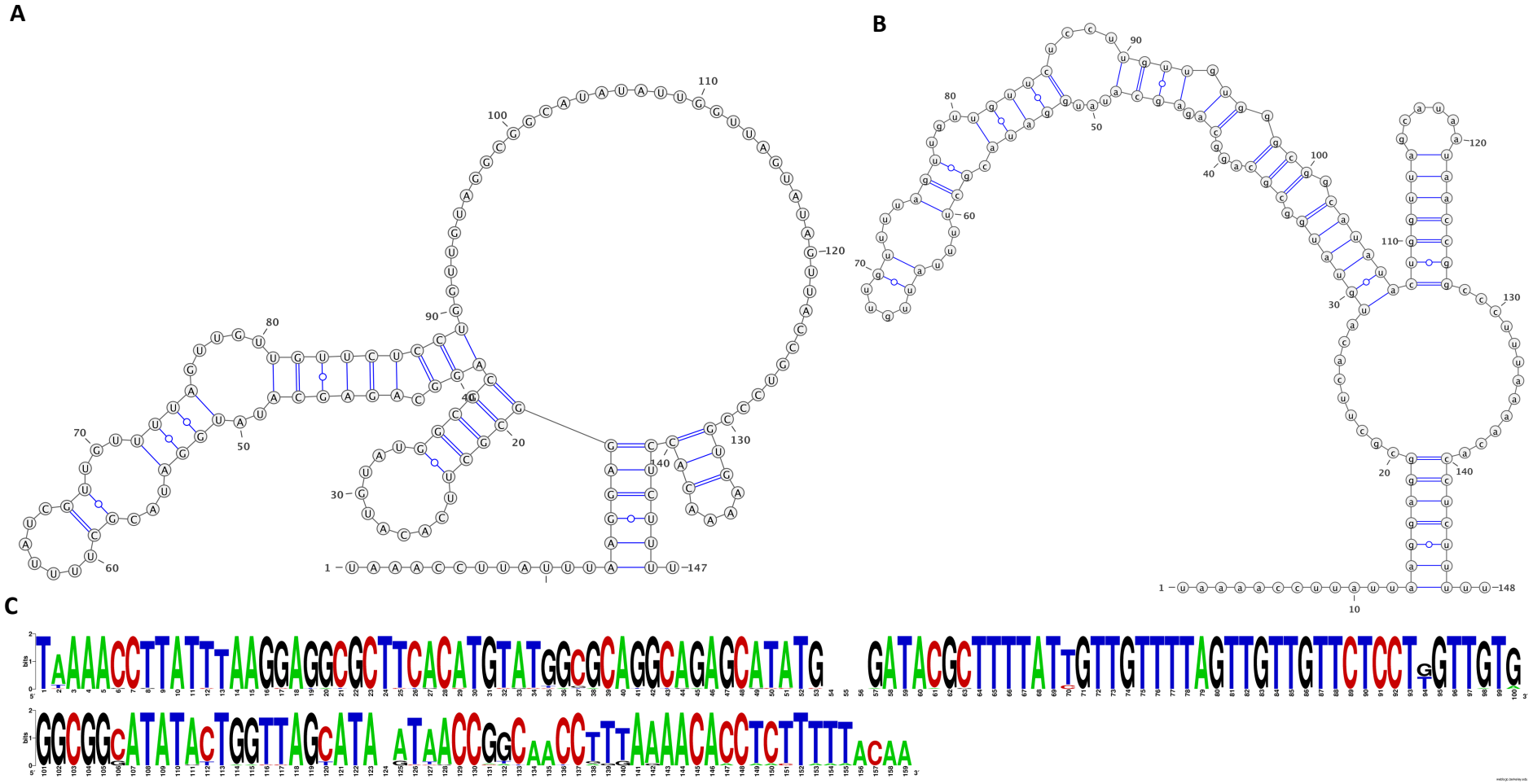


Figure 23 – S968 structure and conservation. (A) Multiple sequence alignment by ClustalW was used to compare homologous sequences before structural prediction by RNAfold. (B) Secondary structure of sRNA sequence from *B. subtilis* was conducted by RNAfold. (C) Probability logo shows per-base conservation of sequence. The conservation was analysed utilising Weblogo, with size of the letter indicating conservation.

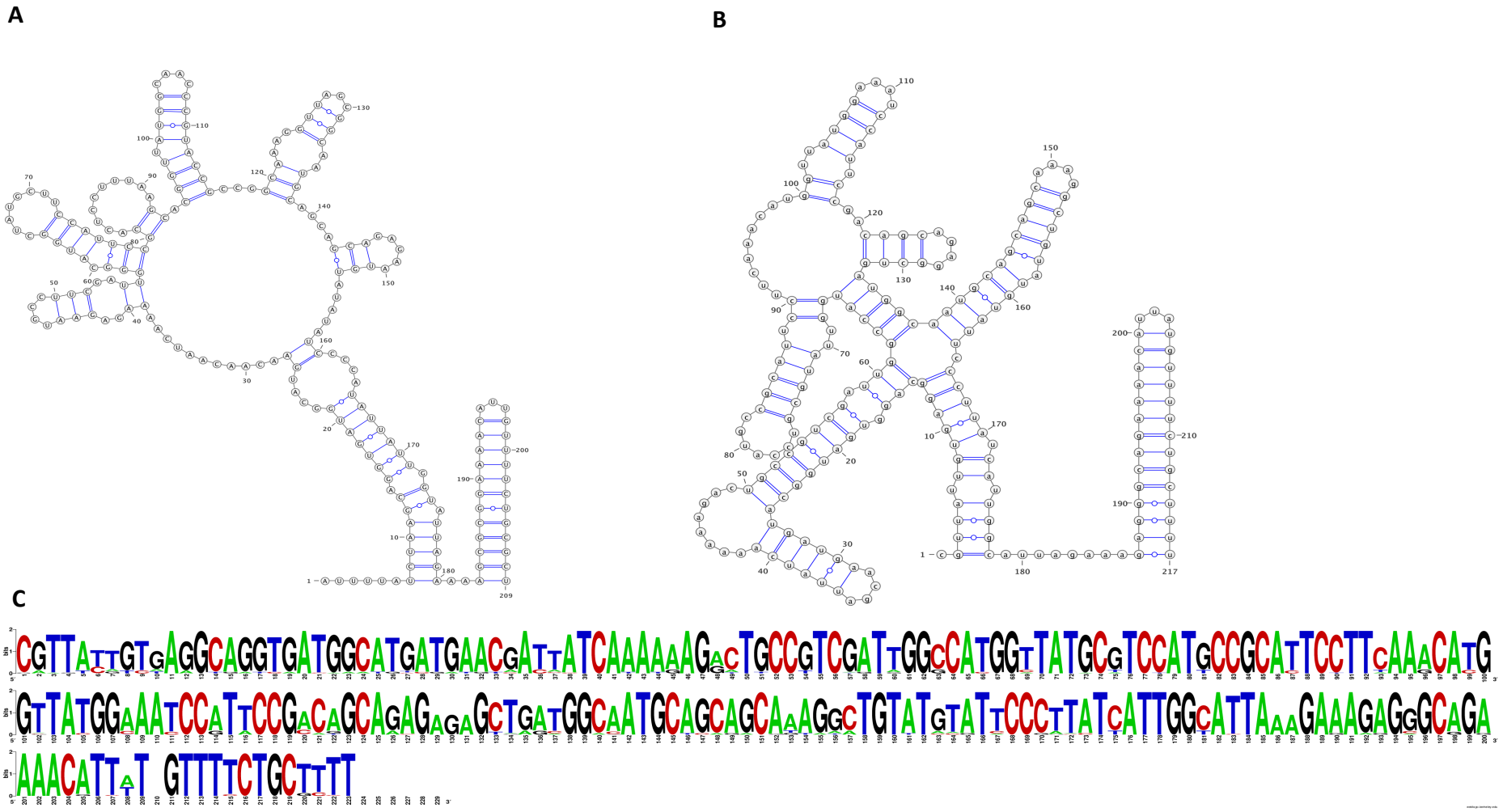


Figure 24 – S1024 structure and conservation. (A) Multiple sequence alignment by ClustalW was used to compare homologous sequences before structural prediction by RNAfold. (B) Secondary structure of sRNA sequence from *B. subtilis* was conducted by RNAfold. (C) Probability logo shows per-base conservation of sequence. The conservation was analysed utilising Weblogo, with size of the letter indicating conservation.

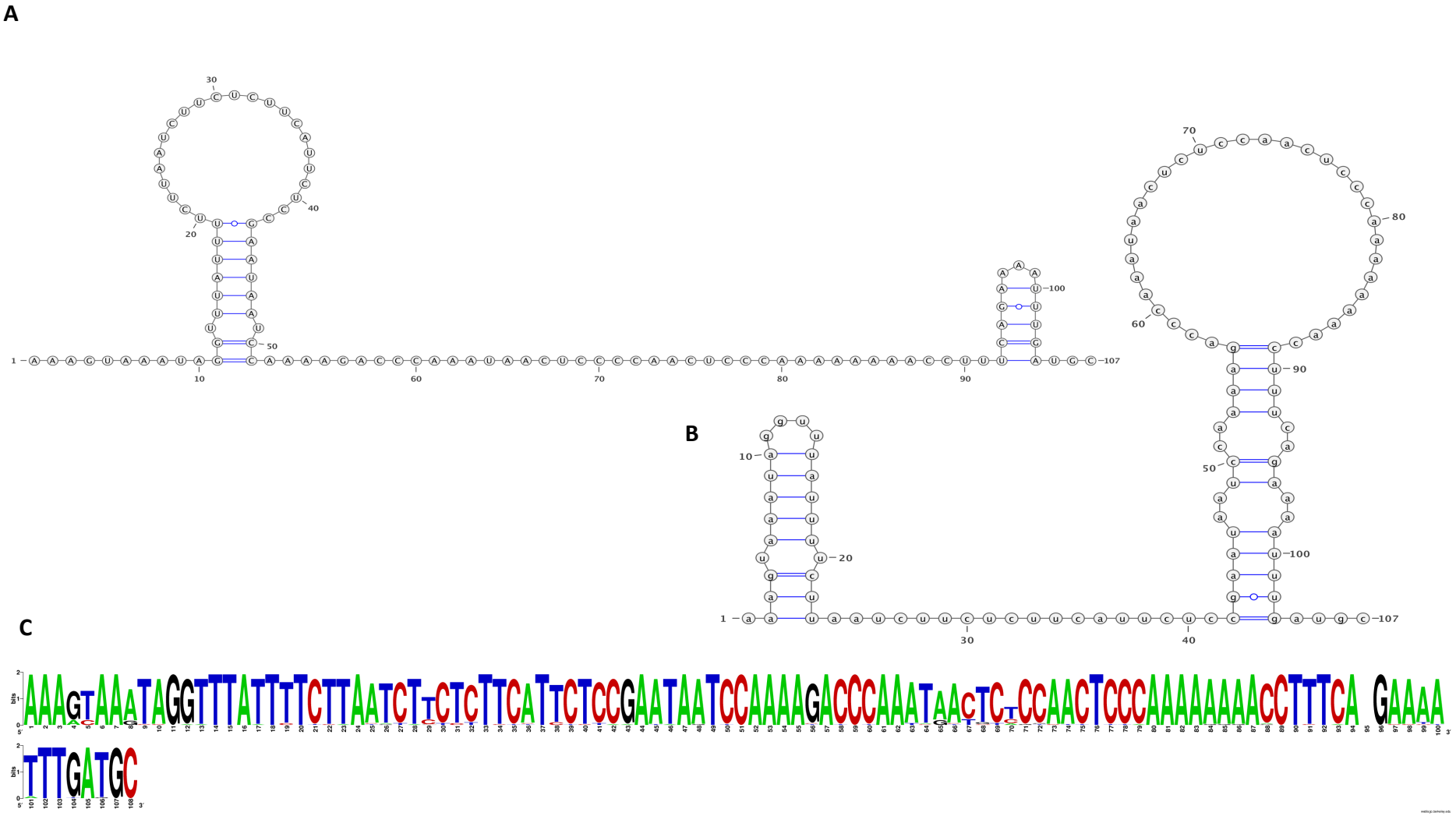


Figure 25 – S1227 structure and conservation. (A) Multiple sequence alignment by ClustalW was used to compare homologous sequences before structural prediction by RNAfold. (B) Secondary structure of sRNA sequence from *B. subtilis* was conducted by RNAfold. (C) Probability logo shows per-base conservation of sequence. The conservation was analysed utilising Weblogo, with size of the letter indicating conservation.

3.3.3 Analysis of sporulation-related next-generation sequencing data

3.3.3.1 Reanalysis of transposon sequencing data

Whole genome sequencing started with *Haemophilus influenzae Rd* in 1995 and now covers thousands of genomes, collected and freely available via GenBank (Fleischmann, R. D. *et al.*, 1995). The complexity of bacterial life has been revealed by the genomes being available for a diverse range of species, with comparative genomics allowing insights into genome evolution (Koonin, E. V. & Wolf, Y. I., 2008). Whilst knowing the nucleotide sequence is a crucial starting point, it is not enough to imply true understanding of function. As such, key to this understanding and keeping up with the high-throughput sequencing of genomes is the advent of high-throughput genotype-phenotype studies, such as transposon sequencing (Tn-seq) (van Opijnen, T. & Camilli, A., 2013).

Tn-seq is a high-throughput method that enables identification of genes with essential roles in a certain condition. Transposons are mobile genetic elements, segments of DNA that can be "cut and pasted" into target sites. This then interrupts genes by insertion, hence disrupting their function. Transposon mutants can be generated using Himar I Mariner transposon transformation in all non-essential genes, as if an essential gene was disrupted the resulting mutant will fail to thrive. By pooling a library of these mutants, with insertions evenly across all non-essential genes, genome-wide phenotype studies can be carried out (van Opijnen, T. *et al.*, 2015). Data analysis consists of counting the number of reads mapping to genes and hence the number of times a mutant appears in the control vs the tested condition, with the number of reads corresponding to the frequency of that mutant. For example, a decrease in the number of reads for a specific mutant is determined to be a decrease in the mutant's fitness to survive in a condition, and therefore the gene in question could be essential for that condition (van Opijnen, T. *et al.*, 2009)

Meeske, A. J. *et al.* (2016) utilised Tn-seq to identify essential genes involved in the process of endospore formation of *B. subtilis*. By comparing mutants from the start of sporulation, after 5 hours, and after 24 hours of sporulation they could also

further characterise those genes involved in the timing of sporulation development. Whilst new discoveries of function are encouraging for known annotated genes, Meekse *et al.* did not analyse the potential for sRNAs to be involved, and therefore this study reanalysed the data to identify any non-coding regions that could be important during sporulation. Utilising a pipeline for the analysis of high-resolution transposon-insertion sequences technique (ARTIST), first a genome map of *B. subtilis* insertion sites was created and all the reads mapping to transposon sites were counted. This was repeated for the three time points: Control taken prior to the start of endospore development, hour 5 of sporulation and after 24 hours.

28 putative S segments were identified to have a large ratio change from the control to the sporulation conditions and can be seen in Table 9. Eleven of these S segments were identified by the workflow presented previously in this chapter (Table 5, highlighted in yellow in Figure 12). Whilst little is known currently about the majority of these RNA species, some give more indication of potential roles during sporulation, such as those genes to which they are antisense and therefore noted in Table 9. Further validation is required to confirm the deficiencies in sporulation for the putative sRNAs within Table 9.

Table 9 – The number of reads mapped to each putative sRNA with a log2 ratio higher than 1.5 in at least one condition when comparing the control experiment to the sporulation conditions. Highlighted in yellow text within the table are those that are also in the selected sRNAs for future analysis.

sRNA	Control	T5	T24	log2 Ratio T5	log2 Ratio T24	Sigma factor	Notes
S1027	5	0	0	-2.58	-2.58	SigA, SigA	
S111	16	4	26	-1.77	0.67	SigEF	• S112 antisense – 5' UTR of *aroK*
S1227	7	0	2	-3.00	-1.42	SigK, SigA	
S1243	7	1	5	-2.00	-0.42	SigB	
S1251	22	24	3	0.12	-2.52	0	• upregulated during the later stages of sporulation.
S1292	22	35	4	0.65	-2.20	SigA, SigA, SigA	• opposite *yvaV*, overlapping at the 3' end. • *yvaV* is unknown but its product is purported to be similar to transcription factor OpcR • OpcR is a transcriptional repressor which aids the regulation of choline uptake (Lee, C. H. *et al.*, 2013).
S1359	6	0	0	-2.81	-2.81	SigK	
S140	11	1	11	-2.58	0.00	SigA, SigB, SigA	• Development? Read-through from upstream – *ydaG* • gene expression patterns between S140 and *ydaG* do overlap significantly, apart from a few conditions such as germination and cold stress. • Neither is particularly upregulated during sporulation

S1459	17	25	0	0.53	-4.17	SigB	• Antisense to *bacA*
S1579	22	0	0	-4.52	-4.52	SigA	• Antisense to *spoIIIJ* and *jag*
S163	33	16	6	-1.00	-2.28	SigD, SigB, SigK	• Antisense to *rsbRA* and *ndoA.* P • Purported to work on 3' region *ndoA*
S198	75	36	16	-1.04	-2.16	SigA	• Riboswitch for *vmlR*
S214	12	0	9	-3.70	-0.38	SigA	• Antisense *ydzW* – a putative phosphomannomutase, which is active in the initial stages in sporulation. • S214 not strictly active in sporulation
S235	28	1	16	-3.86	-0.77	0	• S235 appears to not be upregulated in sporulation but is high in biofilm forming conditions. • Opposite is *sapB*
S249	115	36	39	-1.65	-1.54	SigK	
S25	27	3	6	-2.81	-2.00	SigWXY	• S25 is antisense to both *yabD* and *yabE*.
S265	26	1	23	-3.75	-0.17	SigEF	• S265 is in the selected list of sRNAs, being present in cluster C2. • Antisense *yflH*, *yflG* and *yflI*.
S309	66	20	15	-1.67	-2.07	SigA	
S326	16	3	3	-2.09	-2.09	SigGF	
S348	26	139	8	2.37	-1.58	SigA, SigA	• Antisense – another new feature
S423	19	0	5	-4.32	-1.74	SigK, SigEF	

S458	6	1	0	-1.81	-2.81	SigA, SigEF	
S462	119	34	49	-1.78	-1.26	SigWXY	• Peculiar profile during sporulation
S562	6	0	0	-2.81	-2.81	SigA	
S796	13	2	2	-2.22	-2.22	SigA	• S796 is part of cluster C67 • shows upregulation of expression during late sporulation. • S796 is located upstream of another S segment, S797 • appears not have its own terminator.
S9	5	1	0	-1.58	-2.58	SigB	
S903	10	1	9	-2.46	-0.14	SigA, SigA	• pseudogene, involved in zinc limited conditions as an alternative ribosomal protein. • S903 is upregulated during hour 2 of sporulation.
S977	150	33	95	-2.15	-0.65	SigA, SigB	• *bsrH* – alternative name

3.3.3.2 Germination

Although metabolically dormant, endospores do contain RNA which can be used for the initial burst of activity needed for germination (P, Setlow & A, Kornberg, 1970). Previous studies have identified two main functions of this RNA. The first, which rapidly disappears during outgrowth, are those transcripts essential for the start of germination (Keijser, B. J. *et al.*, 2007) and the second, presumably left over RNA transcripts from the later stages of sporulation, displays a decline in RNA levels as dormancy is maintained and is due to the RNA within being degraded by RNase Y (Segev, E. *et al.*, 2012). Degraded RNA can then be used for *de novo* synthesis of RNA for a rapid germination response (Segev, E. *et al.*, 2012).

Nagler, K. *et al.* (2016) studied the transcriptional landscape of endospores during outgrowth via RNAseq, including in high-salt environments. Whilst a thorough analysis was performed on protein-coding genes, sRNAs were not analysed. Data was extracted from the short reads archive (SRA) of the NCBI for wild-type endospores and the expression data from this study was specifically examined for putative sRNAs.

3.3.3.2.1 Independent RNA features can be found in endospores

S357 is the sixth most abundant RNA found in a dormant endospore and follows a similar expression pattern throughout germination comparable to *rnpB*, the fifth most abundant RNA, which is an RNA component of RNase P. RNase P directs the cleavage of precursor sequences from the 5' ends of pre-tRNAs (Guerrier-Takada, C. *et al.*, 1983). S357 could have similar roles to that of the most highly expressed genes in a dormant endospore, which appear to deal with translation or are active in the later stages of sporulation. The ten most abundant putative sRNAs expressed during sporulation can be found in Table 11.

3.3.3.2.2 Abundant RNA transcripts in a dormant endospore belong to non-coding RNAs

Genes in the list of the ten most abundant genes are likely to either be from the last stages of endospore formation, or important for germination. Genes in the top 10

most highly expressed genes in a dormant endospore include four other new RNA features, or S segments, identified by Nicholas *et al.* (2011) (Table 10). One of these is an intergenic region, S1554, that is associated with a tRNA operon, and the 5' and 3' regions of *scr*, S17 and S18 respectively. *scr* itself is the most abundant RNA present in a dormant endospore. *scr* is a small cytoplasmic RNA that is transcribed with S17 and S18, which is then cleaved to reveal a functional product (Yao, S. *et al.*, 2007) that then causes RNA-driven elongation arrest by interacting with 23S rRNA. When deleted, this gene has a mild sporulation defect (Nishiguchi, M. *et al.*, 1994).

Further non-coding RNAs featured in the ten most abundant genes within a dormant endospore is both *ssrA* and *rnpB* (Table 10). *ssrA* encodes tmRNA, which functions as both a tRNA and mRNA, and rescues ribosomes stalled in translation. *ssrA* deletion results in a defect in endospore formation (Abe, T. *et al.*, 2008). This is followed by *rnpB,* another ncRNA dealing with "housekeeping" functions that is the RNA component of RNase P.

Sporulation-related transcripts are also present in high amounts in a dormant endospore (Table 10). First, *yhcV*, the function of which is unknown, but has been shown to be expressed in the forespore and is a member of the SigG regulon (Wang, S. T. *et al.*, 2006). *sspF* is another highly expressed sporulation-associated gene that encodes for one of the minor small acid-soluble endospore proteins (SASP). SASPs are important for protection of dormant endospores from UV irradiation. Finally, within the ten most abundant genes is *yfhD*, coding for a general stress protein usually involved in the response to ethanol stress and low temperature. However, it is also upregulated during sporulation and has an antisense RNA shown by Nicolas *et al*.

Table 10 – Number of mapped reads of an RNA a dormant endospore from Nagler, K. *et al.* (2016). Ten most abundant transcripts.

sRNA	Depth of coverage
scr	1521370
S17	1479542
ssrA	1258456.5
S18	1041395.5
rnpB	457547.5
S357	311046
yhcV	80393
sspF	59725.5
S1554	50262.5
yfhD	49457

Table 11 – Numbver of mapped reads of a dormant endospore from Nagler, K. *et al.* (2016). Five most abundant putative sRNAs of previously selected independent features from this study after S357.

sRNA	Depth of coverage	Additional information
S357	311046	S357 is the sixth most abundant RNA found in a dormant endospore and follows a similar expression pattern throughout germination comparable to *rnpB*
S31	23111.5	S31 is present in large amounts in the dormant endospore, then gets downregulated during germination, but still present in large quantities. S31 is upregulated during hour 3 of sporulation and continues to be present in high numbers until the end of sporulation. S31 is also active during glucose exhaustion (Nicolas, P. *et al.*, 2012). S31 is antisense to *purR, yabJ* and *ispE. purR* and *yabJ* are both involved in the regulation of purine biosynthesis.
S275	15924.5	S275 is present in large amounts in the dormant endospore, but the RNA level is greatly reduced 30 mins after the onset of germination. S275 is not antisense to another gene. S275 is transcribed as part of two transcriptional units.

S1279	1497.5	S1279 is antisense to "BSU_misc_RNA_54" and partly antisense to *yvsH* – a putative lysine transporter. S1279 is massively upregulated during germination.
S1136	1081.5	S1136 is antisense to another sRNA, S1137, and *rpsD*, a ribosomal protein S4. S1136-S1134 have been shown to reduce expression of *rpsD* during ethanol stress (Mars, R. A. *et al.*, 2015a). S1134 and *rpsD* both do not have a large presence in a dormant endospore, but after 30 minutes of germination the expression escalates rapidly in similar levels for both. S1136 follows the opposite pattern, being high in a dormant endospore but reducing after initiation of germination, suggesting that in this instance S1136 is not working with S1134. S1137 is partially antisense to S1136 follows the same pattern as S1136 during germination.
S1180	257	Whilst being present in relatively large amounts within a dormant endospore, S1180 is further upregulated during germination. S1180 is antisense to *glgB*, which is part of an operon controlled by SigE in the mother cell that encodes enzymes involved in glycogen metabolism (Kiel, J. A. *et al.*, 1994). Whilst being expressed during sporulation after hour 2, *glgB* is not active during germination nor present in dormant endospore.

3.4 Discussion

3.4.1 Initial selection of sRNAs

There are many different programs available for the characterisation of putative sRNAs, with a variety of different features to use as potential signs to predict the likelihood of a sequence being an sRNA. As such, the proposed workflow was utilised to select putative sRNAs which warrant experimental investigation in the lab.

Vital to being expressed during sporulation is the possession of a promoter motif which is recognised by one of the sporulation-specific sigma factors. Reanalysis of the data available from Nicolas *et al.* indicated that 61 such promoter sequences could be found for selected sRNAs by this study. Additionally, the average expression profile indicated that genes and sRNAs in both clusters C25 and C74 are likely later acting in the process of sporulation (Figure 7). Indeed, when arranged into the given clusters by Nicolas *et al.* this study found that many of the predicted promoters were for SigK, the late acting mother cell specific sigma factor (Table 6B).

DBTBS was used to further identify transcriptional motifs, including transcriptional regulators. One such predicted regulator is SpoIIID, an early regulator of mother cell genes (Eichenberger, P. *et al.*, 2004). Two putative sRNAs were revealed to have a potential SpoIIID binding site, S1042 and S645, however S1042 did not meet the cut-off for further study (Table 5). Whilst identification of this important sporulation-specific regulator would typically be a positive identifier for a good candidate, S1042 is antisense to *yrrC*, a 5'-3' DNA helicase gene (Walsh, B. W. *et al.*, 2014). Being antisense to *yrrC* would make it difficult to carry out functional studies as interruption of *yrrC* has been shown previously to have an increase in spontaneous mutations (Walsh, B. W. *et al.*, 2014). S645 expression appears to be specific to sporulation, starting at hour 3 and peaking at hour 7 (Nicolas, P. *et al.*, 2012). S645 is antisense to another S segment, S644, which is the 3' UTR of *yncM,* a secreted protein of unknown function (Voigt, B. *et al.*, 2009).

This study identified six putative sRNAs with predicted CodY binding motifs (Figure 3). Differential gene expression analysis of previously published data showed that two of the predicted CodY binding sites for sporulation-related sRNAs, S1234 and S1543, were indeed under the control of CodY (Table 7). In addition to this, a further six independent transcripts were considered significantly altered and therefore likely warrant further study under conditions CodY typically acts in. Neither S1234 nor S1543 met the cut off score to warrant further study (Table 5), for reasons set out below.

S1234 is only expressed after hour 4 of sporulation (Nicolas, P. *et al.*, 2012) and appears to be in an operon with more S segments – S1235 and S1236. S1234 is expressed antisense to *yuiC*, which is a stationary phase survival and sporulation gene, expressed in the forespore under control of SigF (Wang, S. T. *et al.*, 2006). *yuiC* is also repressed by CodY (Mader, U. *et al.*, 2002b). *yuiC* and S1234 display differing expression patterns and do have differing numbers of transcripts in both wild-type and *codY* null mutant. S1234 also overlaps the 3' end of *yuiB,* a hypothetical protein that is also under CodY control. For these reasons, it is possible that this sRNA is acting in antisense.

S1534 appears to be the start of a large operon comprising of the genes *yxbB*, *yxbA*, *yxnB*, *asnH* and *yxaM*. This operon is known to be expressed in the transition to stationary phase. Two mRNAs are produced that cover this region. One transcript covers the whole operon, while the other results in a shorter transcript consisting of the first three genes: *yxbB*, *yxbA* and *yxnB* (Morinaga, T. *et al.*, 2010). Both are under the control of CodY and AbrB, regulators of stationary phase and sporulation genes. Previously, studies of the 5' region of *yxbA* identified a long sequence triplication segment, consisting of two conserved 118 bp units and a less conserved 129 bp unit (Yoshida, K. *et al.*, 1995). This was then attributed to lending stability to the shorter transcript consisting of three genes, in a similar manner to *ompA* in *E.* coli (Morinaga, T. et al., 2010) whereby the 5' region contains stem-loop structures and two RBS, which stabilise the mRNA through binding and preventing translation (Yao, S. &

Bechhofer, D. H., 2009; Morinaga, T. *et al.*, 2010). This 5' region is S1534 and therefore its function is probably as a regulatory 5' UTR.

Extraction of data available from Nicolas *et al.* indicated that two of the selected sRNAs had elongated transcripts in a Δ*rho* background. However, an elongation of a transcript of less than 50 nucleotides is rather short to be defined as truly *rho* dependent. Both S1202 and S951 are long transcripts and both can be seen to diminish transcriptional activity towards the 3' end. Genetic disruption of S951 is also likely to affect the essential genes *sigA* and *dnaG* to which it is antisense and as such, this sRNA does not fit the criteria for inclusion in this study. Additionally, S1202 was also predicted to have a *rho*-independent termination region by this study and therefore unable to predict the exact mechanism of termination for this sRNA.

Two further Indep-NT segments, those shown by Nicolas *et al.* to not possess a downshift in expression, were predicted by this study to have a region capable of rho-independent termination. One of these, S111, upon analysis of the tiling array profile can be seen to suggest some termination with a clear partial downshift in expression (Figure 10). As such, it is feasible that S111 is an indep segment. The second of these is S145, and its expression profile by Nicolas *et al.* is unclear and therefore it is not possible to conclude the termination mechanism.

In summary, clustering by Nicolas *et al.* facilitated the capture of putative sRNA sequences to create a list of 85 newly annotated S segments that are likely to be independently expressed (Table 5). The reanalysis of published data is an invaluable source of information and can give a good basis for the determination of a putative sRNA, particularly when combined with other available tools. Using basic criteria such as expression profiling with promoter and termination predictions successfully narrowed down potential candidates, but experimental evidence is essential for confirmation. Regardless, it is possible to take those identified here and tailor experiments to further characterise putative sRNAs involved in the various processes of endospore formation.

3.4.2 Further characterisation

3.4.2.1 sRNA homologs

sRNAs have been found to be widely conserved across species (Peer, A. & Margalit, H., 2014), but also can be strain specific (Dugar, G. *et al.*, 2013). This may be due to the selective pressures of its targets. There are many essential genes in the process of endospore formation that when disrupted block the formation of the endospore (Dembek, M. *et al.*, 2015; Galperin, M. Y. *et al.*, 2012). These genes are required to be expressed and their products act in a sequential manner, which leaves this process vulnerable to evolutionary disruption (Steil, L. *et al.*, 2005; Eichenberger, P. *et al.*, 2004). *Listeria* and *Staphylococcus* species lack almost all essential sporulation genes and have previously been used to exclude genes when attempting to find a minimal sporulation gene set (Eichenberger, P. *et al.*, 2003). Auxiliary functions during sporulation, however, can be important in the process of endospore formation despite not blocking sporulation entirely, and may not be conserved throughout all endospore formers.

Within *Bacillus,* all the sRNAs tested were found to present homologs in *B. subtilis* and its close relatives, with the addition of the more distantly related *B. cereus sensu lato* group reflecting previous observations (Mars, R. A. *et al.*, 2016). Interestingly, genes early on during sporulation (SigE-controlled) are more conserved than those later in sporulation (SigK-controlled) (Eichenberger, P. *et al.*, 2004). This has been hypothesised to be due to the potential of environmental differences, as those switched on later are in contact with the environmental stresses experienced by the end-product (i.e. coat proteins). As such, the sRNAs within cluster 3 could be early acting, or those not involved with genes involved in the protection against environmental stresses.

One of the most conserved putative sRNAs is S1574. It should be noted, however, that antisense to S1574 is *exoA*, a DNA-repair endonuclease that is important in endospore dormancy, which could be the reason why this sRNA is so widely conserved (Moeller, R. *et al.*, 2014; Moeller, R. *et al.*, 2011).

3.4.2.2 Target predictions

One of the foundational characteristics of sRNAs is that they accomplish their effects via base pairing with their targets. Knowing this is advantageous, but is not overly useful, since the base pairing required is often short and imperfect. Therefore, target prediction algorithms cannot solely rely on sequence-matching as many false positives are generated due to the imperfect mechanism of base pairing (Tjaden, B. *et al.*, 2006). Conservation is now incorporated into target prediction searches (Wright, P. R. *et al.*, 2014) and methods of predicting targets of an sRNA are improving but, as very little is known for the sRNAs in this study, target predictions was not performed at this stage.

3.4.2.3 Reanalysis of next-generation sporulation data

3.4.2.3.1 Transposon sequencing can indicate essential sRNAs

Transposon mutagenesis coupled with deep sequencing facilitated Meeske, A. J. *et al.* (2016) in the identification of many essential genes in sporulation, including those which cause a developmental delay. Reanalysis of this data revealed 28 putative S segments with significant ratio changes from control conditions to sporulation conditions, indicating that these mutants are lacking a fitness advantage and therefore are lost during the sporulation process.

S198 is in cluster C17 and as such one of the selected putative sRNAs analysed during the previous part of this chapter. Mutants in S198 are slightly less present after 5 hours (-1.04), but more so after 24 hours (-2.16). It is, however, upstream of the antibiotic efflux ABC transporter *vmlR*, which is slightly upregulated during sporulation. *vmlR* has been previously shown to have a riboswitch element (Dar, D. *et al.*, 2016; Irnov, I. *et al.*, 2010; Ohki, R. *et al.*, 2005). Whilst it seems like S198 is the regulatory element for *vmlR*, it is also conceivable for it to function independently as a *trans*-acting sRNA.

S1459 appears to have no developmental delay but an absolute one. *bacA* is antisense to S1459 that encodes a bacilysin biosynthesis protein. *bacA* is upregulated

during the first hour of sporulation and might be active as a defence mechanism during the initial stages of entering sporulation. This, however, does not explain the reason S1459 might be essential during sporulation. However, it is also expressed in the late stages of sporulation, and as such might have a crucial role there. Transcription of S1459 would not have been activated at hour 5 and therefore this accounts for the presence of the Tn mutants at hour 5.

S977 is also called *bsrH* and has been previously identified in searches for sRNAs in *B. subtilis* (Rasmussen, S. *et al.*, 2009; Saito, S. *et al.*, 2009). BsrH is known to be expressed during growth conditions and is not upregulated in sporulation (Saito, S. *et al.*, 2009; Nicolas, P. *et al.*, 2012). There is however a decrease in Tn mutants at hour 5 (-2.15) compared to the control, in comparison to the other putative sRNAs suggesting a potential role.

S265 is another antisense sRNA to three genes, two of unknown function *yflH* and *yflI,* and *yflG* a methionine aminopeptidase. Expression of S265 is activated when expression of *yflG* drops during sporulation conditions. This could indicate S265 to be acting in antisense to downregulate levels of *ylfG*.

Antisense to S25 is *yabE*, a protein of unknown function, but has similarity to a cell wall binding protein. It has previously been demonstrated that *yabE* is controlled by a cis-acting antisense RNA (S25) which it itself is controlled by SigX and SigM (Eiamphungporn, W. & Helmann, J. D., 2009). A *yabE* deletion mutant yielded no substantial deficiency in sporulation (Eiamphungporn, W. & Helmann, J. D., 2009), but this could be exacerbated in this experimental condition due to the competitive element present in the Tn-seq experiment.

Lack of S235 transposon mutants could be attributed to its antisense gene *sapB.* It has been previously reported that when *sapB* is mutated it affects formation of alkaline phosphatase in sporulation conditions, independently of SigF and E (Whalen, M. B. & Piggot, P. J., 1997), which could mean that it is the antisense disruption which causes the drop.

S163 is antisense to both *rsbRA* and *ndoA. rsbRA* is a member of the "stressosome" allowing the integration of several stress signals and eventually the activation of SigB (Reeves, A. *et al.*, 2010). *rsbRA* is typically constitutively expressed, but is down regulated during sporulation, despite it having a SigA promoter which is typically always expressed. *ndoA* encodes RNase EndoA (Pellegrini, O. *et al.*, 2005), to which S163 has been predicted to function in an antisense manner due to overlapping with the 3' region (Rasmussen, S. *et al.*, 2009). As such, S163 is not likely to be a good candidate for further study.

S1579 is antisense to sporulation-related genes *spoIIIJ* and *jag*. SpoIIIJ is essential for SigG activity, facilitating the "feeding tube" assembly from the mother cell to the forespore, allowing for intercellular signalling and the mother cell to direct SigG activity (Camp, A. H. & Losick, R., 2008). Jag is a SpoIIIJ associated protein with unknown function, but has had an antisense RNA shown which is presumably part of S1579 (Rasmussen, S. *et al.*, 2009; Irnov, I. *et al.*, 2010).

S1251 clusters in C41 which is why it has not been included in the previous analysis. Whilst S1251 is not an antisense RNA, the genes in the surrounding region are involved in purine utilisation and are not upregulated in sporulation. The gene *yuzJ* follows a very similar pattern of expression to S1251. Interestingly *yuzJ* has been shown to have reduced expression during germination under osmotic stress compared to typical germination in lab conditions (Nagler, K. *et al.*, 2016). *yuzJ* and S1251 do not appear to overlap and are transcribed from different strands and therefore could be independent RNA species.

3.4.2.3.2 Germination could involve sRNAs

Nagler, K. *et al.* (2016) studied the transcriptional landscape of endospores during outgrowth via RNAseq. Whilst a thorough analysis was performed on protein-coding genes, sRNAs were not analysed and as such this study aimed to change this. As such, this study identified S segments to be present and highly abundant. The sixth most abundant RNA species detected in a dormant endospore is S357, making S357 a likely

suitable candidate for further study. Other genes which could be putative sRNAs and are present in a dormant endospore in high amounts are S31 and S275.

S31 is antisense to three genes, of which have links to germination. *purR* is the transcriptional repressor of the *pur* operon (Weng, M. *et al.*, 1995). *yabJ* is required for activity of *purR*. Whilst not present in a dormant endospore in high abundance, both are rapidly expressed during the first 30 mins of sporulation. *ispE* is an essential gene, encoding a kinase within the MEP pathway of isoprenoid biosynthesis (biosynthesis of lipids). Again, small amounts are present in dormant endospore, but it is upregulated after germination initiation. S31 could therefore be working as an antisense RNA to one of these genes.

During ethanol stress S275 is the final gene of an operon controlled by SigB. However, during sporulation it is transcribed as an independent segment. S275 is transcriptionally active between hours 4 and 5 of sporulation. Expression profiles match perfectly to the operon containing *yfkH*, apart from during sporulation, anaerobic growth and fermentation. Therefore, S275 may have an independent sigma factor controlling its expression for these conditions.

3.4.3 Final summary

This study curated and compiled information on putative sRNAs in addition to using specialised software to characterise and select sRNAs. This study successfully identified many putative sRNAs that warrant further exploration. This study concludes that already published data is valuable source of information and has successfully identify putative sRNAs which could be interesting for future study, a prime example of which is S357. It is possible to now take those identified here and tailor experiments to further characterise and further confirm activity during sporulation specifically.

Table 12 – Summary table of most interesting putative sRNAs

Name	Classification	Sigma Nicolas	Sigma DBTBS	TF DBTBS	Antisense gene	Antisense overlap	Arnold terminator prediction	Rho terminator prediction	Score	Length	Conservation	log2 Ratio T5	log2 Ratio T24	top 5 S segment in dormant spore
S111	Indep-NT	SigEF	NA	MntR, DegU	NA	0	Yes	0	5	148	3rd	-1.77	0.67	
S1024	indep	SigK, SigA	SigX, SigW	NA	NA	0	Yes	0	4	216	4th			
S1227	Indep-NT	SigK, SigA	NA	ComK	NA	0	0	0	4	106	4th	-3	-1.42	
S2	Indep-NT	SigK, SigEF	NA	NA	NA	0	0	0	4	273	2nd			
S357	indep	SigGF	NA	NA	NA	0	Yes	0	4	296	3rd			1st
S423	indep	SigK, SigEF	NA	NA	NA	0	Yes	0	4	186	3rd	-4.32	-1.74	
S547	indep	SigGF	SigG, SigF	ComA	NA	0.76	Yes	0	4	123	2nd			
S645	indep	SigK	NA	SpoIIID, DegU, AhrC	NA	0	Yes	0	4	160	4th			
S849	indep	0	SigG	NA	NA	0	Yes	0	4	74	4th			

S968	indep	SigK, SigA, SigB	SigW	NA	NA	0	Yes	0	4	147	3rd			
S1009	indep	SigK, SigA, SigA	NA	NA	NA	0	0	0	3	151	3rd			
S1027	indep	SigA, SigA	NA	NA	NA	0	Yes	0	3	219	3rd	-2.58	-2.58	
S1202	Indep-NT	SigB	SigF, SigE, SigB	NA	*yugH*	0.91	Yes	Yes	3	984	3rd			
S145	Indep-NT	SigEF, SigB	NA	ComA	*ydzK*	0.36	Yes	0	3	215	4th			
S198	indep	SigA	NA	NA	NA	0	Yes	0	3	197	2nd	-1.04	-2.16	
S249	indep	SigK	NA	NA	NA	0	0	0	3	560	4th	-1.65	-1.54	
S275	indep	SigGF, SigA, SigB	NA	AbrB	NA	0	0	0	3	156	3rd			3rd
S612	indep	SigA	NA	DegU	NA	0	Yes	0	3	133	4th			
S665	indep	SigK	NA	NA	NA	0	0	0	3	109	3rd			
S731	indep	SigEF	NA	LexA	NA	0	0	0	3	498	3rd			
S863	indep	SigK, SigA, Sig-	NA	NA	NA	0	0	0	3	199	3rd			
S877	indep	SigEF	NA	GltR	NA	0	0	0	3	107	3rd			

S912	indep	SigWX Y, SigK	NA	NA	NA	0	0	0	3	170	4th			

4 Chapter Four: Validation and characterisation of putative sporulation specific sRNAs

4.1 Introduction

Understanding cellular organisation and interactions within networks is still a major interest in cell biology. Small RNAs (sRNAs) are part of the large regulatory network evolved to predict and cope with a constantly changing environment (Gottesman, S. & Storz, G., 2011), from sugar metabolism (Gorke, B. & Vogel, J., 2008) to quorum sensing (Kay, E. *et al.*, 2006).

Trans-encoded sRNAs are transcribed distant from their target and implement their effects by imperfect base pairing with their targets. Due to the distant relationship with their targets, *trans*-encoded sRNAs often have imperfect complementarity, but still can elicit their regulatory effects. Therefore, there is often no single obvious target but many potential targets of a *trans*-encoded sRNA.

Transcriptional profiling of *Bacillus subtilis* by Nicolas, P. *et al.* (2012) mapped the transcriptional landscape of the whole genome over 104 different conditions including a time course through the process of sporulation. Interestingly, this showed that 36% of the variability of gene expression could be attributed to sporulation using a Principle Component Analysis (PCA) projection when comparing all 269 transcriptomes. With such large variability attributed to endospore formation, the flexibility of gene expression is emphasised.

Previous studies have already identified the expression of several sRNAs during sporulation in *B. subtilis* (Silvaggi, J. M. *et al.*, 2006; Schmalisch, M. *et al.*, 2010). Although no specific regulatory role for these sRNAs has yet been identified, the fact that they are differentially expressed during sporulation suggests they may play a role in the process. As such, it is now important to systematically analyse the identified potential sporulation-specific sRNAs to determine if they do indeed play a role in the process of sporulation. Further to this, if the sRNAs do play a regulatory role in the process, how does their role fit within the large regulatory network needed to coordinate sporulation?

To answer these and other hypotheses regarding the potential role of sRNAs in regulation of sporulation in *B. subtilis*, sRNAs which were previously identified (Chapter Three) to have high potential to be sporulation specific were analysed using several methods. To first confirm their expression during sporulation, promoter fusions to GFP were utilised and their activity was determined. This also allowed for identification in which compartment the expression of the sRNA is present. To identify specific roles within the process of sporulation, a selection of the active sRNAs was chosen to be investigated in further detail. Assessments of each sRNAs role in growth was then determined, to ensure activity is specific to sporulation by utilisation of deletion mutants. Timing of endospore production is one of many crucial factors which rely on the many cascades of gene expression and as such was assessed using GFP promoter fusions to known sporulation genes in combination with sRNA deletions. Furthermore, deletion mutants were probed for differences in sporulation resistance properties and finally germination efficiency, crucial for resuming normal homeostasis.

4.2 Results

4.2.1 Promoter activity

To test expression activity of the different sRNAs during the endospore process, transcriptional promoter fusions to *gfp* were created using the single copy integration vector pBaSysBioII. pBaSysBioII has previously been successfully used to analyse gene expression in *B. subtilis* (Botella, E. *et al.*, 2010; Buescher, J. M. *et al.*, 2012). To ensure complete cloning of the sRNA gene, the upstream region and 20 bp after the potential start site of the sRNA were incorporated. The expression location within the developing endospore of the chosen sRNAs was monitored using fluorescence microscopy.

4.2.1.1 Identifying forespore or mother cell expression location

The strains carrying pBaSysBioII containing selected promoter regions from the putative sRNAs were forced to undergo sporulation in 2x Schaeffer's with glucose (SG) medium (Goldrick, S. & Setlow, P., 1983), a nutrient rich medium that facilitates endospore formation after successful growth, and images were taken during a late stage of sporulation. Compartment-specific expression was observed for eleven of the shortlisted sRNAs and one sRNA (S612) exhibited mixed compartment activity (Figure 26). Eight sRNAs displayed mother cell specific expression and four showed forespore specific expression (Figure 27). For S612, fluorescence was visible in both compartments in a small sub-population of the cells, but *gfp* activity was much stronger in the forespore.

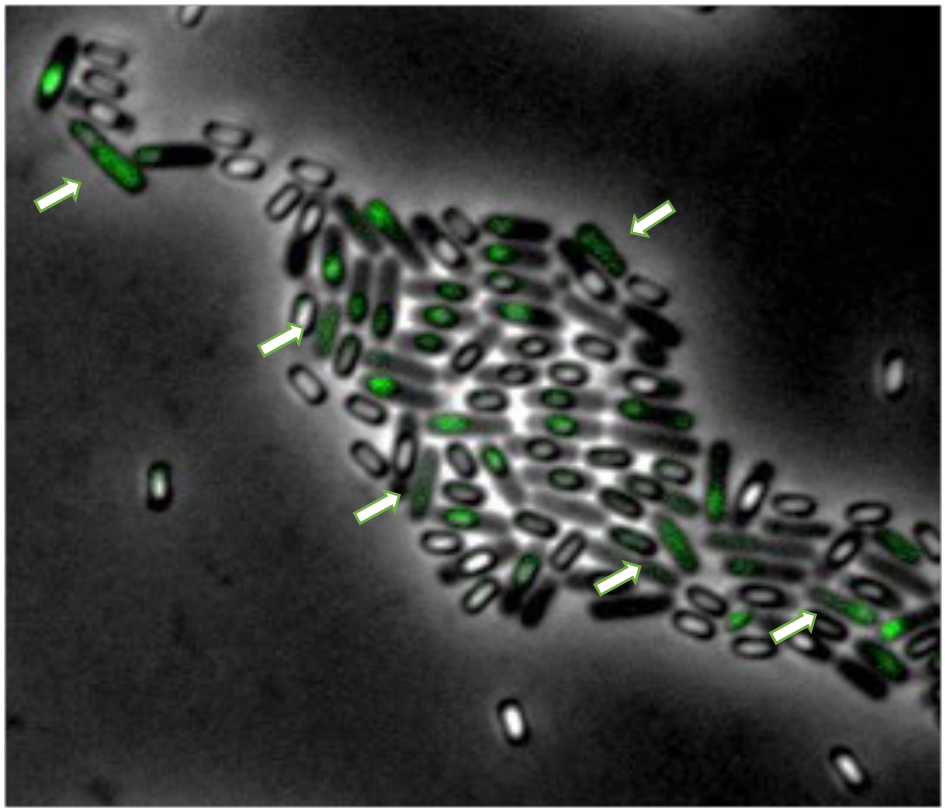


Figure 26 – Imaging of *B. subtilis* pS612-GFP during sporulation – overlay of phase contrast and FITC channels. Promoter activity of S612 is predominantly located in the forespore and weakly in the mother cell. Images were taken on a Leica DMi8 inverted fluorescence microscope at 100x magnification. Arrows indicate S612 in the mother cell.

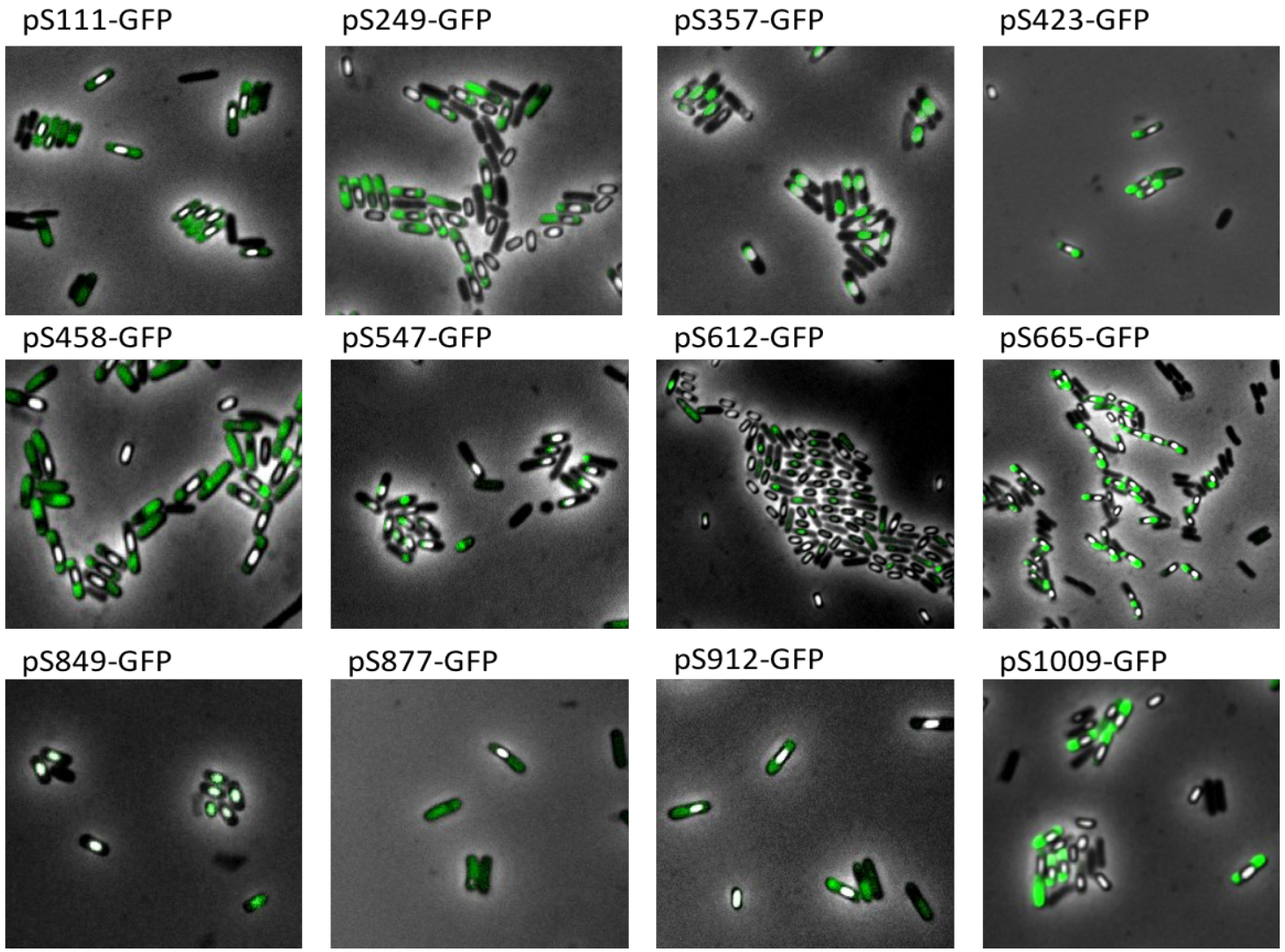


Figure 27 – Images of *B. subtilis* psRNA-GFP strains during sporulation – overlay of phase contrast and FITC channels. Promoter activity of sRNAs is displayed in green. Images were taken on a Leica DMi8 inverted fluorescence microscope at 100x magnification. Only sRNAs that exhibited GFP after active growth curve and therefore exhibting sporulation-specific activity shown.

Table 13 – Summary table of features identified from Chapter Three.

Name	Classification	Sigma Nic	Sigma DBTBS	TF DBTBS	Arnold termin	Rho termin	Score	Length
S357	indep	SigGF	NA	NA	Yes	0	4	296
S547	indep	SigGF	SigG, SigF	ComA	Yes	0	4	123
S612	indep	SigA	NA	DegU	Yes	0	3	133
S849	indep	0	SigG	NA	Yes	0	4	74

4.2.1.2 Timing of expression during sporulation

To determine the timing of expression for sRNAs shown to be expressed during the sporulation process, this study chose to further analyse a sub-group of forespore-specific, putative sRNAs – S357, S547, S612 and S849. These sRNAs were chosen based on both functional predictions outlined in Table 5 from chapter three and on microscopy observations of promoter fusion clones, described earlier in this chapter (Figure 27). Sporulation specific sigma factors have been predicted for three out of four of the selected sRNAs with all predicted to have Rho-independent terminators.

In order to ensure the temporally controlled induction of sporulation, a chemically defined resuspension medium was used in conjunction with GFP reporter fusions (Sterlini, J. M. & Mandelstam, J., 1969). Previous studies have identified three well-characterised genes as being expressed at different time points during the process of sporulation: *spoIIQ* is an early forespore-specific gene controlled by SigF, *spoIID* is an early mother cell gene activated by SigE and finally *gerE* is a late mother cell gene within the SigK regulon. A late acting forespore specific gene attached to *gfp* is unreliable, as GFP becomes less stable as the pH in the eventual endospore drops (Doherty, G. P. *et al.*, 2010). Therefore, this study has used these genes as a point of comparison for the temporal expression of the putative sRNAs examined in this study. As such, the promoter region of these genes was fused to *gfp* utilising pBaSysBioII. GFP expression was monitored during sporulation utilising a chemically defined medium.

All sRNA promoter activity was comparatively low when compared to the reporter promoter fusions. Therefore, for comparison purposes, expression values are normalised to the maximal GFP fluorescence levels for each of the promoter fusion constructs (Figure 28 and Figure 29). S357 and S547 increase in expression levels just after *spoIID* expression can be detected. S612 has a minor wave of activity present at a similar time to *spoIIQ* with a secondary increase in expression, contributing to more of its total resulting fluorescence, slightly later at the same time as *spoIID*. This suggests that S357, S547 and S612 are active during the early part of the sporulation process. A final increase in transcription is seen at the same time the expression of

gerE increases, for all sRNAs tested. This can potentially be attributed to the cross talk of sporulation-specific sigma factors where SigF and SigG have one base difference in their consensus sequence (Sierro, N. *et al.*, 2008). S849 follows a different pattern of expression, with a minor burst of transcription early in sporulation, but then a much larger gradual increase in transcription is detected in the later part of the sporulation process, after *gerE* is activated.

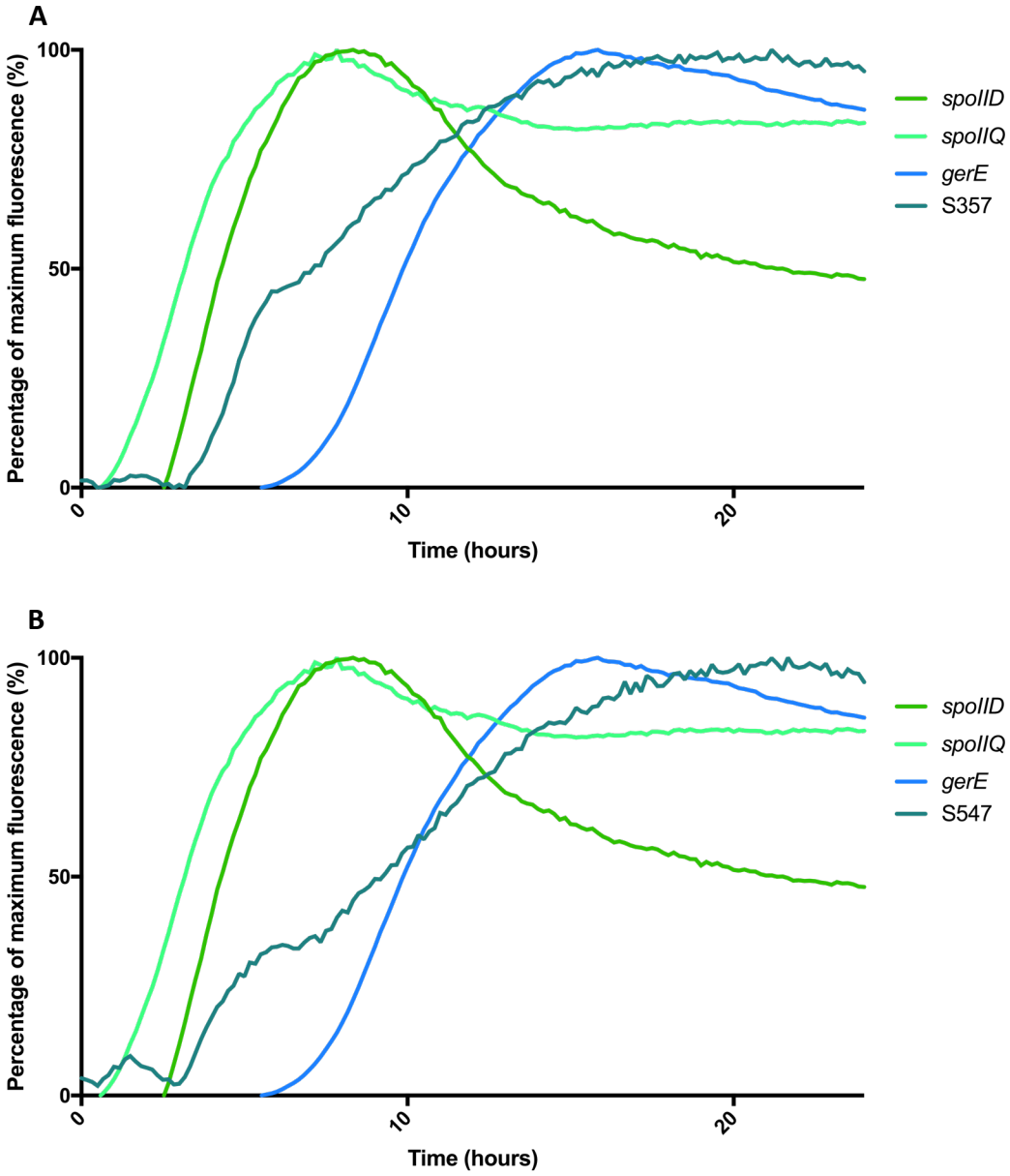


Figure 28 – Promoter activity of the putative sRNAs S357 (A) and S547 (B) in comparison with the activity of benchmark sporulation genes. All expression values are normalised to present the percentage of the maximal GFP level.

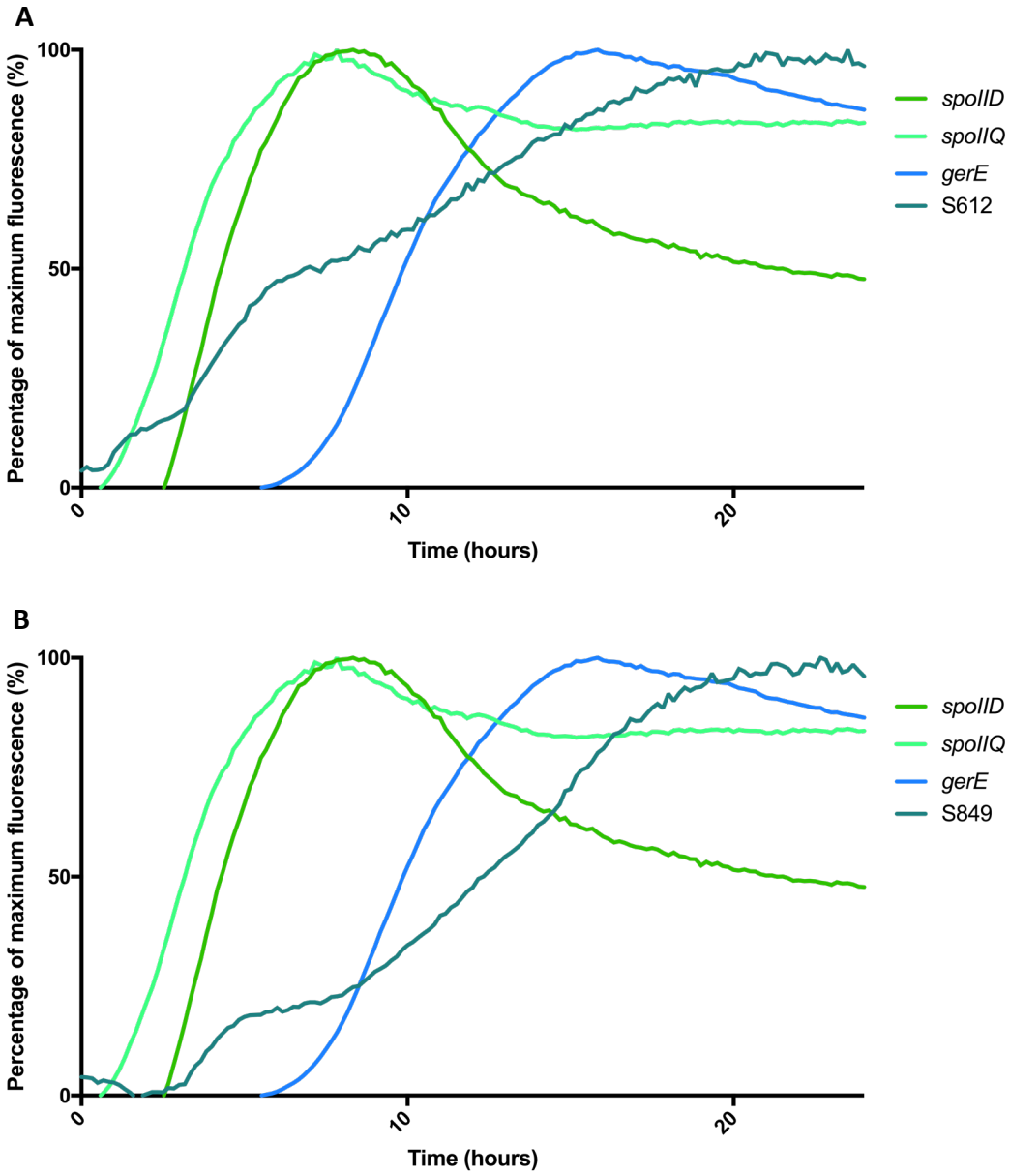


Figure 29 – Promoter activity of the putative sRNAs S612 (A) and S849 (B) in comparison with the activity of benchmark sporulation genes. All expression values are normalised to present the percentage of the maximal GFP level.

4.2.1.3 Towards determination of sigma factor specificity

Promoter specificity is driven by sigma factors and other transcriptional regulators (Paget, M. S., 2015). If the sigma factor controlling a gene is deleted, the expression and activity of its regulon is also removed. The *B. subtilis* genome encodes four sigma factors involved in regulating the sporulation process (Fimlaid, K. A. & Shen, A., 2015). To understand the sigma factors involved in regulating the expression of the four selected sRNAs, promoter fusions of the sRNAs were combined with deletion strains of the sigma factors previously shown to be involved in sporulation.

Measuring fluorescence and optical density (OD) during sporulation can be technically challenging due to numerous factors. Besides normal auto-fluorescence of a culture, the endospore under production becomes phase bright as the layers develop and therefore optically refractive during the late stages of sporulation. In addition, GFP becomes less stable as the pH in the eventual endospore drops (Doherty, G. P. *et al.*, 2010). It should also be noted that of the sigma factor deletion mutants it is only the *ΔsigK* strain that can form a phase-grey forespore, of which gene expression is not blocked until stage IV, with the rest unable to reach that point of sporulation (Stragier, P. & Losick, R., 1990).

Figure 30 shows representation of one individual experiment, performed on three replicates of each strain. This experiment was repeated 3 times independently. Figures 30A-5C show the changes in the expression levels of the genes used for benchmarking in this study in response to deletion of the different genes encoding the sigma factors. As expected, deletion of *sigF* abolished *spoIIQ* activity and deletion of *sigE* removed the activity of *spoIID.* Deletions of all relevant sigma factors prevented activity of *gerE,* as this is a late active gene, which would require the early sigma factors to initiate the cascades required for its expression. Only minimal activity can be seen for the *sigG* mutant; however, the transcription level is lower than the levels observed in the wild-type P*gerE-gfp* strain.

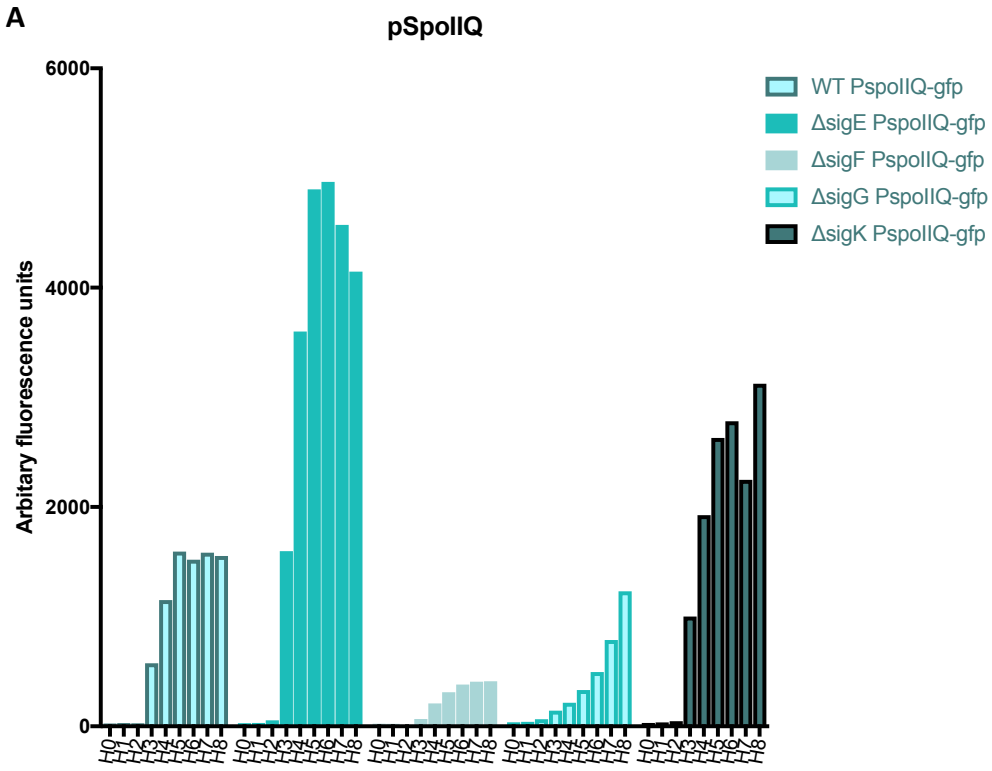
During sporulation, *sigG* is activated first in the forespore; its activation then initiates a signalling pathway via SpoIVFB through to the mother cell to assist activation of

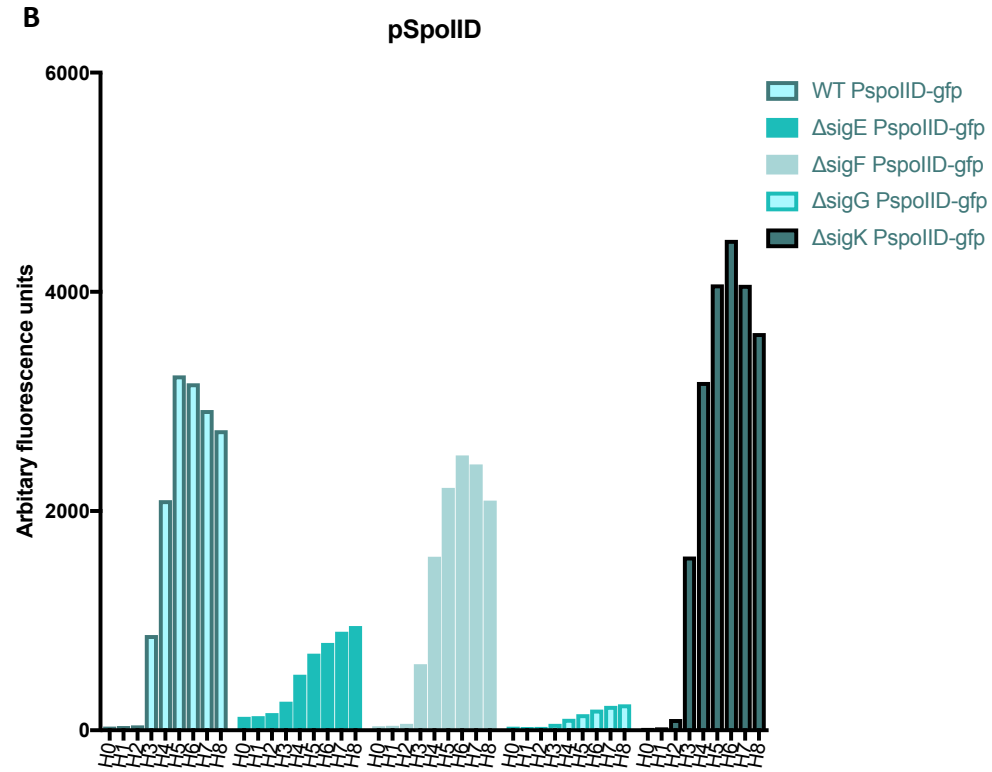
sigK. In the absence of *sigG*, *sigK* activation is incomplete and as such, there is no increase in the transcription levels of the SigK regulon. Interestingly, one observable phenotype of the *sigE* mutant is the creation of a double forespore (Eichenberger, P. *et al.*, 2001), which can explain the increased expression of *spoIIQ* in the *sigE* deletion strain as there are now two areas in the cell in which expression is induced.

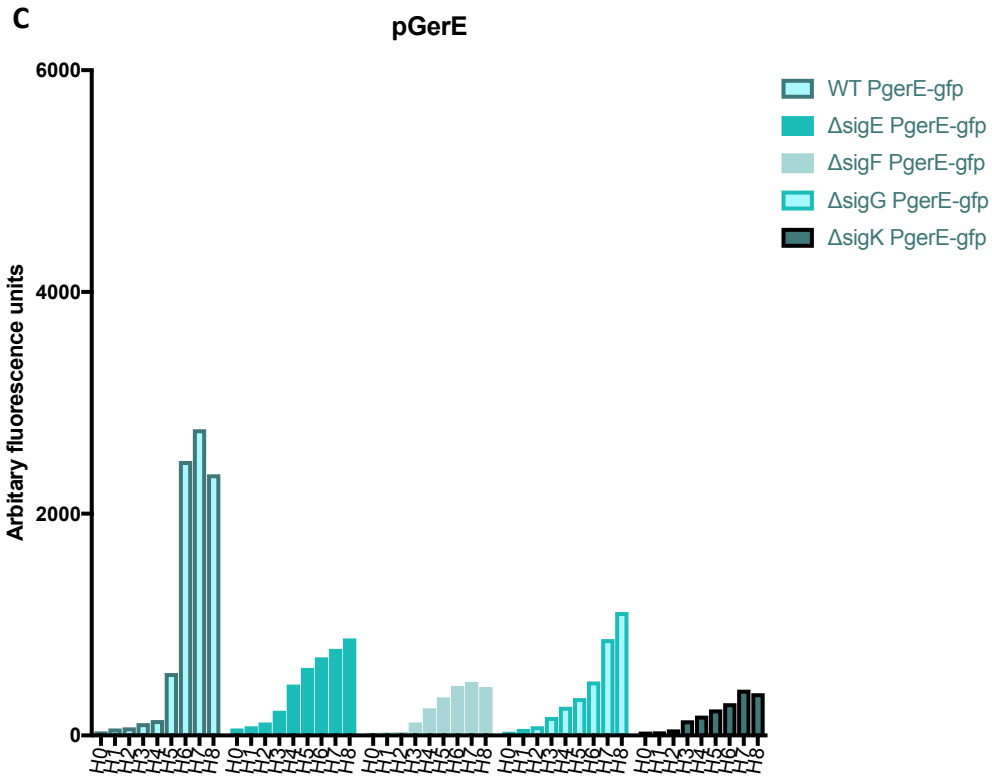
In the wild-type background S357 is shown to be activated after hour 3, and expression steadily increases as sporulation progresses (Figure 30D). The same pattern of activity can be seen for each of the sigma factor deletions, curiously with Δ*sigF* having the least impact and Δ*sigK* having the most. Both *sigE* and *sigK* are mother cell specific sigma factors, however, both require the cooperation of their forespore-specific counterparts for activity.

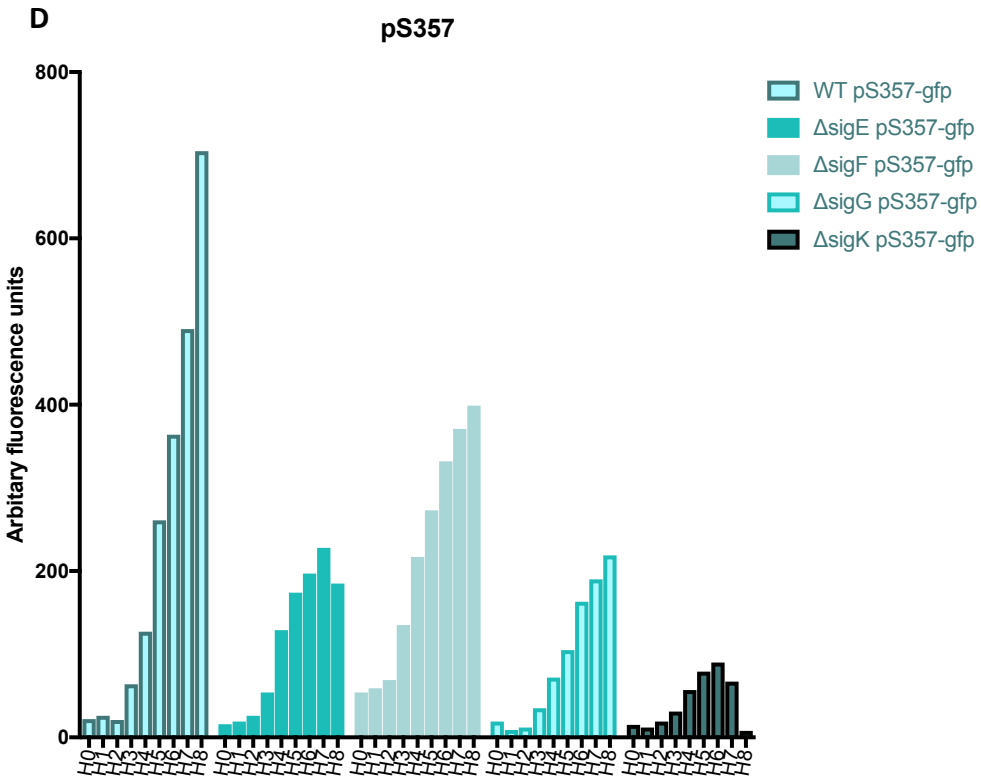
S547 is also activated during the 3rd hour of sporulation with levels increasing as sporulation progresses (Figure 30E). Deletion of *sigE* increases the rate of expression early on, with expression levels at hour 3 and 4 at least double that of the wild type, presumably explained by the double forespore. Deletion of the later acting sigma factors *sigG* and *sigK* had no effect on activity, as did deletion of *sigF*.

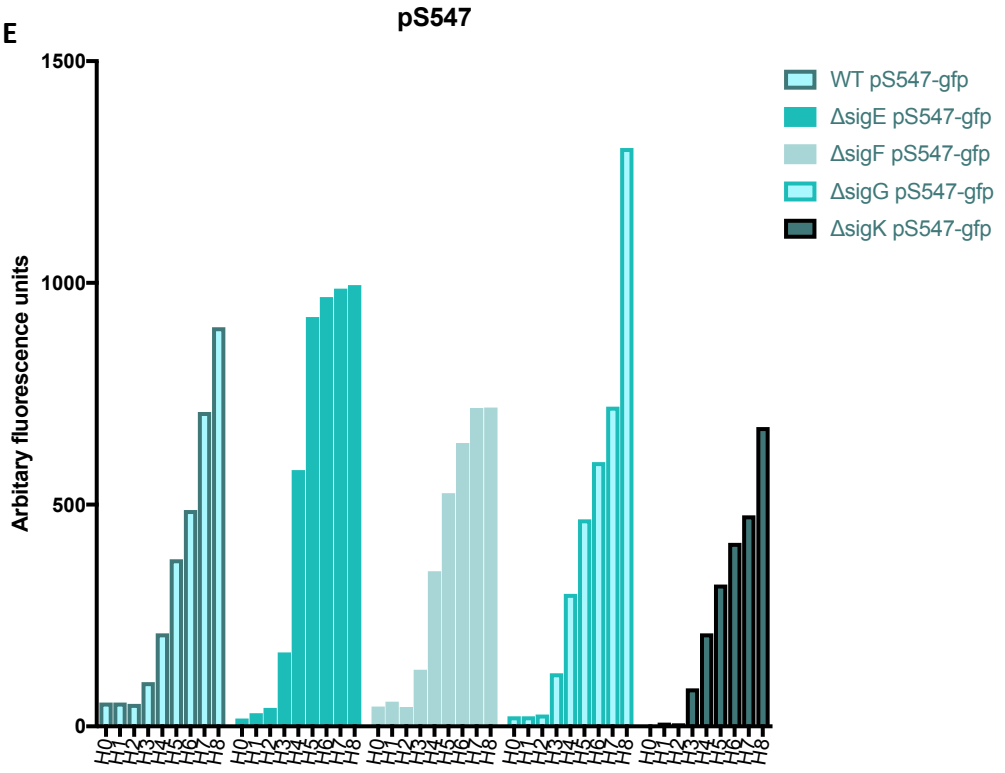
The other two sRNAs, S612 and S849, displayed similarly puzzling patterns of promoter activity. Expression of S612 seemed to start at hour 2 onwards. Deletion of *sigE* is the only change which seemed to not have any effect on the transcription pattern of S612. Deletion mutants in the other sigma factors, however, reduced expression dramatically. S849, in contrast, as seen in Figure 30 is expressed in the latter part of the sporulation process, from hour 4 onwards. Deletion of sigma factor mutants seemed to increase activity.

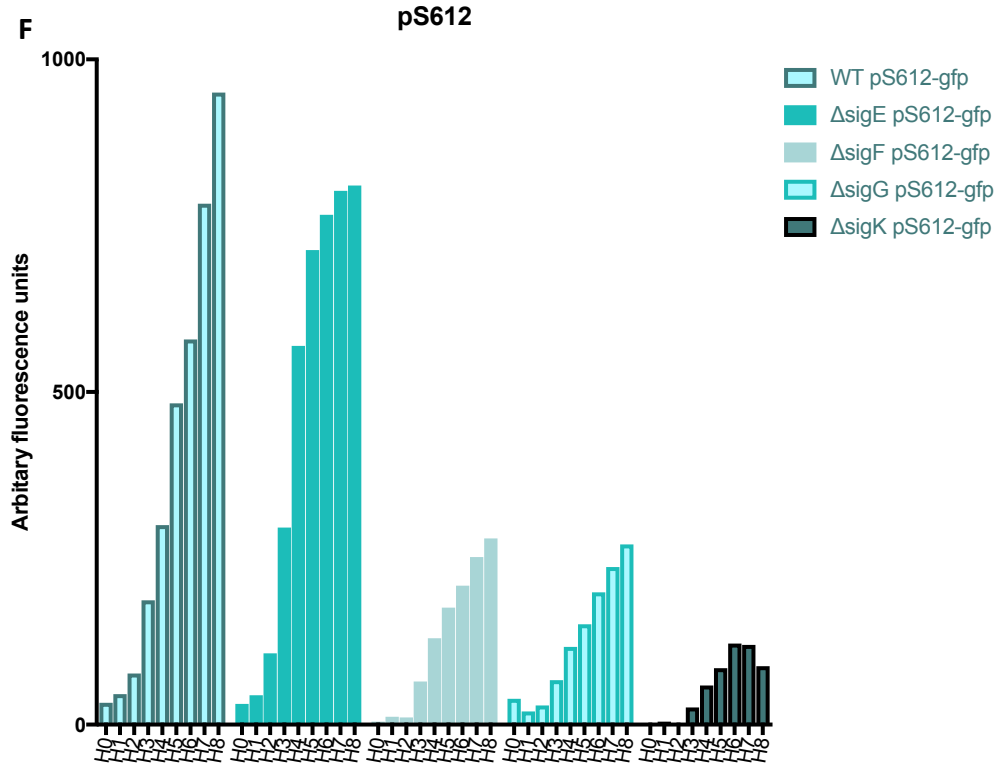

A
pSpoIIQ
Arbitary fluorescence units
6000
4000
2000
0
WT PspoIIQ-gfp
ΔsigE PspoIIQ-gfp
ΔsigF PspoIIQ-gfp
ΔsigG PspoIIQ-gfp
ΔsigK PspoIIQ-gfp
H0 H1 H2 H3 H4 H5 H6 H7 H8 H0 H1 H2 H3 H4 H5 H6 H7 H8 H0 H1 H2 H3 H4 H5 H6 H7 H8 H0 H1 H2 H3 H4 H5 H6 H7 H8 H0 H1 H2 H3 H4 H5 H6 H7 H8


B
pSpoIID
Arbitary fluorescence units
6000
4000
2000
0
WT PspoIID-gfp
ΔsigE PspoIID-gfp
ΔsigF PspoIID-gfp
ΔsigG PspoIID-gfp
ΔsigK PspoIID-gfp
H0 H1 H2 H3 H4 H5 H6 H7 H8 H0 H1 H2 H3 H4 H5 H6 H7 H8 H0 H1 H2 H3 H4 H5 H6 H7 H8 H0 H1 H2 H3 H4 H5 H6 H7 H8 H0 H1 H2 H3 H4 H5 H6 H7 H8


C
pGerE
Arbitary fluorescence units
6000
4000
2000
0
WT PgerE-gfp
ΔsigE PgerE-gfp
ΔsigF PgerE-gfp
ΔsigG PgerE-gfp
ΔsigK PgerE-gfp
H0 H1 H2 H3 H4 H5 H6 H7 H8


D
pS357
Arbitary fluorescence units
800
600
400
200
0
WT pS357-gfp
ΔsigE pS357-gfp
ΔsigF pS357-gfp
ΔsigG pS357-gfp
ΔsigK pS357-gfp
H0 H1 H2 H3 H4 H5 H6 H7 H8


E
pS547
Arbitary fluorescence units
1500
1000
500
0
WT pS547-gfp
ΔsigE pS547-gfp
ΔsigF pS547-gfp
ΔsigG pS547-gfp
ΔsigK pS547-gfp
H0 H1 H2 H3 H4 H5 H6 H7 H8


F
pS612
Arbitary fluorescence units
1000
500
0
WT pS612-gfp
ΔsigE pS612-gfp
ΔsigF pS612-gfp
ΔsigG pS612-gfp
ΔsigK pS612-gfp
H0 H1 H2 H3 H4 H5 H6 H7 H8

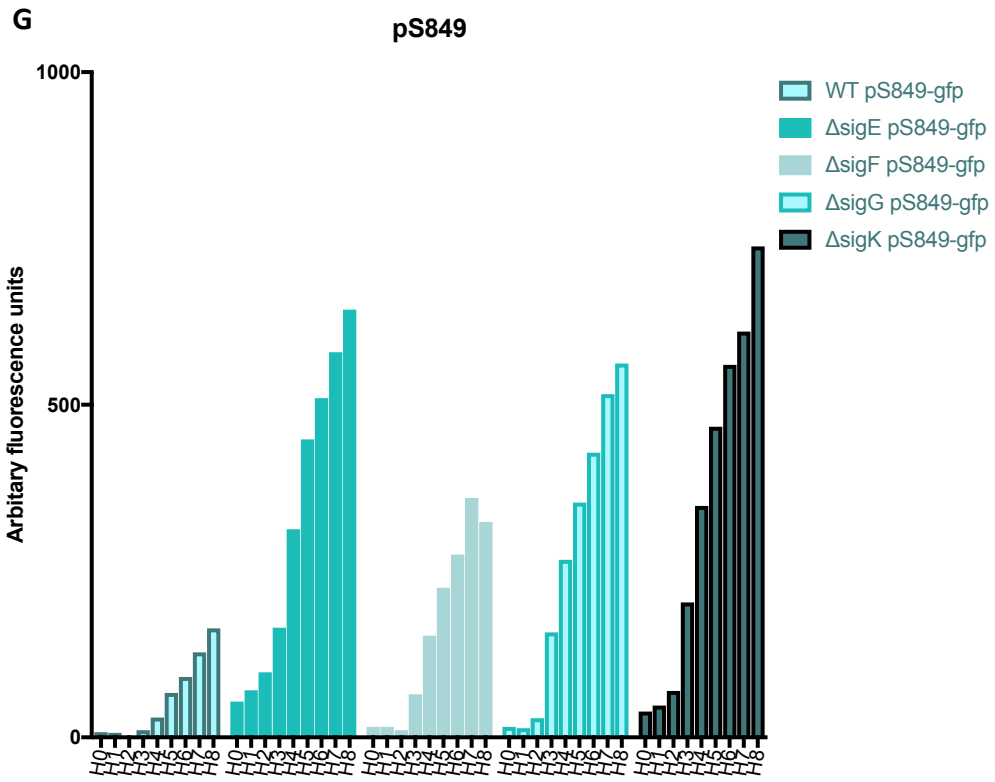


Figure 30 – Promoter activity of benchmarking sporulation genes, *spoIIQ* (A), *spoIID* (B) and *gerE* (C) expressed in different sigma factor deletion mutant strains. Promoter activity of the sRNAs S357 (D), S547 (E), S612 (F) and S849 (G) in the presence of sigma factor deletion mutant strains. Values showed per hour of sporulation, from hour 0 through to hour 8, via the resuspension method of endospore formation.

4.2.2 Assessing sRNA deletion mutant's growth capabilities

Classical molecular biology works on a bottom-up approach, assigning functions based on phenotypes. A widely accepted method is to determine what occurs when a gene of interest is deleted. As such, this study created deletion mutant strains of *B. subtilis* in which S357, S547, S612 and S849 were replaced by a phleomycin antibiotic resistance cassette. The phenotypes of these strains under different experimental conditions were investigated.

Approximately 1500 nucleotides both upstream and downstream of the region to be deleted was amplified. This was then joined via overlapping PCR to a phleomycin antibiotic resistance cassette. This linear construct was then transformed into *B. subtilis* utilising homologous integration, where selection for antibiotic resistance yields a strain that has replaced the native locus with the resistance cassette (Fabret, C. *et al.*, 2002).

4.2.2.1 Growth in rich medium

This study firstly examined the ability of each sRNA deletion mutant to grow in standard laboratory media (LB) to ensure that any phenotype this study identified was due to the role the sRNA played during sporulation and not during growth. (Figure 31).

Three of the deletion mutants tested were characterised by displaying similar growth curves to the wild-type. The one exception was the deletion mutant of S849, which displayed a longer lag time (Figure 31D). This phenotype disqualifies the S849 deletion mutant from future experiments, since sporulation experiments often require growth prior to the start of the sporulation phase and therefore was not pursued further.

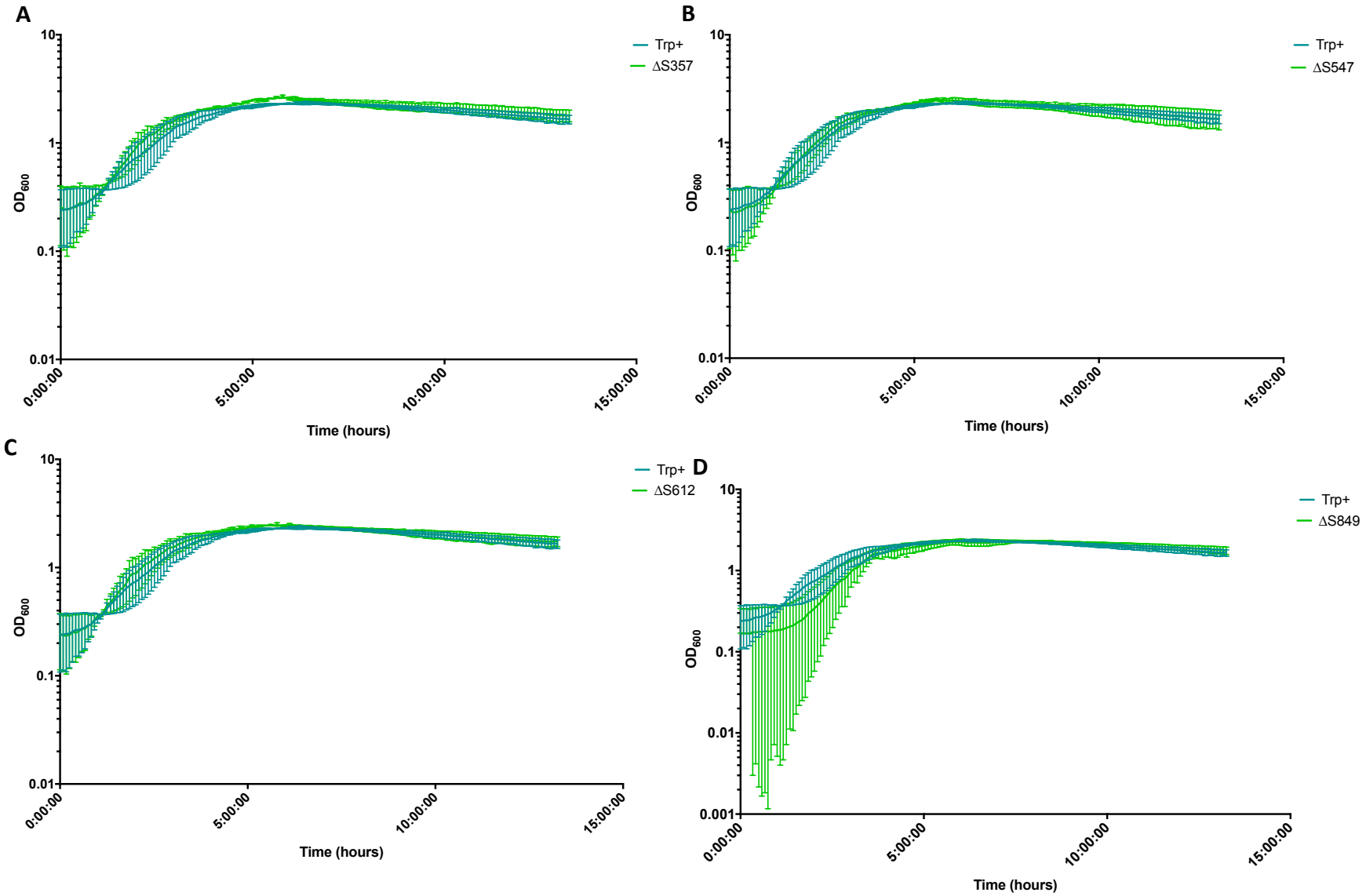


Figure 31 – Growth of sRNA deletion strains ΔS357 (A), ΔS547 (B), ΔS612 (C) and ΔS849 in LB. OD $_{600\ nm}$ readings was taken every 10 minutes. Average of two independent experiments with three biological replicates. Error bars represent standard error mean.

4.2.2.2 Growth in defined media

Deletion strains of sRNAs were checked for growth in minimal media, where the carbon source available for growth is defined. M9 is a minimal media routinely used for the growth of *B. subtilis* where it is advantageous to control the nutrients available. In this medium, *B. subtilis* follows a pattern of lag, exponential growth, stationary phase and subsequently cell death, where in complex medium such as LB little or no cell lysis occurs.

Growth of the deletion mutants and the wild-type was compared under three different carbon sources – glucose (Figure 34), fructose (Figure 35), and pyruvate (Figure 36). While using glucose as a sole carbon source produced no differences in growth curves between the sRNA deletion mutants and the wild-type strain, there were pronounced differences when growing on other carbon sources. Growth in M9 with fructose as its sole carbon source resulted in a different growth behaviour of ΔS357 compared to the wild-type (Figure 35A).

Growth in M9 supplemented with pyruvate was also tested due to the observation from expression profiles for growth in minimal medium from Nicolas, P. *et al.* (2012) that the sporulation specific sigma factors are upregulated in M9 where pyruvate or gluconate and succinate are the sole carbon source. Other sigma factors not involved in sporulation do not exhibit the same upregulation (Figure 32). Interestingly, many sRNAs selected initially in chapter three are also upregulated in the same pattern (Figure 33). Figure 36 however shows there is no growth defect, even though upregulation of the sRNAs in this growth condition occurs.

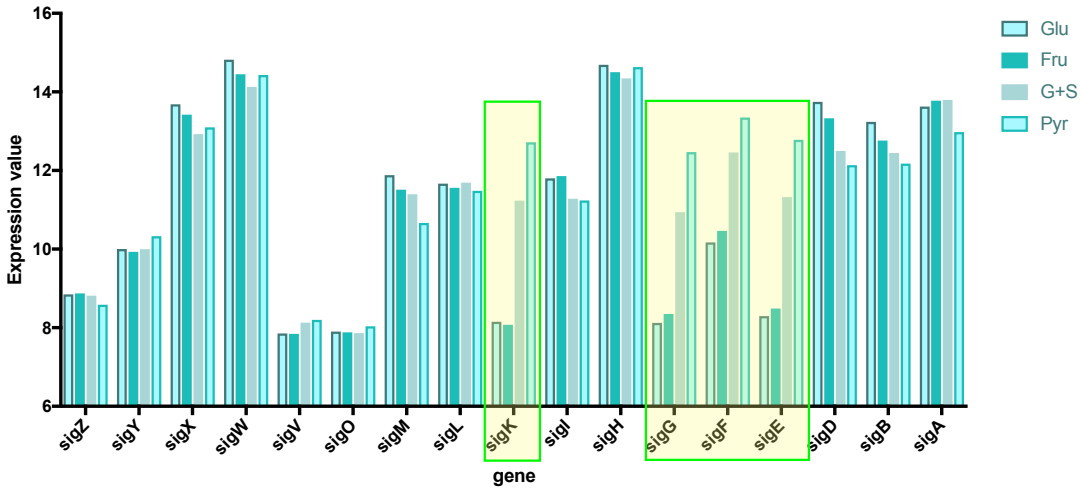


Figure 32 – Expression levels of *B. subtilis* sigma factors in M9 medium supplemented with different carbon sources. Glu - glucose, Fru - fructose, G+S - mixture of gluconate and succinate and Pyr – pyruvate.

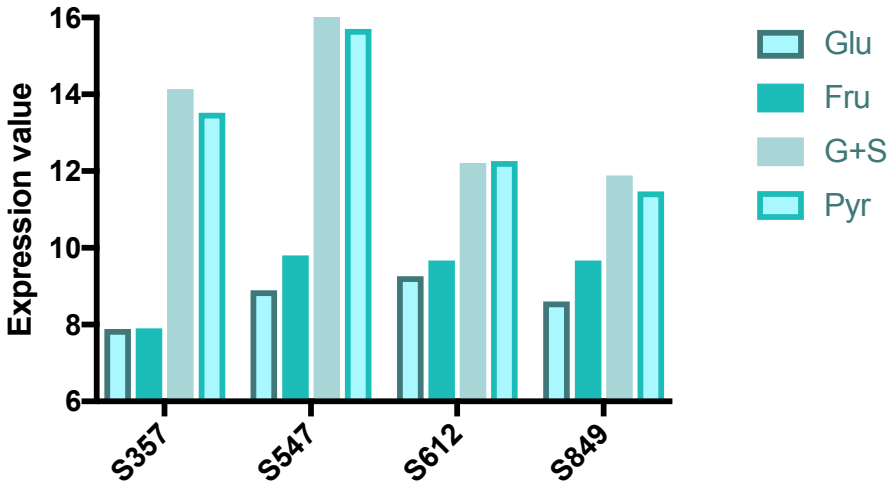


Figure 33 – Expression pattern of sRNAs in M9 medium extracted from data available by Nicolas *et al*. Glu - glucose, Fru - fructose, G+S -mixture of gluconate and succinate and Pyr - pyruvate

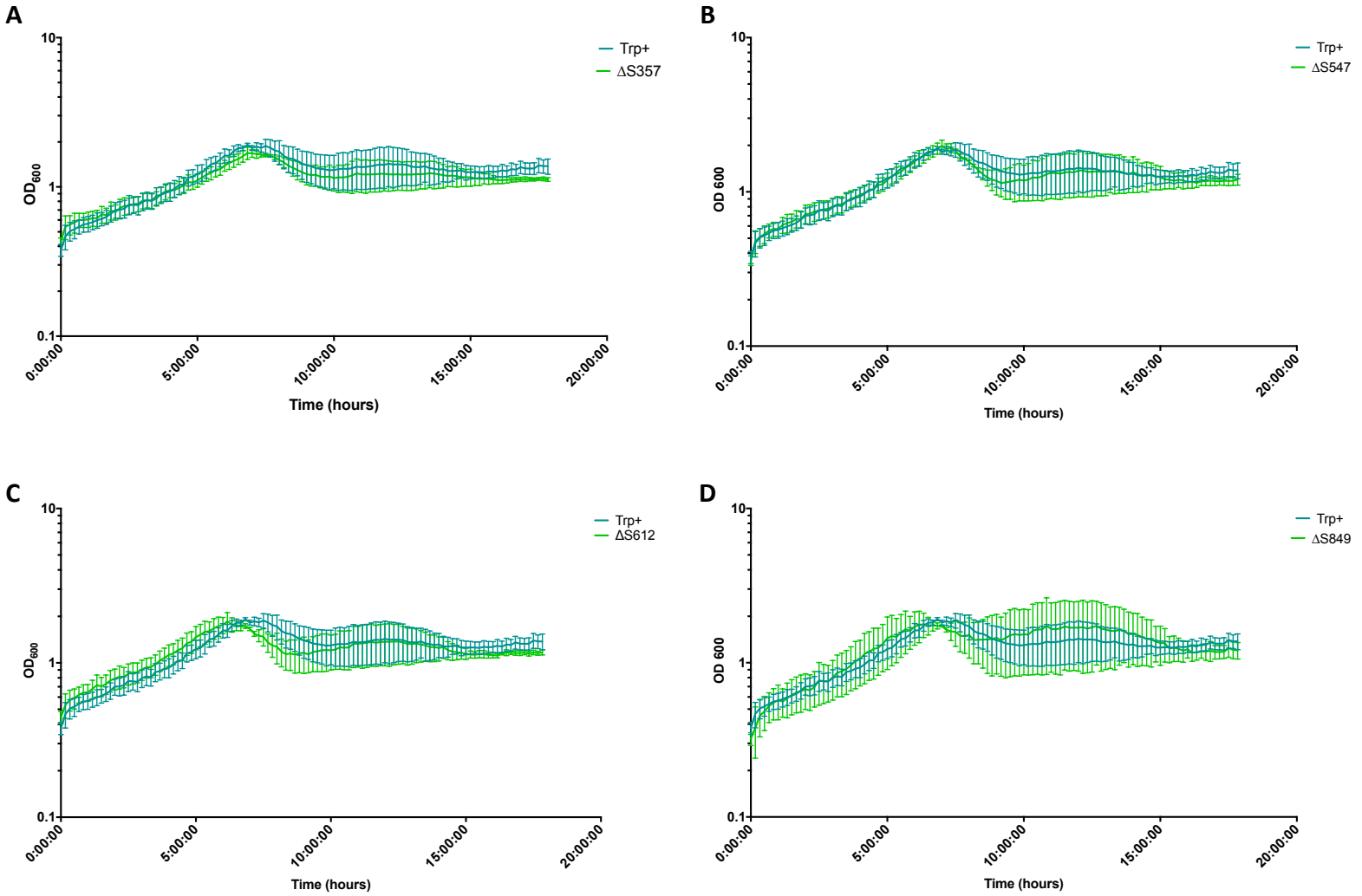


Figure 34 – Growth of sRNA deletion mutant strains ΔS357 (A), ΔS547 (B), ΔS612 (C) and ΔS849 in M9 glucose. OD $_{600\ nm}$ readings were taken every 10 minutes. Average of two independent experiments with three biological replicates.

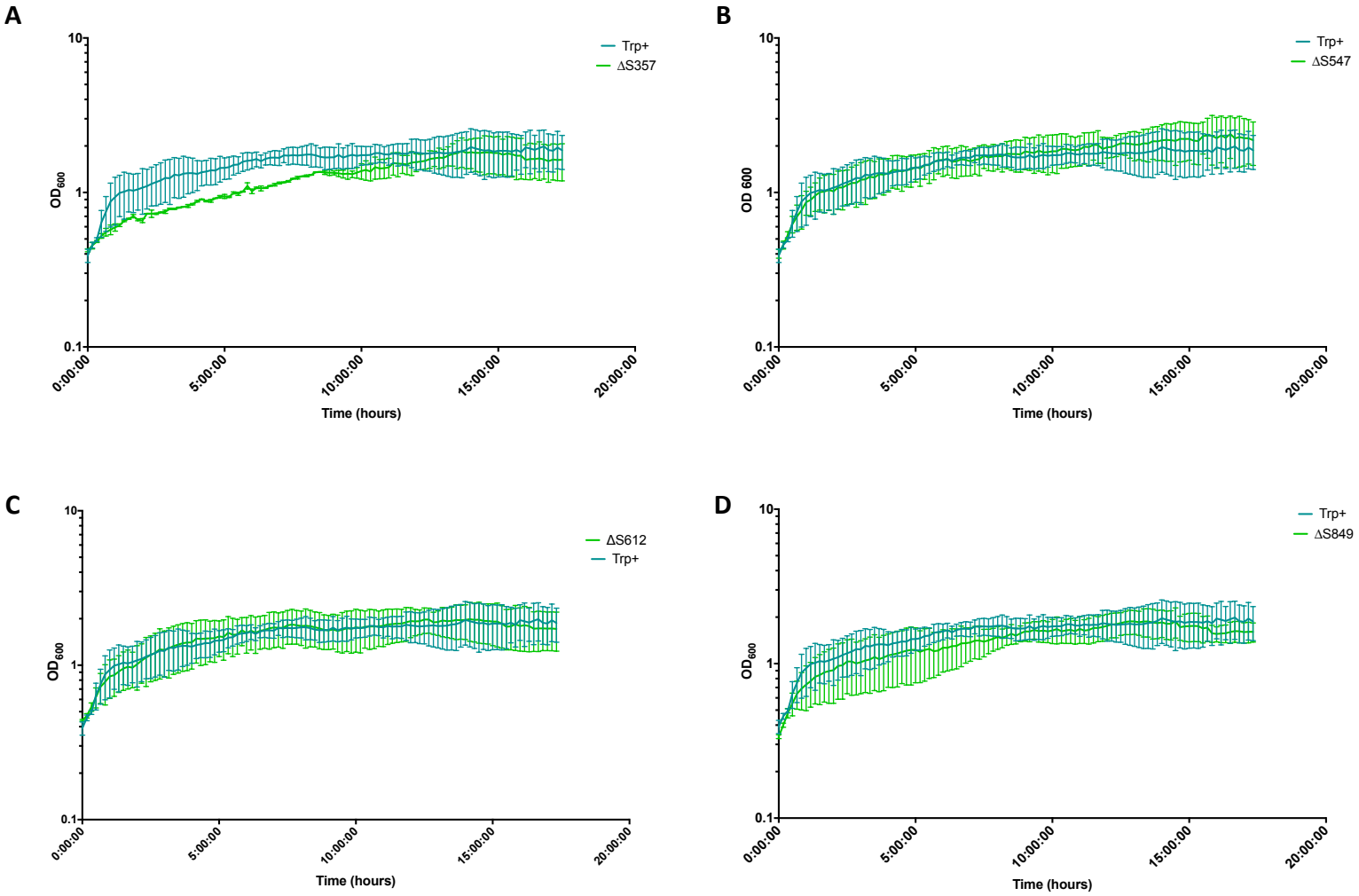


Figure 35 – Growth of sRNA deletion mutant strains ΔS357 (A), ΔS547 (B), ΔS612 (C) and ΔS849 in M9 fructose. OD $_{600\ nm}$ readings were taken every 10 minutes. Average of two independent experiments with three biological replicates.

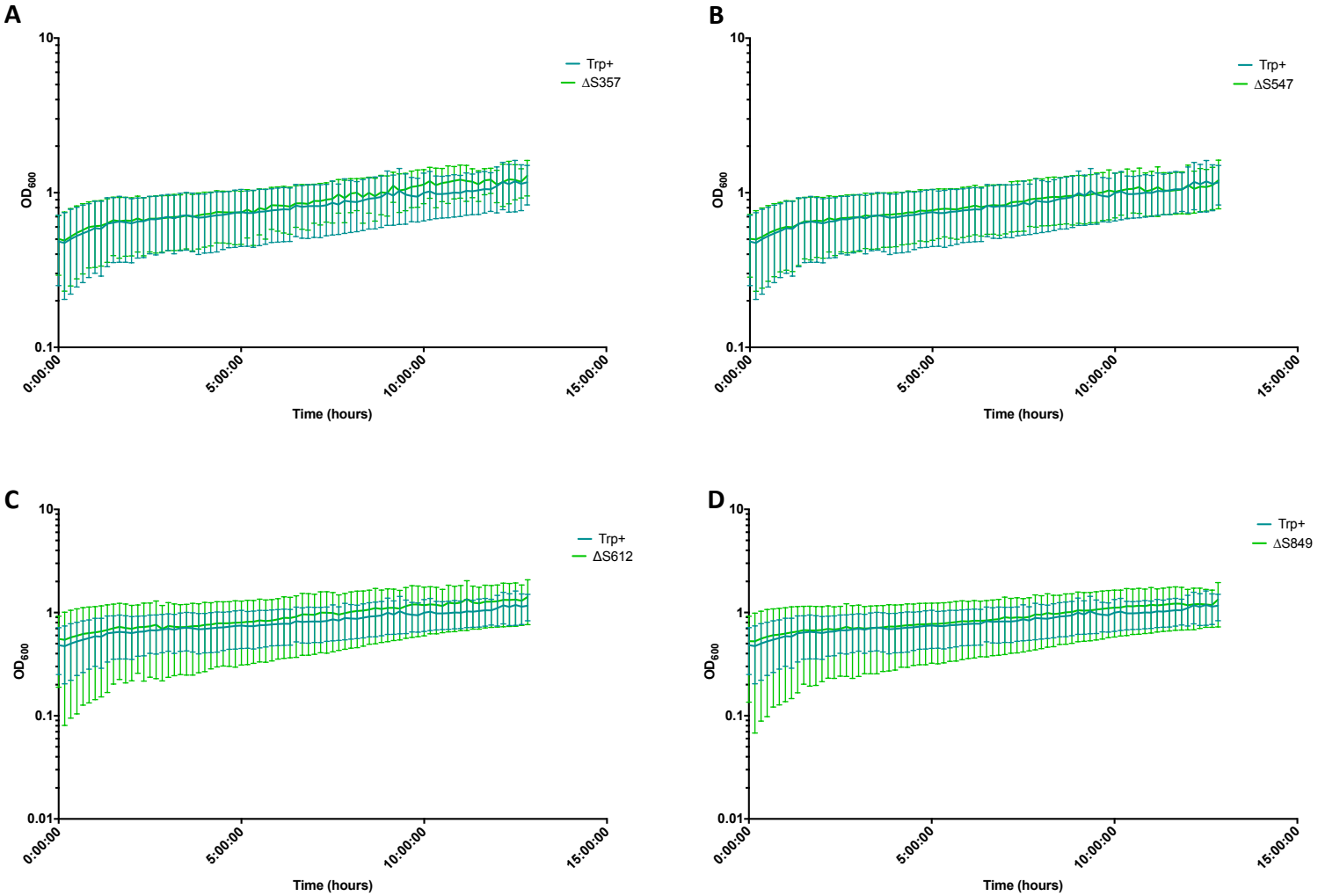


Figure 36 – Growth of sRNA deletion mutant strains ΔS357 (A), ΔS547 (B), ΔS612 (C) and ΔS849 in M9 pyruvate. OD $_{600\,nm}$ readings were taken every 10 minutes. Average of two independent experiments with three biological replicates.

4.2.3 Sporulation

4.2.3.1 Sporulation efficiency

The efficiency of sporulation was then checked for its ability to produce endospores. The natural endospore resistance properties are typically used as a method for distinguishing between developed endospores and vegetative cells. Commonly heat resistance is used, where endospores can survive prolonged wet heat and those that are not resistant to heat are killed, typically vegetative cells and those not having completed sporulation. As such, by plating pre- and post-heat treatment the ratio of endospores to the entire population of cells can be distinguished. The wild-type and deletion mutants were inoculated into DSM medium and followed the natural depletion method of endospore production. After 24 hours, a sample was taken from the population and split into two. One sample was plated without treatment and the other was heat shocked, killing the remaining vegetative cells. From this, the percentage of endospores present and capable of completing the entire sporulation process were calculated in triplicate. This method includes the entire lifecycle of sporulation; from a vegetative cell, through sporulation to successfully produce an endospore and finally complete germination.

Experiments were carried out with three technical and three biological replicates. Figure 37 shows that endospore formation of the wild-type was within the previously reported range (Maughan, H. & Nicholson, W. L., 2004). None of the deletion strains for the sRNAs (S357, S547, S612 and S849) showed a statistically significant decrease in sporulation efficiency.

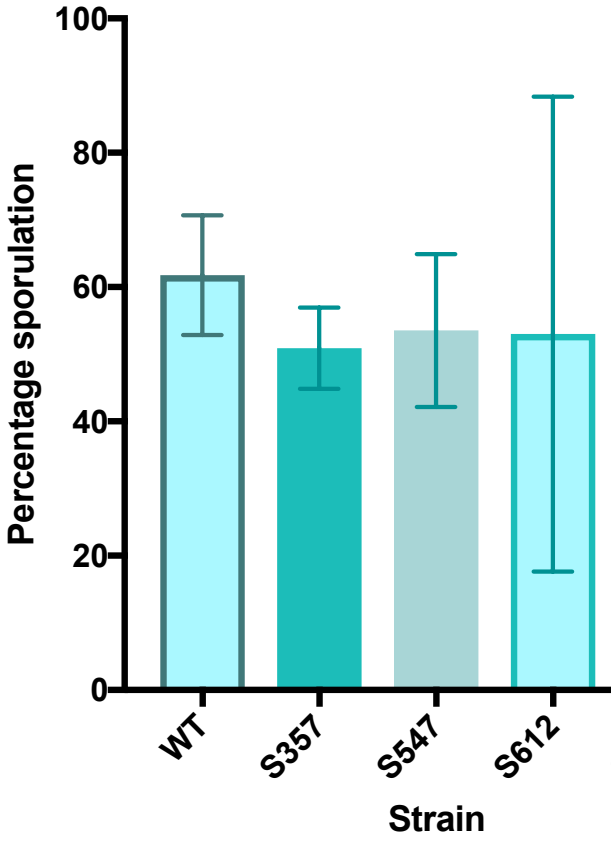


Figure 37 – *B. subtilis* sRNA deletion mutants sporulate as efficiently as the wild-type. Percentage of the population of *B. subtilis* which successfully completed the sporulation cycle when subjected to wet heat and thus removing vegetative cells. Counts of colonies formed after heat treatment were normalised to colony counts before heat treatment. Error bars indicate standard error of three experiments.

4.2.3.2 Testing mutant endospore resistance properties

Whilst heat treatment is successfully used to remove vegetative cells and endospores that are not heat resistant, endospores possess many other resistance properties. This, in turn, can be used to evaluate the specific deficiencies of different endospores. Endospores are not solely resistant to wet heat, but capable of surviving many different chemical assaults. Purified endospores were therefore tested with alternative treatments, to ensure the effect measured from the starting population was direct to endospore resistance and not the killing of non-endospores. Ethanol, chloroform and hydrogen peroxide are all chemicals which affect the endospore in different manners. For example, hydrogen peroxide is an oxidising agent, presumably killing endospores by damaging external layers (Melly, E. *et al.*, 2002; Young, S. B. & Setlow, P., 2003). Chloroform resistance has previously been shown to be effected by deletion of *stoA*, encoding a thiol-disulphide oxidoreductase resulting in endospores that are deficient in the cortex (Crow, A. *et al.*, 2009) but it is unknown how this is involved in chloroform resistance. Endospores were treated with sub-lethal concentrations of each treatment and instead of counting total numbers of endospores, dilution spot plates of endospores were plated (Thomas, P. *et al.*, 2015).

All treatments reduced the total population of endospores by at least one dilution factor (Figure 38), as a certain percentage of endospores in the population are always sensitive. Ethanol and hydrogen peroxide treatment revealed the difference between the number of resistant endospores in wild-type and mutants is insignificant. In the case of treatment with chloroform, however, a deletion of S612 appears to increase survival, relative to the wild-type. Chloroform addition reduced the wild-type by a mean of 6 dilution factors but S612 by only 3 dilution factors.

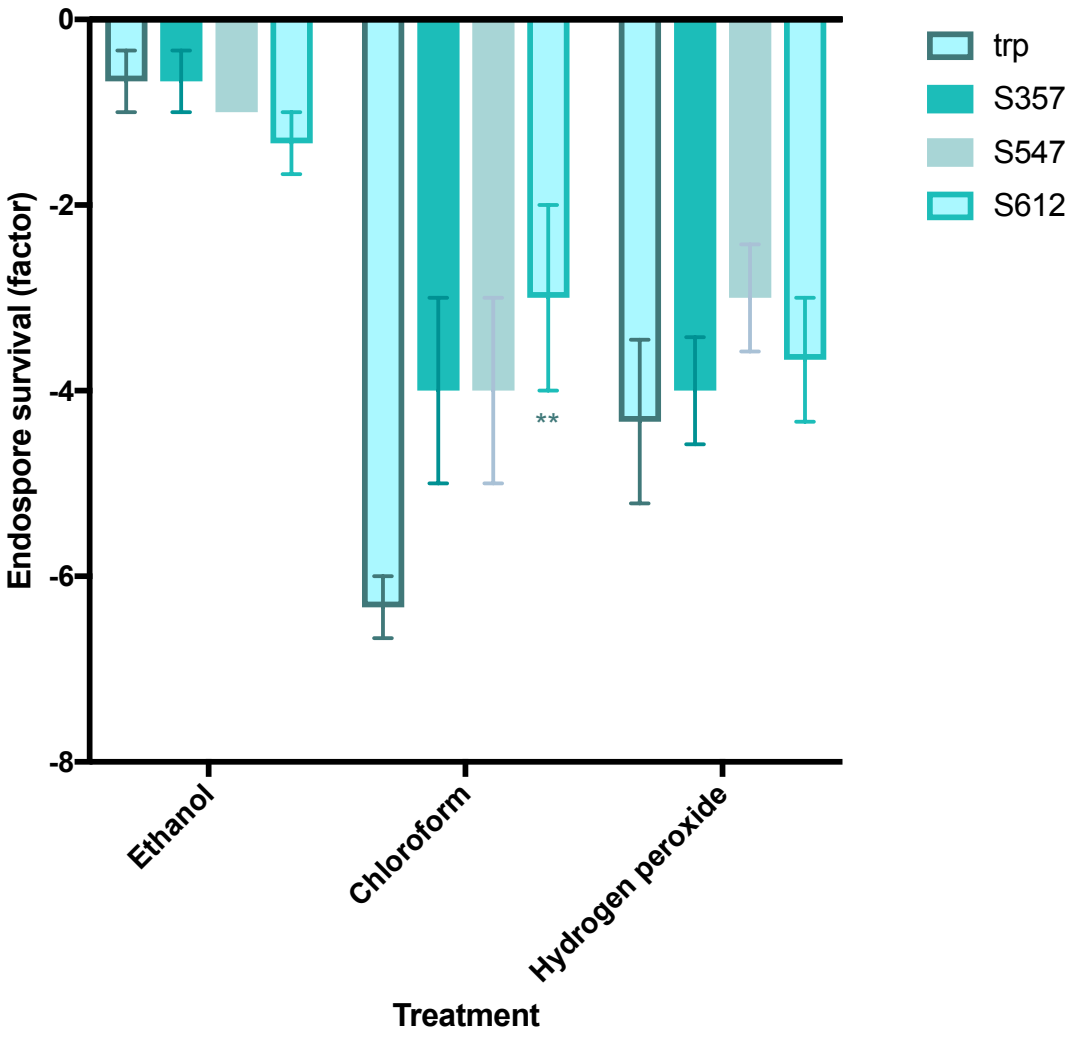


Figure 38 – Survival of endospores to chemical treatments. Treatments were administrated and normalised to no treatment. Endospore survival was measured as a 10x dilution series. Three experiments were averaged, error bars represent standard error mean. ** = p-value - <0.05.

4.2.3.3 Assessing timing of sporulation

General efficiency experiments utilising heat shock and subsequent growth allow time to compensate for minor deficiencies in regulatory processes, this study questioned whether there was any developmental delay during sporulation. As such, this study used previously created promoter fusions to *gfp* of three known sporulation genes from earlier in this study, *spoIIQ, spoIID* and *gerE,* and combined these with stains containing deletion of the sRNAs. If the strain had a developmental delay the promoters would be activated at erroneous times or periods compared to the wild-type. The strains were created using the promoter fusion as the background

strain and the marked deletion of the sRNA was added to this strain. This was to minimise the possibility of a different copy number of the GFP reporter constructs being present and therefore should have a similar level of activity during resuspension in a chemically defined sporulation medium.

The first sporulation gene to be checked was *spoIIQ,* an early forespore-specific gene (Figure 39A). The period of expression for *spoIIQ* was comparable to the wild type for stains harbouring deletions of ΔS357 or ΔS612. However, both had increased levels of expression. Interestingly, ΔS547 had more than double the *spoIIQ* activity in addition to an extended period of activity.

Secondly, the SigE regulated *spoIID* gene was used as a reporter for early mother cell gene expression (Figure 39B). The wild-type level and timing of expression for *spoIID* and ΔS547 was identical. ΔS612, whilst having a similar period of activity, exhibited slightly increased levels of expression of *spoIID*. ΔS357 displayed an earlier drop in promoter activity of *spoIID* as well as reduced levels of expression.

Finally, the last benchmarking gene to be activated is *gerE,* a late mother cell gene regulated by SigK (Figure 39C). The wild-type strain initiated activity at the same time as the deletion strains and exhibited a similar transcription pattern. The strain containing ΔS612 was the most similar to wild-type in terms of maximal expression and the timing of repression. ΔS547 seems to require a slightly longer time period to reach maximal activity, which is somewhat lower compared to that of wild-type. ΔS357 and appears to have a lower level of activity for a relatively shorter period, switching off earlier, compared to the other strains.

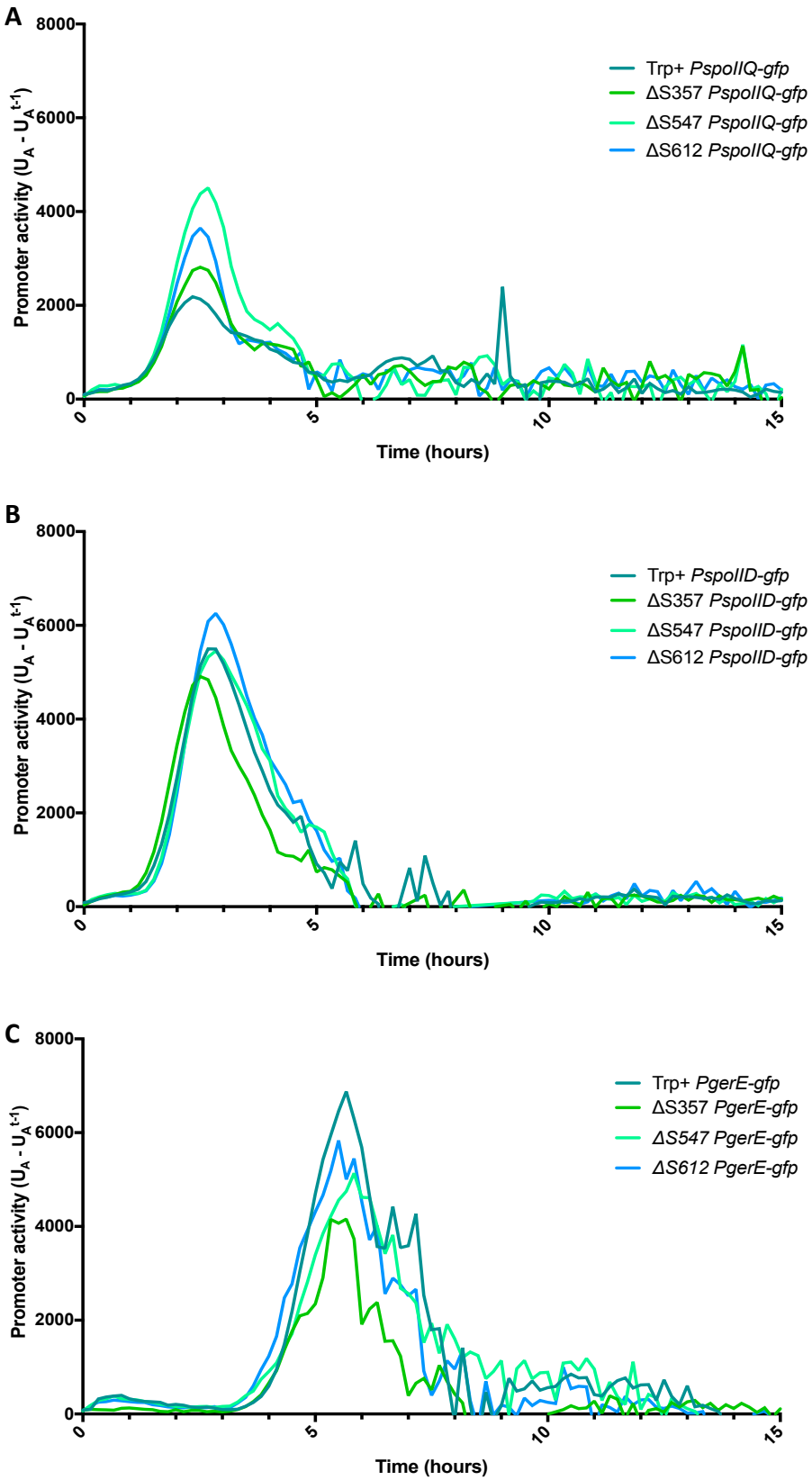


Figure 39 – Promoter activity of benchmark sporulation genes in sRNA deletion mutants. Representation of repeated experiments. (A) expression of *spoIIQ*, an early mother cell specific gene, (B) expression of *spoIID*, an early forespore specific gene and (C) expression of *gerE*, a late mother cell specific gene. GFP was monitored every 10 minutes.

4.2.4 Nutrient germination

The ability to revert from an endospore back to a vegetative cell is an important feature of a bacterial cell capable of sporulation. This is a complicated process involving the shedding of the important structures protecting the endospore from the unfavourable environment, reviving metabolism and other essential pathways for its standard lifespan and elongation into the final vegetative cell (Sinai, L. *et al.*, 2015). The first, but non-essential, step in this revival is the monitoring of the surrounding environment to initiate the cascade of events where germination is triggered. Key players in this monitoring process are the germinant receptors embedded in the inner membrane of the endospore. Small molecules are assumed to passively diffuse via the endospore coat to the inner membrane and then interact with these receptors (Setlow, P., 2014).

The genome of *B. subtilis* encodes two types of germination receptor; one activated in the presence of alanine or valine (Hudson, K. D. *et al.*, 2001), and the other that is activated via a mixture of fructose, glucose, potassium ions, and asparagine (Paidhungat, M. & Setlow, P., 2000). The former is comprised of the GerA nutrient receptor, and the latter is comprised of the nutrient receptors GerB and GerK. All are tricistronic operons, with the A, B and C subunits hypothesised to work in a complex (Ross, C. & Abel-Santos, E., 2010). For example, the GerA germination complex is made up of the three proteins GerAA, GerAB and GerAC (Hudson, K. D. *et al.*, 2001), all of which are essential for germination in the presence of alanine or valine.

To investigate the role that the sRNAs may have in the process of germination, this study purified endospores of the sRNA deletion mutants and measured the germination rates in response to different germinants. As the endospores progress through germination, they lose the phenotypic characteristic of being phase bright and therefore become less optically refractive, resulting in a drop in $OD_{600\ nm}$. By measuring this, it was found that the rate and quantity of the population reviving can be tracked.

4.2.4.1 Endospore response to germinants triggering GerA

Rates of germination in response to alanine were measured via loss of OD to essentially pure endospore crops as previously described (Ghosh, S. & Setlow, P., 2009). A deletion mutant of *gerAA* was utilised as a negative control, as the deletion completely removes the endospores ability to germinate in the presence of alanine. The germination response was found to be deficient in two of the three sRNA deletion strains, ΔS357 and ΔS547 (Figure 40).

ΔS357 was shown to have a mild delay in germination, with levels not returning to wild-type levels after 4 hours. This was phenotype was complemented back to wild-type levels upon reintroduction of S357 at the *amyE* locus. A more severe germination deficiency was found for ΔS547, with a long delay and a mild overall germination efficiency. This could also be complemented back to wild-type germination levels upon reintroduction of the sRNA in an ectopic location. ΔS612 revealed identical levels of germination in endospores, however, a mild delay in response was found. Complementation of ΔS612 did not recover this phenotype and the complementation strain had high variation. This could be for several reasons, notably, the complementation failed as the phenotype was the result of unintended genetic disruption. Upstream of S612 is *mutL*, involved in the initiation of DNA mismatch repair (Liu, J. *et al.*, 2016) and antisense is other new S segments identified by Nicolas *et al.*

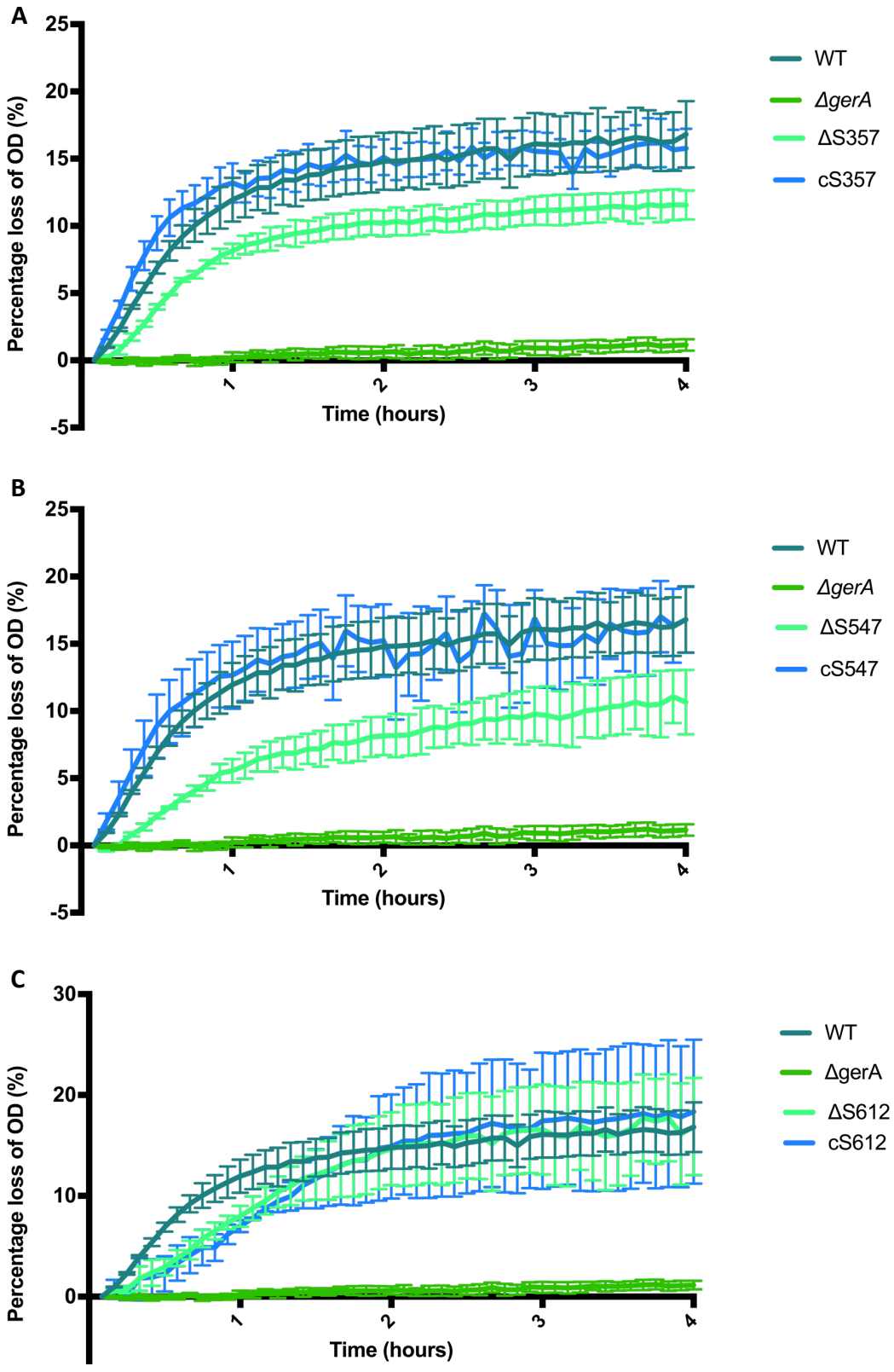


Figure 40 – Germination of wild-type and sRNA deletion mutants in the presence of the germinant alanine. (A) response of ΔS357 (B) ΔS547 and (C) S612. Data from three biological replicates is shown. Error bars represent standard error mean.

4.2.4.1.1 Measurement of dipicolinic acid release

Slow germination could be due to numerous factors. Loss of OD measures the loss of the refractive qualities of an endospore. The endospore reverts from being phase-bright to phase-dark following the rehydration of the endospores' core, which is a consequence of release of the endospore's store of dipicolinic acid (DPA) and cortex hydrolysis. Measurement of DPA release measures this event alone, and not the other events essential to germination. DPA release is considered the step after commitment to germination and is the trigger for the cortex lytic enzymes to start cortex hydrolysis (Paidhungat, M. *et al.*, 2001).

Testing of deletion mutants for DPA release in response to valine, which activates the GerA receptor in the same manner as alanine was performed. The first 30 minutes shows that DPA release was delayed for all deletion mutants (Figure 41). After this, ΔS547 and ΔS612 continue to wild-type levels and curiously, ΔS357 surpasses wild-type levels.

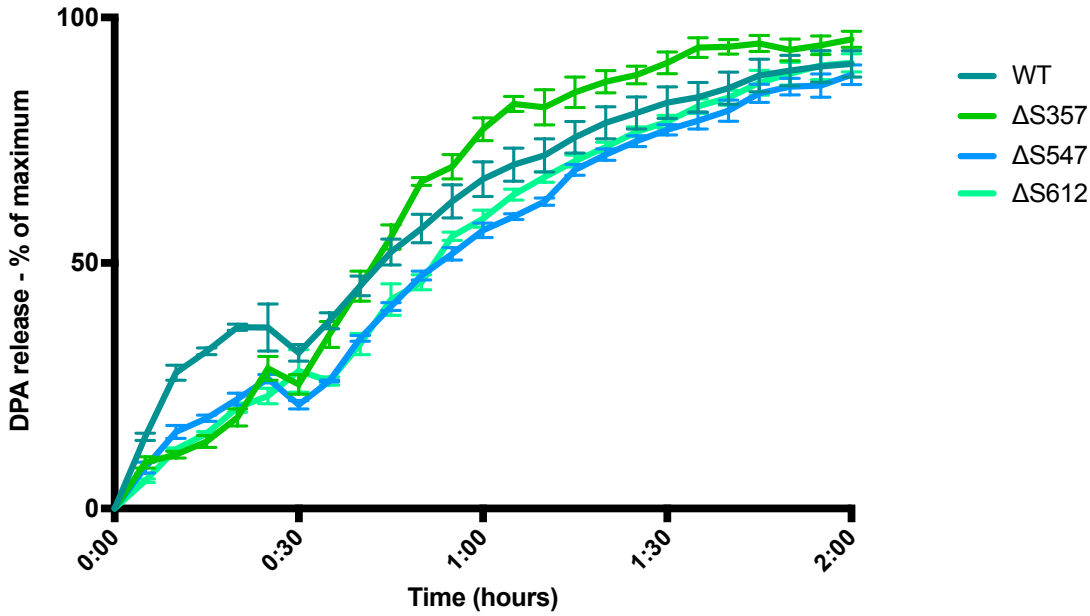


Figure 41 – Germination of wild-type and sRNA deletion mutants in the presence of the germinant valine. Endospores of strains were purified as described in materials and methods. DPA release in response to valine was performed as described previously by the Setlow laboratory (Ghosh, S. *et al.*, 2012). Data from three biological replicates is shown, error bars represent standard error mean.

4.2.4.2 Endospore response to germinants triggering GerB

In addition to alanine, *B. subtilis* has germination receptors that respond to a mixture of asparagine, glucose, fructose and potassium, namely the GerB receptor. As such, the response to AGFK was tested via loss of OD to check the alternative nutrient receptors. The response of ΔS357 was found to be the same as wild-type, with the deletion having no effect on timing and overall efficacy of germination. ΔS547, however, appeared to have a mild deficiency in both levels of endospore germination and initial response. Assays using the complemented strains described above, yielded varied responses, therefore this study was unable to confirm if this is a true phenotype or not at this time.

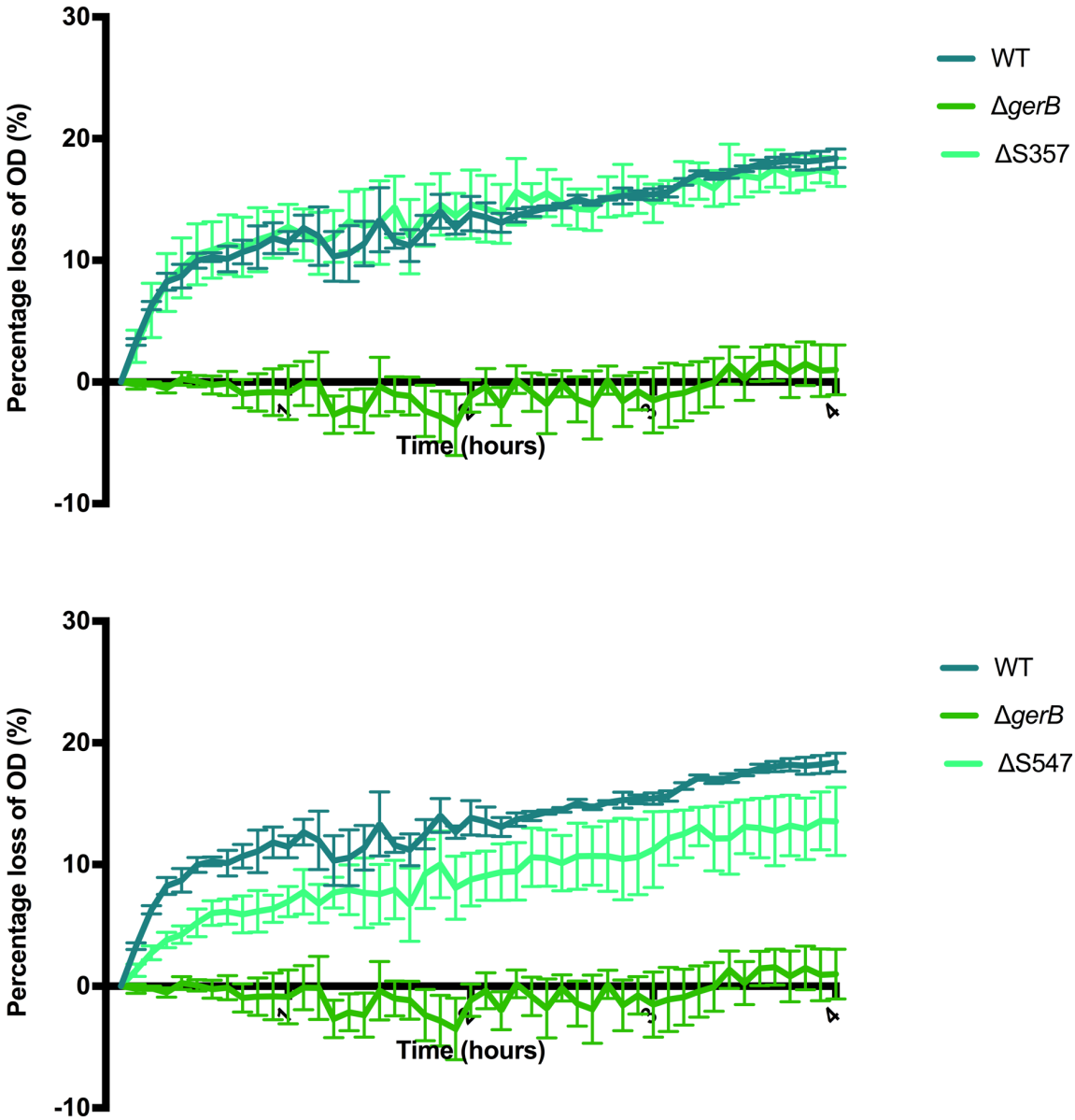


Figure 42 – Germination of wild-type and sRNA deletion mutants in the presence of the germinant mixture AGFK. Endospores of strains were purified and germination simulated as described in materials and methods. Data from three biological and three technical replicates is shown. Error bars represent standard error mean.

4.2.5 Understanding germination further

4.2.5.1 Endospore structure

The endospore coat is essential in germination. This is due to its role in not only allowing the germinants to permeate through it to the inner membrane for access to receptors (Griffiths, K. K. *et al.*, 2011), but also because it contains the lytic enzymes essential for the shedding of endospore structures, permitting eventual outgrowth (Bagyan, I. & Setlow, P., 2002).

EM microscopy allows for greater resolving power and higher magnifications than a standard light microscope and has successfully been used to identify the endospore structures (Waller, L. N. *et al.*, 2004). To determine the endospore structure of the deletion mutants, both the wild type and sRNA deletion mutants were purified and observed using electron microscopy. A minimum of 15 pictures were taken at random and Figure 44 is a representation of those images.

Endospores produced from strains ΔS357 (Figure 44A) and ΔS547 (Figure 44B) are similar in their appearance to wild type endospores. However, endospores purified from ΔS612 (Figure 44C), appear to be lacking in cortex, which typically appears white. This suggests that the cortex is not present, and the coat structure is extended. Alternatively, the cortex could be of an altered composition and as such its appearance is altered.

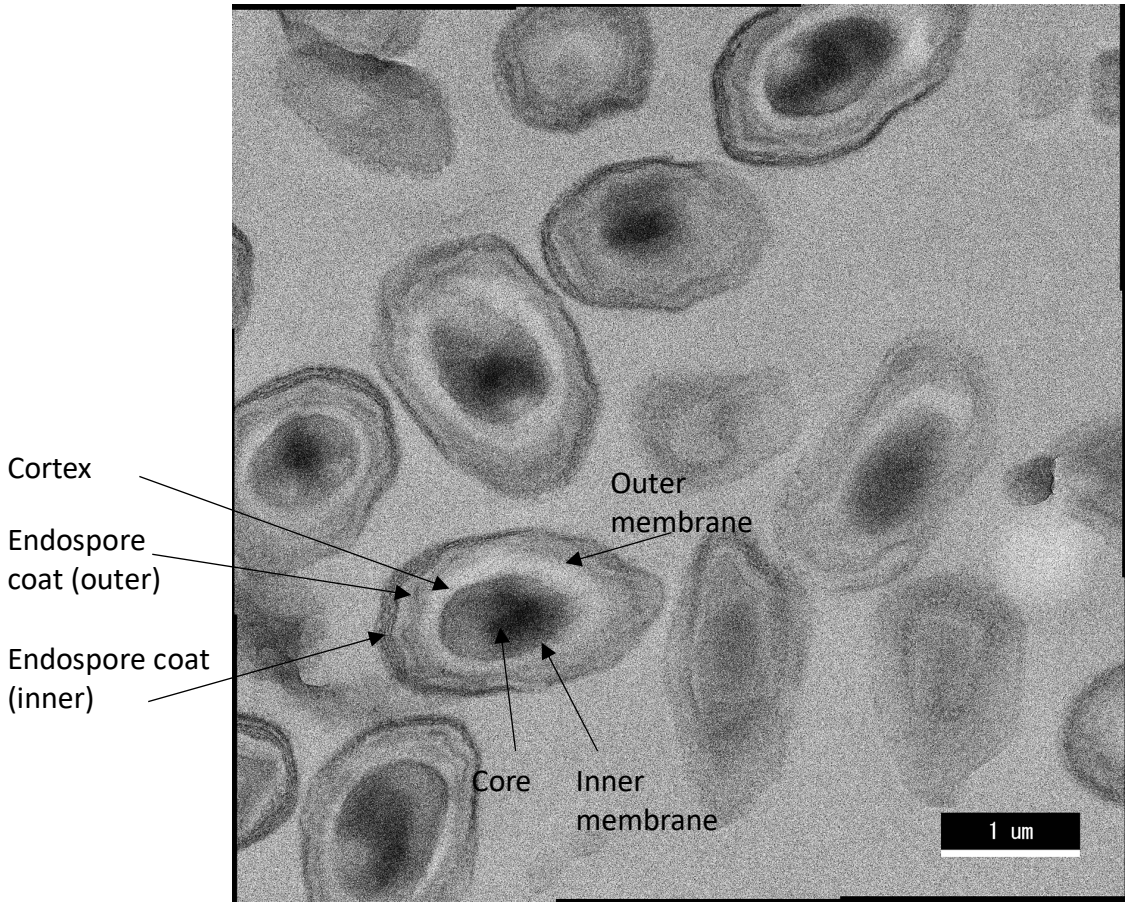


Figure 43 – EM image of wild-type endospores. Endospores were purified and captured as per materials and methods. Arrows indicate different endospore structures.

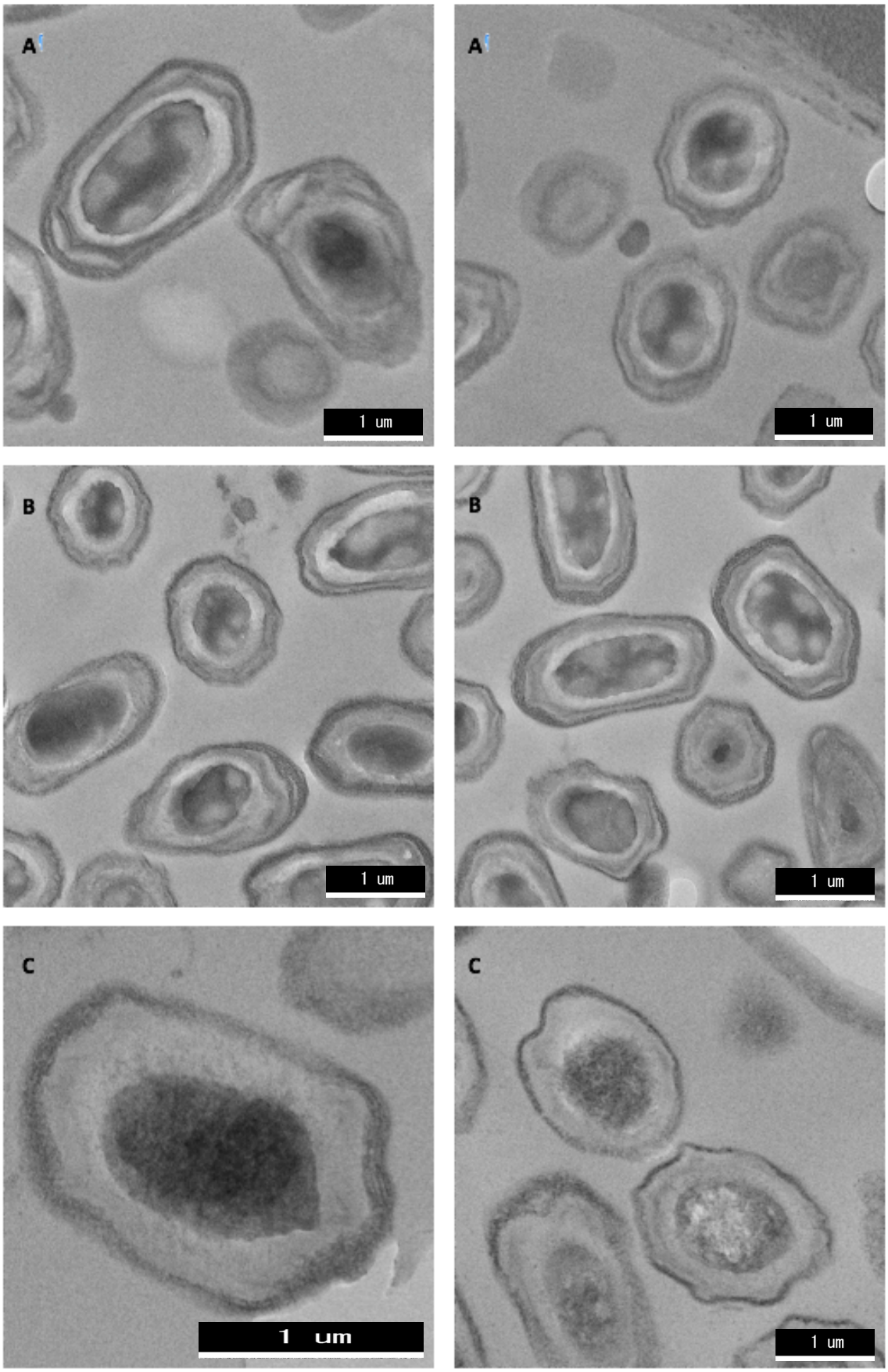


Figure 44 – EM images of endospores produced by sRNA deletion mutants. A = ΔS357 B = ΔS547 and C = ΔS612

4.2.5.2 Identification of nutrient receptor levels

One of the key potential factors effecting germination is the level of germination receptors. Receptor levels have previously been studied by deletion and overexpression of the receptors and measuring the germination response to germinant stimulation (Paidhungat, M. & Setlow, P., 2000; Cabrera-Martinez, R. M. *et al.*, 2003). For example, it was found that manipulation of the GerA receptor proteins did not affect the endospores typical response to AFGK, but did alter the rates of germination in response to the presence of alanine (Cabrera-Martinez, R. M. *et al.*, 2003).

The deletion mutants and wild-type were analysed for the presence of key factors involved in nutrient receptor-mediated responses using Western blot. GerAA and GerAC are involved in the response that is specific to alanine, and GerBC and GerKA are involved in the response to the germinant mixture AGFK. GerD is the protein responsible for the scaffold of the germinosome and clustering of nutrient receptors (Griffiths, K. K. *et al.*, 2011). SpoVAD is a likely germination protein, responsible for the uptake and release of DPA (Stewart, K. A. *et al.*, 2012). Figure 45 shows that all deletion mutants tested had no observable changes in the levels of key germination proteins.

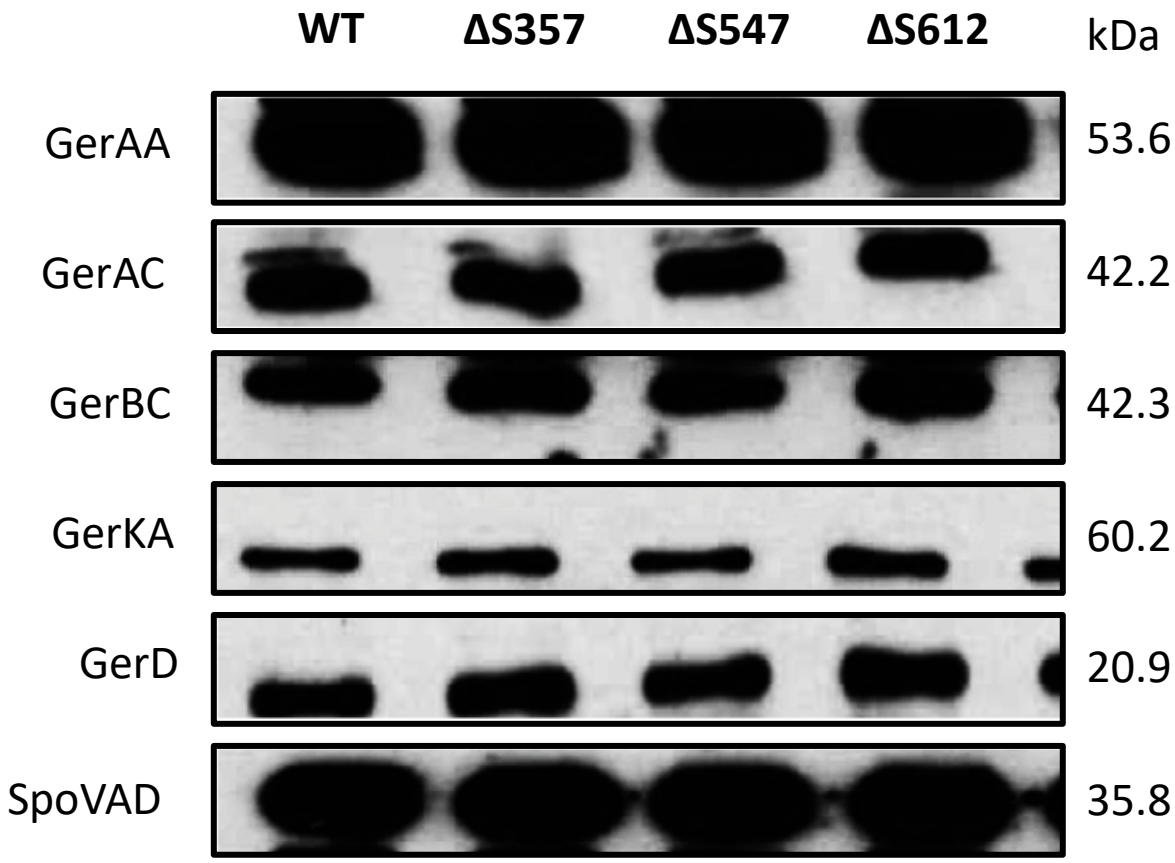


Figure 45 –Western blot of germination protein levels in wild-type (WT) and sRNA deletion mutants No differences in germination related proteins were seen. Performed by Setlow laboratory. Representative image from 3 separate biological repeats.

4.3 Discussion

Chapter three described a bioinformatics approach for the identification of putative sRNAs which are likely to be active during the process of sporulation. The workflow detailed previously yielded a total of eleven putative sRNAs with compartment specific activity, indicating its success. Four of these sRNAs were chosen to pursue the function of and are detailed further below.

The role of sRNAs in some of the key aspects in sporulation was probed for differences. First, compartment specific activity is a major indicator of expression of an sRNA during sporulation. The populations overall sporulation efficiency was tested, as well as a selection of resistance properties of the endospores. Finally, crucial for the recovery of growth from an endospore is germination, where without such an endospore would be forever dormant.

4.3.1 S357

S357 was selected to be highly likely to be an independent sRNA involved in sporulation based on the workflow in Chapter three. This was due to S357 not being antisense to another gene, and having its own distinct upshift and downshift, and being upregulated during sporulation (Nicolas, P. *et al.*, 2012). Notably, S357 RNA was found to be highly abundant in a dormant endospore by RNAseq analysis (Chapter three).

S357 was predicted to encode a forespore-specific sigma factor site for SigF or SigG (Chapter three). Another prediction utilising the web server DBTBS with more relaxed parameters than utilised previously in chapter three, allowing 5% threshold p-value, reveals a prediction of a SigF binding site one nucleotide down from transcriptional start site predicted by Nicolas, P. *et al.* (2012). At 10% threshold p-value the SigG sigma factor binding site can be predicted in the same region. Analysis of the promoter activity of S357 in the presence of sigma factor deletions, however, did not clarify the specific sigma factor involved in S357 transcriptional expression and therefore it is not clear which sigma factor regulates S357. S357 begins

transcription prior to hour 3 of the resuspension protocol and therefore is an early acting sporulation gene (Figure 30). In the future, Northern blotting of RNA extracted from wild-type, Δ*sigF* and Δ*sigG* strains using a S357 specific probe could be used to clarify the regulatory mechanism. Promoter activity of S357-*gfp* is comparatively low compared to the genes utilised as a benchmark, indicating that a more sensitive method is needed to ensure any changes in GFP levels is distinguishable above the natural noise of a plate reader detection of fluorescence.

A deletion mutant of S357 was subjected to growth in both LB and M9 with varying carbon sources. M9 supplemented with fructose was the only condition in which deletion of S357 showed a different growth pattern compared to the wild-type, whereby ΔS357 exhibited a slower exponential growth (Figure 35). Fructose is one of the 6-carbon sugars utilised in glycolysis that leads in to the Krebs cycle. Not only are these cycles important for the release of energy in the form of ATP etc. for growth, but the Krebs cycle has been found essential for sporulation (Fortnagel, P. & Freese, E., 1968). Considering the other strains and the wild–type exhibited expected growth, and that multiple experiments have shown the same deficiency, it can be expected that this is a true deficiency indicating S357 may have a role in fructose metabolism. Mutants of many of the steps in the Krebs cycle are blocked at the early stages of sporulation, with pH and divalent cation fluctuations the primary causes (Matsuno, K. *et al.*, 1999). The S357 deletion mutant is not blocked in sporulation, so if S357 does indeed play a role, it could be a subtle one. Further experiments to elucidate the potential role of S357 should include measurement of pH changes, in addition to tracking the expression of genes involved in the Krebs cycle.

Evaluation of the S357 deletion mutant in its capability to produce resistant endospores was studied, a common practice with genes implicated in sporulation processes, despite the assay being very narrow. Allowing the strains to complete the entire sporulation process once, which includes growth twice leaves room for compensational changes, and thus only major defects will show a deficiency. Furthermore, genes involved in essential resistances alternative to heat treatment or timing of sporulation and germination will be missed. ΔS357 was statistically the

same in comparison to the wild type (Figure 37). In addition, other resistance properties of the dormant endospores, specifically chemical sensitivity, revealed to be statistically consistent to wild-type levels (Figure 38). Taken together, these data indicate that neither the resistance properties or the efficiency of endospore formation are effected upon deletion of S357.

The timing of sporulation was checked by combining the S357 deletion mutant with promoter fusions to known sporulation genes (Figure 39). From this, ΔS357 appeared to increase levels of *spoIIQ* in the forespore and decrease the amount of time *spoIID* was active in the mother cell, both early in sporulation. In addition, *gerE* expression was lowered for a reduced period. Considered concurrently, it could be concluded that the normal pattern of expression for *spoIIQ* is altered upon deletion of S357. This could again tie into a potential Krebs cycle deficiency, where a subtle role as opposed to total block of sporulation could be the slowing down of essential cascades of expression.

spoIID is repressed by the transcriptional regulator SpoIIID by binding of DNA, presumably interfering with RNAP binding with its recognition sequences overlapping -35 regions (Himes, P. *et al.*, 2010). *spoIIID* is upregulated at the same time as *spoIID* as it is also activated by SigE (Eichenberger, P. *et al.*, 2004). Previously, 10 genes were identified to be transcriptionally activated by SpoIIID, however most genes regulated by SpoIIID are transcriptionally repressed. These 10 genes are important for the next sigma factor SigK (Eichenberger, P. *et al.*, 2004). ΔS357 appears to have very slight developmental issues, first with the early termination of *spoIID* expression and then with *gerE* expression reduced in both transcription levels and length of time in which the promoter is activated. Taken together, it is possible that a currently unknown process is affecting the expression levels at a later stage of early expression, causing a knock-on effect in the later stages of sporulation within the mother cell. With the feeding tube model, genes are known to influence both compartments, as such it is possible that the forespore-specific sRNA S357 could play a role in this finding. Alternatively, it could be a part of the wider scaffolding required to gradually layer the forespore with the structural requirements for endospore

structure, such as GerM from the mother cell being essential for localisation of SpoIIQ from the forespore to create the transmembrane channel linking the mother cell to the forespore.

To continue the search for phenotypes, ΔS357 was assessed for its capability to germinate under nutrient simulation. ΔS357 was shown to exhibit a deficiency in germination phenotype in response to alanine, which could be complemented back to wild-type levels upon reintroduction of S357 at a different genomic locus. As such, cortex lysis/core swelling is both delayed and not complete to wild-type levels. Further to this, initial DPA release is delayed in the first 30 minutes but then increases exponentially (Figure 41). DPA release occurs during the decision to commit to germination, offering as much as 15% of total DPA stores (Yi, X. & Setlow, P., 2010). Taken together, deletion of S357 causes delay in endospore commitment to germination. This is probably caused by the delay in GerA receptor activity, causing in turn a knock-on effect in overall germination efficiency. Whilst there is a long lag time to significant DPA release, the time taken to release 50% of maximum DPA is the same as wild-type. Furthermore, receptor levels are comparable to wild-type (Figure 45). This implicates the signalling between germinant receptors and lytic enzymes to be one of the possible reasons for impairment, somewhere in the commitment step to germination.

Previously, high levels of S357 were found to be present in a wild-type endospore (Chapter three). This would suggest that the high levels of S357 could be significant within the first 30 minutes of germination. The mechanism of signal transduction is currently unknown. No metabolism has been detected in the early stages of germination, requiring core rehydration and prior to rehydration small acid soluble proteins hold the DNA so they are protected (Segev, E. *et al.*, 2013). This indicates the signalling cascade must be present in the dormant endospore.

In summary, the evidence suggests that S357 is crucial only to GerA signalling and results in increased lag time to the commitment to germination and a reduction in endospores being able to turn phase dark. Could there be specific cascades for each

receptor? Or could S357 be working in the very first stages of receptor recognition? The very first stages could be the reason why DPA release is delayed but eventually restored.

4.3.2 S547

As described in Chapter 3, S547 was another sRNA selected for further investigation, due to high likelihood of being involved in the sporulation process. S547 has been described previously by Marchais, A. *et al.* (2011) and named CsfG. S547 has been described as an early forespore-specific transcript in accordance with findings here (Figure 27 and Figure 28). S547 was shown to lack activity when *sigF* is deleted and a deletion of *sigG* abolished the second wave of activity (Marchais, A. *et al.*, 2011), indicating a SigF sigma factor binding site. Conversely to what was previously reported, my results show that S547 activity is not affected by deletion of *sigF* (Figure 30).

Deletion of S547 displayed no effect on the growth capacities in both LB and M9 media supplemented with various carbon sources. In accordance with previous findings, deletion of S547 displayed no sporulation efficiency alterations in comparison to wild-type efficiency (Figure 37) and was no more sensitive than the wild-type to other chemical treatments. However, it has been reported that upon competition of the ΔS547 strain with the wild-type, the deletion mutant is steadily out-competed, highlighting a potential subtle deficiency (Marchais, A. *et al.*, 2011).

The competition experiment involved mixing two strains, one the wild-type and the other the deletion mutant of S547, to successive rounds of growth, sporulation and germination. This forces a subtle phenotype to be revealed as each round the strains must compete for nutrients, particularly during germination, outgrowth and then finally growth. A sample is taken after 24h of growth and sporulation in a nutrient rich media which supports sporulation and is heat shocked, selecting for endospores, and reintroduced into fresh nutrient media to start the round again. Marchais, A. *et al.* (2011) found that the ratio of wild-type to mutant dropped from 50:50 to 98:2.

The timing of sporulation upon deletion of S547 from Figure 39 was equivalent to wild-type in the mother cell, with *spoIID* and *gerE* activity largely unchanged. Within the early forespore, however, *spoIIQ* expression was double that of the wild-type and expression was maintained for longer. This could be indicative of disrupted cascades of expression. The increase of *spoIIQ* in the ΔS547 background is unusual. No studies have been done on the overexpression of *spoIIQ* so the effects of the increase of *spoIIQ* transcription levels cannot be predicted. However, both S547 and *spoIIQ* are expressed in the forespore and are activated at the same time therefore suggesting the potential for interaction is possible (Nicolas, P. *et al.*, 2012).

The next sporulation stage that this study investigated was germination. With alanine as the sole germinant, deletion of S547 resulted in a strong defect in germination which was then successfully complemented back to wild-type levels (Figure 40). Measurement of DPA release when the same nutrient receptor is triggered resulted again in a delay, which takes a long time to return to the same level as wild-type. Furthermore, when tested for response to AGFK, which triggers the other set of nutrient receptors, S547 consistently had a slower response time in comparison to wild-type (Figure 42). Taken together, these data indicate that endospores of ΔS547 are impaired in general.

The deficiency in germination levels in S547 supports the previous findings of Marchais, A. *et al.* (2011) where despite there being no overall sporulation deficiency, a severe defect was found upon competition with rounds of sporulation with the wild-type. Taken together, the data indicates that the decrease in germination efficiency is the cause for the ratio to shift in favour of the wild-type, eventually out-competing the ΔS547 strain.

These experiments suggest that the role of S547 in sporulation is in the germination stage of sporulation. The mechanisms of control that S547 has on germination are unknown at this stage and further experiments should be performed on non-nutrient stimulation of germination, to confirm the germination phenotypes are

nutrient receptor-specific to ultimately confirm the role of S547 in the process of germination.

4.3.3 S612

As described in Chapter 3, S612 was another sRNA selected for further investigation, due to the high likelihood of being involved in the sporulation process (Figure 27). In contrast to the other sRNAs taken further, S612 promoter activity can also be seen in the mother cell. This can be explained by the expression profile found in Figure 29, where a small amount of expression can be seen to occur prior to septation.

The growth of ΔS612 was unaffected in LB and M9 with various carbon sources. Furthermore, sporulation efficiency was unaffected, albeit inconsistently (Figure 37). Sensitivity of the endospore to chloroform stress was decreased in contrast to the wild-type (Figure 38). Endospore killing via chloroform has been shown to be altered due to a deletion of *stoA,* a membrane bound thiol-disulphide oxidoreductase important in the construction of the cortex (Erlendsson, L. S. *et al.*, 2004). The deletion of *stoA* decreased survival to chloroform, opposite to the effect seen here. However, evidenced further by EM pictures in Figure 44, endospores of ΔS612 lacked a noticeable cortex layer, like inactivation of *stoA*. As such, the role of S612 may be similar to that of *stoA* regulating a gene in the synthesis of the cortex.

The germination response to alanine as a sole germinant was slower in the ΔS612 than the wild-type but eventually did get to wild-type levels. However, this could not be complemented successfully. DPA release was also shown to be significantly impaired whereas levels of germination proteins were found to be unchanged. Further work should focus on the endospore structure and composition of ΔS612, particularly the cortex.

4.3.4 S849

S849 is a putative sRNA that is upregulated during sporulation and now shown to be active in the forespore (Figure 27). Analysis of its activity during sporulation indicated

active in the later part of sporulation, primarily being upregulated only one hour prior to *gerE* (Figure 30). Deletion of the sporulation-specific sigma factors, however, did not abolish *gfp* activity. A deletion mutant was created but its growth in LB was not ideal, suffering from a long lag time and therefore was not ideal for studying further. The role of this sRNA in the sporulation of *B. subtilis* remains unclear. Further experiments are needed to elucidate its role. Prior to this, its growth capabilities in complex media in comparison to defined media is puzzling and therefore further analysis of growth in LB should be studied.

5 Chapter five: exploration of methodology for finding targets of sRNAs

5.1 Introduction

5.1.1 Sporulation is a heterogeneous process

Bacillus subtilis is a bacterial model organism for cellular differentiation due to its genetic tractability and the many different cell states that are achieved through various developmental processes, including sporulation. Upon nutrient starvation, a population of cells is not homogeneous; each cell makes its own individual decision for survival, resulting in part of the population opting to undergo sporulation to form an endospore whilst others remain as vegetative cells and choose different survival methods including using the stringent response (Narula, J. *et al.*, 2012). There are many tactics available for survival; *B. subtilis* may become more motile to search for nutrients, or it may secrete a variety of toxins to destroy potential competition for nutrients and forage available proteins (Schultz, D. *et al.*, 2009). It has been suggested that this behavioural variability is a "bet-hedging" mechanism that ensures that part of the population commits to the time consuming, irreversible, and metabolically costly process of sporulation to ensure survival in extreme conditions, whilst another portion of the population is ready to grow quickly if the environmental conditions become more favourable (de Jong, I. G. *et al.*, 2010). Other environmental cues that influence bacterial lifestyles are considered together with nutrient limitation, for example detection of quorum sensing peptides which are present in the extracellular environment to help coordinate gene expression across the bacterial population in a cell density dependent manner (Schultz, D. *et al.*, 2009).

There is no on/off switch to start endospore formation, but rather a series of signals aggregated into a decision. The sporulation "master regulator", Spo0A, is activated upon nutrient stress via a phosphorelay process creating phosphorylated Spo0A (Spo0A~P) (Stragier, P. & Losick, R., 1996). Spo0A-P accumulation is measured to determine the progression of the sporulation initiation response amongst other signals (Narula, J. *et al.*, 2012). Due to this, sporulation has a variable clock rate, meaning that the likelihood of sporulation proceeding is dependent upon each cell individually responding to fluctuations in the concentrations of Spo0A-P which varies as the outside environment changes (Schultz, D. *et al.*, 2009). Genes under the

control of Spo0A-P with a low threshold are those required for other survival tactics such as biofilm formation or cannibalism. This low threshold means that only a small amount of Spo0A-P is required to activate them and thus these genes are activated before others in the regulon, to try other means of survival before committing to the costly process of sporulation (Chastanet, A. *et al.*, 2010). Genes that are under the control of Spo0A-P with a high threshold are needed for the initiation of sporulation. Therefore, in a *B. subtilis* culture subjected to nutrient starvation, there will be a varied population of cells expressing different sets of genes required for various stress-response mechanisms.

5.1.2 RNA preparation during sporulation results in variable signals and noise

To further the current knowledge of endospores, studies have moved from classical microbiology to 'omics' tools, including genomics and transcriptomics. Previous studies with transcriptomic approaches have been carried out (Iwanicki, A. *et al.*, 2013; Molle, V. *et al.*, 2003) and this enabled the identification of many of the genes involved in sporulation. For example, the genes activated within the forespore by the major sporulation-specific sigma factors (SigF and SigG) have been identified via microarrays (Wang, S. T. *et al.*, 2006). However, the application of next generation sequencing to study sporulation has proven to be problematic (Eijlander, RT *et al.*, 2016). Eijlander *et al.* (2016) demonstrated the regulatory activity of SpoVT, which is essential for sporulation in *Bacillus cereus*, but acts as a fine-tuning regulator in *B. subtilis* and therefore is non-essential. As part of their study, they utilised RNAseq with a deletion mutant of *spoVT* and the wild-type at three sporulation time points. In doing so, they observed large variation in gene expression values between strains and biological replicates. As such, when comparing Spearman Rank correlation, only one time point was accepted with two biological replicates clustering together despite originally seeking to use three time points.

5.1.3 Dormant endospores could hold the key to germination issues

Endospores of pathogenic bacterial species are capable of surviving through standard sterilisation procedures, therefore posing a risk to health. Consequently, dormant endospore characterisation is of major interest to industry where further understanding is hoped to ensure the safety of food and medicines, with further insight needed before preservation methods can be improved upon. Our understanding of the process of endospore formation remains limited, impeded by the very resistance properties of the endospore which are essential for survival but renders common experimental techniques unsuitable.

The dormant endospore is comprised of several distinctive structures: the coat, outer membrane, cortex, germ cell wall, inner membrane and the central core (Figure 46). At the centre of an endospore is the core, comprised of a complex mixture of Ca^{2+}-dipicolinic acid (DPA), DNA bound to small acid soluble proteins (SASPs), RNA, proteins such as ribosomes and enzymes, and divalent cations (Setlow, P., 2007). The make-up of an endospore core is essential for certain resistance properties of an endospore against stresses, such as UV and heat resistance (Nakashio, S. & Gerhardt, P., 1985; Setlow, P., 2001). The core of an endospore is dehydrated in comparison to the standard cell (Beaman, T. C. *et al.*, 1982) and is thought to protect the numerous enzymes in the endospore and provide resistance in dormancy (Sunde, E. P. *et al.*, 2009). DPA has been linked to the maintenance of low water content within an endospore, with removal of DPA lessening resistance properties (Paidhungat, M. *et al.*, 2000).

Engulfment of the forespore during endospore formation leads to a double membrane surrounding the forespore. These two membranes are termed the inner and outer membrane and are present in the final endospore. The function of the outer membrane is not clear, but the inner membrane is the location of crucial nutrient receptors for germination (Hudson, K. D. *et al.*, 2001). Between the inner and outer membrane is where the primordial germ cell wall is assembled, ready to be the new cell wall upon germination of the endospore (Atrih, A. *et al.*, 1996). Also

between the inner and outer membrane is the cortex, implicated in the maintenance of dehydration of the core, which consists of a thick layer of modified peptidoglycan (Popham, D. L. *et al.*, 1999; Lewis, J. C. *et al.*, 1960). The cortex modifications of typical peptidoglycan allow it to be specifically degraded by cortex lytic enzymes, CwlJ and SleB, that are important for the appropriate timing of germination (Zhang, P. *et al.*, 2012).

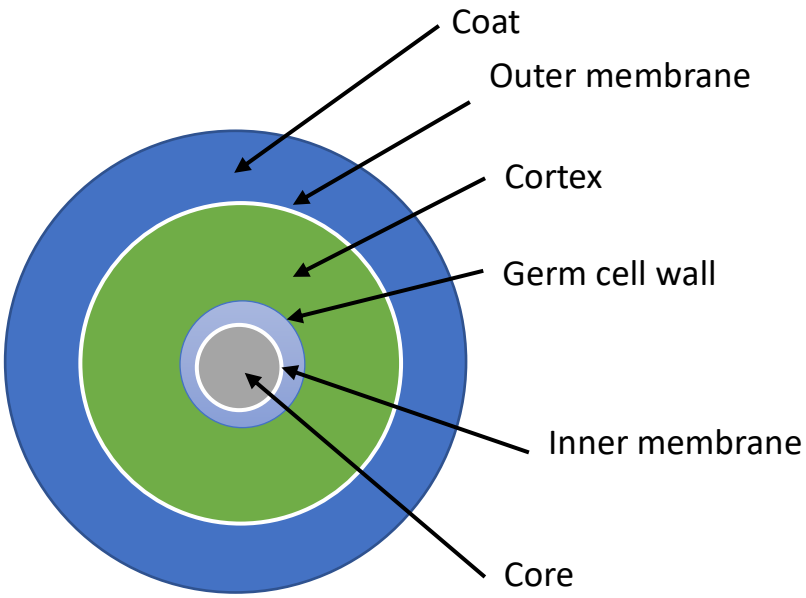


Figure 46 – Schematic of the composition of a dormant endospore.

The coat is the final layer of the endospore, made up of the inner and outer coat, which are both synthesised from the mother cell (Zheng, L. B. & Losick, R., 1990). There is also a basement layer and outermost crust (McKenney, P. T. *et al.*, 2010). Layers of the coat mostly consist of proteins highly linked with disulphide bonds and includes minor carbohydrate components. The coat can be further subdivided in to two fractions; soluble and insoluble (Pandey, NK & Aronson, AI, 1979). The endospore coat is crucial for many of the resistance properties associated with an endospore. The coat is not a barrier to small molecules, instead it acts more as a sieve primarily providing protection from larger molecules which could be damaging to the endospore, such as lysozyme (Takamatsu, H. *et al.*, 2000). In addition, the coat must allow small molecules to pass through for the germination process. Many coat-associated proteins are only temporarily present for the assembly of the coat and

therefore are not present in the final endospore. As such, deletion of many of the coat proteins have previously revealed no observable effect on the phenotypic resistance or structure of an endospore (Donovan, W. *et al.*, 1987). This could be because some of the coat proteins are involved in enzymatic roles, such as protein cross-linking and degrading toxins. There are also other coat-associated proteins, such as CwlJ that is involved in the process of germination through its cortex degrading activity (Ishikawa, S. *et al.*, 1998). The coat structure varies for different bacilli, presumably because of environmental circumstances the different species inhabit. For example, *B. thuringiensis* and closely related organisms have an additional layer, termed the exosporium (Stewart, G. C., 2015).

5.1.4 Techniques to identify proteins in dormant endospores

Many studies focus on looking at individual proteins via western blots or SDS-PAGE of extracts of endospores (Abe, A. *et al.*, 1993; Kuwana, R. *et al.*, 2002; Paidhungat, M. & Setlow, P., 2001). Proteomics aids the high-throughput identification of proteins which can provide a global snapshot of nearly all the proteins present in a sample. Recent advances in mass spectrometry technology have increased the number of proteins that can be readily detected and have increased the sensitivity, making the data gained via proteomic studies even more informative. It is possible to analyse the composition of the multi-layered structures that ensure its survival in harsh conditions by utilising this technique on protein extracts prepared from purified endospores. Dormant endospores pose a unique challenge, with approximately a third of proteins being insoluble in typical extraction procedures due to the heavily cross-linked nature of the coat (Serrano, M. *et al.*, 1999). Despite this, efforts have been made to utilise this method to successfully identify many proteins present in a dormant endospore.

5.1.4.1 Findings from previous proteomics-based studies

5.1.4.1.1 Inner endospore membrane

The most comprehensive study of proteins within an endospore, in terms of number of proteins identified, focussed specifically on the endospore inner membrane (Zheng, L. *et al.*, 2016). A total of 929 proteins were identified in at least two biological replicates, with 334 of those being assigned as membrane proteins. When isolating the inner membrane, it is impossible to not also potentially collect cytoplasmic and core proteins despite efforts made to minimise this contamination. Therefore, the 929 proteins are those found in the inner membrane, but also could be through to the core. The study identified 46 proteins that had not been previously identified in other proteomic studies. The identification of the membrane proteins was based on the results of six bioinformatics tools, but it should be noted that the inner membrane is unique and therefore standard prediction tools may not be entirely accurate for this location. The study also found 1316 proteins for vegetative cells, 880 of which were found in at least two replicates and overlapped the inner membrane significantly.

5.1.4.1.2 Endospore coat

The coat is particularly challenging to analyse due to its inherent resistance properties. As such, many groups have used strong alkalis, detergents and reducing agents alone or in combination to analyse endospore coat proteins (Aronson & Fitz-James, 1976; Jenkinson et al., 1981; Donovan et al., 1987), but many of these early studies were limited by the detection technology despite the strong treatments utilised.

Tightly associated coat proteins can be insoluble when utilising standard protein extraction techniques. Lai, E. M. *et al.* (2003) carried out a preliminary proteomic study on the *B. subtilis* endospore coat and identified 38 proteins. Their extraction procedure involved boiling the purified endospores in SDS or LDS, followed by sonication with urea and detergent. They found that the best results involved freshly collected endospore samples in small volumes of SDS. The principles of coat isolation

have not been improved upon for a substantial time so recent efforts have been made to advance the protein extraction method (Goldman, R. C. & Tipper, D. J., 1978). A gel-free protocol was recently applied to specifically target the insoluble fraction using these chemical disruption principals (Abhyankar, W. *et al.*, 2011). This method involved bead beating purified endospores, the removal of non-covalently linked proteins with NaCl and finally washing the sample with a mixture of SDS and β-mercaptoethanol. This procedure successfully identified 55 proteins.

5.1.4.1.3 Whole endospore

Lagging behind characterisation of the coat is the identification of all the proteins across all layers of a dormant endospore. One of the most comprehensive studies that focussed on identifying novel endospore proteins is by Kuwana, R. *et al.* (2002). In this work, a protein sample of endospores was solubilised in Laemmli loading buffer and boiled, subsequently one-dimensional PAGE and liquid chromatography was utilised to analyse the extracted proteins. They found a total of 154 proteins, which to date is one of the more comprehensive whole endospore proteomics studies. 69 of which had not been discovered in an endospore prior to their study. Many of the endospore coat proteins are found in many gel slices, indicating that these proteins are either heavily processed or cross-linked. Whilst the Laemmli loading buffer can be considered a rather harsh treatment due to its components giving a highly denaturing environment, even with further treatment with SDS and β-mercaptoethanol (that are typically sufficient for disrupting bacterial membranes) it was not enough to solubilise the entire endospore proteome. Similar protocols have also tried and failed to capture the entire endospore proteome, with studies yielding fewer proteins with more complicated methods (Mao, L. *et al.*, 2011). Mao, L. *et al.* (2011) identified 71 proteins from whole endospores, however this study used SDS-DTT treatment, followed by 2D electrophoresis and tryptic digestion.

Whole endospore proteomic protocols were improved recently in *Bacillus anthracis* due to the use of trichloroacetic acid over bead beating (Deatherage Kaiser, B. L. *et al.*, 2015). Mechanical disruption is a commonly used method of disrupting membranes, but has been found to be insufficient for full disruption of endospore

membranes (Thompson, B. M. *et al.*, 2011). Through using this technique together with chemical extraction, however, an increase of 52 proteins was achieved over bead beating alone to give a total of 547 proteins identified for whole-endospore proteomics.

5.1.4.2 Summary of challenges

The protein composition of an endospore is complex, rich in proteins that are resilient to typical approaches of protein extraction, such as those in the endospore coat, which are rich in cysteine residues and many disulphide bonds which strengthen the structure (Goldman, R. C. & Tipper, D. J., 1978). Whilst no single technique can isolate the entire protein content of an endospore, it is possible to minimise the loss.

Contamination of the protein preparation from vegetative cells is possible, despite efforts to ensure this is kept to a minimum. Generally, repeated washing and visually checking for 99% clean endospore crops is acceptable, with detergents used to ensure any vegetative cells present are lysed and any debris and protein left from this is removed through repeated washing (Tavares, M. B. *et al.*, 2013). Extraction from whole, intact endospores is essential over those that are in the process of germinating and/or lysing.

The identification of low abundance proteins is a problem when studying the entire proteome. However, more sensitive tandem mass spectrometry (MS/MS) technologies are improving the detection of these proteins. As these methods rely on ionisation, the creation of as many peptide fragments as possible is an important step. In addition, there are still issues surrounding very small proteins as these do not break into many peptides; this may mean that they do not meet peptide minimum requirements upon analysis, as more peptides being identified strengthen the likelihood of characterisation.

5.1.4.3 Advances in proteomics

To date, quantitative endospore proteomic studies have been limited in the number of proteins that could be identified by the techniques available. *In vivo* techniques, such as stable isotope labelling by amino acids in cell culture (SiLAC) involve labelling proteins via the incorporation of isotopes of an element (Mann, M., 2006; Li, G. W. *et al.*, 2014). Subsequent analysis can then be achieved by comparing the relative intensities between sample fragmentation spectra. SiLAC however, requires the organism to be an auxotroph to the labelled amino acids used, usually lysine or arginine, and these tend to only be used in protein synthesis. This naturally lends itself to be more useful in exponentially growing cells.

Label-free quantification (LFQ) is a method of mass spectrometry that does not require the use of stable isotopes or other labels (Higgs, R. E. *et al.*, 2005). LFQ is analysed by comparing precursor signal intensity profiles to infer protein abundances. Liquid chromatography coupled with MS/MS (LC-MS/MS) is widely applied in proteomics and utilises LC separation and peptide mass. Label-free quantification allows two methods of quantification, "relative" quantification between samples and "absolute" quantification to compare amounts of different proteins/absolute amounts.

5.1.5 Aims

Many sRNAs have been found to be actively transcribed in sporulation, but functions have not been identified (Marchais, A. *et al.*, 2011; Schmalisch, M. *et al.*, 2010). Previous chapters have focussed on identifying the putative sRNAs that are transcriptionally active during sporulation (Chapter Three) and phenotypically characterising the sRNAs that were identified to be present in the forespore (Chapter Four). Chapter Four successfully identified some minor phenotypic and gene expression changes in the deletion mutants ΔS357, ΔS547 and ΔS612. As such, this study plans to build upon previous findings in Chapter Four and further characterise three sRNAs expressed in the forespore during sporulation.

1. Investigate the whole proteome of an endospore and to improve previous methodologies
2. Investigate the function of sRNAs in the formation of an endospore
3. Investigate alternative methods of determining sRNA targets

5.2 Results

5.2.1 Sporulation heterogeneity yields RNA of poor quality

This study sought to further understand how the deletion of the sRNAs S357, S547 and S612 affected the sporulation process. RNA extraction during the growth and sporulation of *B. subtilis* in 2x SG medium yielded RNA with increasingly poor quality as sporulation proceeds (Figure 47). This was due to the heterogeneity of sporulation; the RNA extracted was hypothesised to come from three distinct cell populations that yielded either intact or degraded RNA. The mixed population of cells contained a mixture of vegetative, endospore-forming and dying cells. This population issue also lends itself to other omics approaches, such as proteomics, as the population retrieved is always a mixture. Whilst it would be advantageous to study the genetic implications of deletion of sRNAs, transcriptomics during sporulation was not pursued for the study of sRNA involvement.

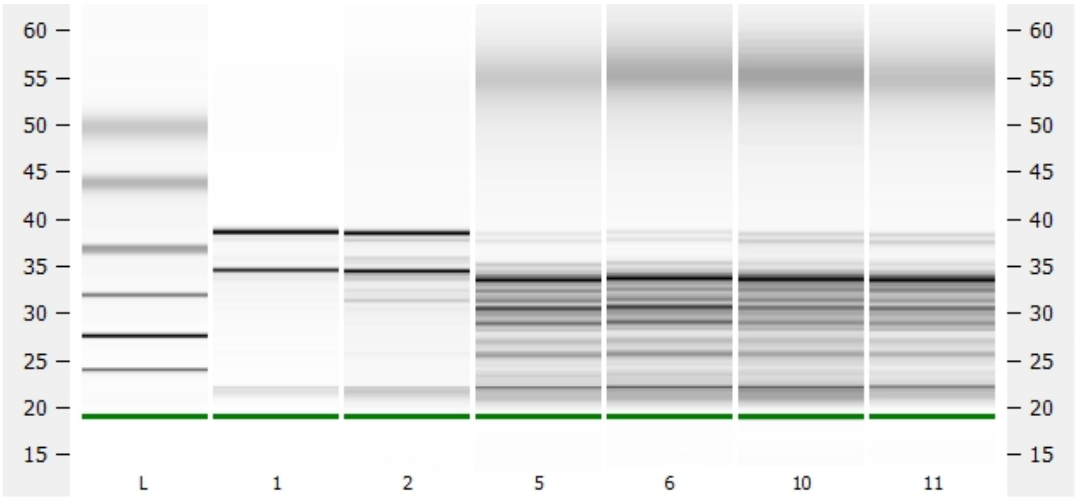


Figure 47 – RNA samples become more degraded as the process of sporulation continues. Bioanalyzer pseudogel images representative of RNA extraction during growth and sporulation in 2x SG medium.
L = ladder, 1 = mid-exponential, 2 =transition phase, 5 & 6 = early sporulation (2h after initiation) and 10 & 11 = late sporulation (7h after initiation)

5.2.2 Optimisation of endospore extraction

The protocols that were used in previous studies to extract proteins from the dormant endospore were compared to determine the optimal method that can then be used to investigate the proteome of the endospore in the wild type strain and in strains containing sRNA deletions.

The initial breaking open of an endospore is an important step for subsequent protein solubilisation. Previous studies of endospore protein composition found that a significant number of proteins are insoluble and require specific care to isolate (Abhyankar, W. *et al.*, 2011). Mechanical disruption via bead beating is a common method to disturb the endospore layers but has been found previously to limit protein yields (Deatherage Kaiser, B. L. *et al.*, 2015). Other methods include pressure breakage using a French press (Harwood, C. R., and S. M. Cutting. , 1990). De-coating solutions made of harsh chemical treatments have long been known to be the best strategy to isolate the coat fraction (Goldman, R. C. & Tipper, D. J., 1978). As such, many studies have relied on a mixture of SDS and DTT at pH 10, which is known to increase the solubilisation of proteins (Goldman, R. C. & Tipper, D. J., 1978). Furthermore, the best results have been found on freshly collected endospores in small volumes of SDS (Lai, E. M. *et al.*, 2003). As such, a combination of bead beating, pressure breakage and chemical de-coating was tested for optimal protein yields.

In general, protein extractions are specifically tailored to the properties of the proteins to be isolated and as such there are numerous options to support the extraction of less soluble proteins. This must be balanced for endospores, as the proteins present are very diverse. Common methods for extracting low solubility proteins include the use of NaCl as it maintains the ionic strength of buffers which facilitates the solubilisation of membrane proteins. In addition, the solubilisation of proteins is encouraged by using the detergent SDS. β-mercaptoethanol is another chemical used to facilitate the solubilisation of proteins, which acts as an agent to reduce disulphide bonds. An alternative reducing agent commonly used is DTT,

which is also known to reduce oxidation damage. A combination of the chemicals described above was tested for increased protein yields, as judged by SDS-PAGE.
A representation of the results obtained from the various methods can be viewed in Figure 48, as detailed in Table 14. Both pressure breakage utilising a cell disrupter and beat beating was found to be very ineffectual resulting in very little protein yield (Figure 48A). In addition, microscopy revealed that the endospores were still intact after several rounds of pressure and even after treatment with lysozyme. The data drawn upon here indicates that chemical treatment of the coat is indeed the superior method of reducing an endospore, as reported previously (Goldman, R. C. & Tipper, D. J., 1978; Deatherage Kaiser, B. L. *et al.*, 2015).

Chemical de-coating and subsequent treatments to provide denaturing conditions, such as boiling in the presence of urea, was not sufficient to substantially improve protein yield (Figure 48A, lane 3). However, once the endospores are de-coated, the endospore is sufficiently sensitive to lysozyme to allow degradation of the germ cell wall and cortex according to previous findings (Zheng, L. *et al.*, 2016). Lysozyme is, however, sensitive to pH changes and as such would not survive in the de-coating solution. For this reason, it was found that the supernatant from the coat was best to be extracted and purified separately to the rest of the proteins within the endospore. After lysozyme treatment, samples were boiled for 1 hour and then combined with Laemmli buffer and further boiled. This was observed to be very effective (Figure 48, lane 4). An additional sonication step was attempted to improve protein solubility, but was not found to be significantly helpful in increasing protein extraction (Figure 48A and B).

Table 14 – Combinations of treatments utilised. Three initial cell breakage methods were utilised in addition to a de-coating solution.

	Breakage			Treatments		De-coating solution		
Lane	Pressure	Beat beating	Sonication	Lysozyme	urea	SDS	DTT	Na
1	✓							
2	✓			✓				
3					✓	✓	✓	✓
4				✓		✓	✓	✓
5			✓	✓		✓	✓	✓
6		✓						
7				✓				

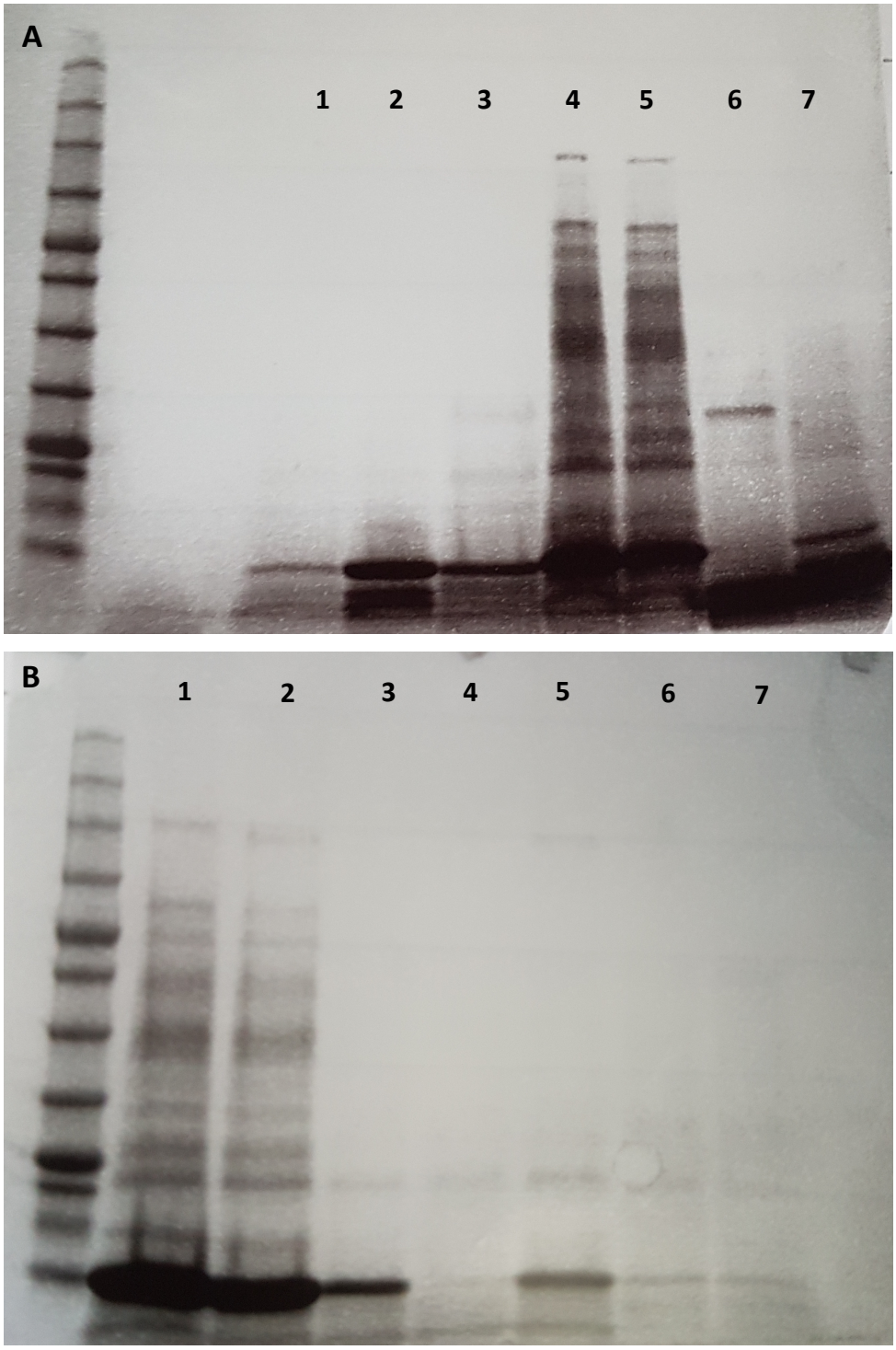


Figure 48 – SDS-PAGE of different protein extraction methods tested on *B. subtilis* endospores. (A) From left to right is ladder, a blank row then treatments 1 – 7 as detailed in Table 10. All treatments were boiled for 10 minutes in Laemmli loading buffer prior to loading. The chemical treatments, in lanes 4 and 5, were the most effective for protein extractions from the endospore, yielding significantly more proteins. (B) Combinations of the treatments shown in Table 14. These methods were found to have decreased protein yields. All bands in both gels were visualised with Coomassie intant blue.

5.2.3 Mass-spectrometry analysis of endospore preparations

SDS-PAGE followed by in-gel trypsin digestion was used to extract proteins and spilt them into peptides. Subsequently, extracted peptides were analysed by Nano-LC-ESI-MS/MS. MaxQuant software was then used for protein identification and quantification against *B. subtilis* in the Uniprot database (http://www.uniprot.org/proteomes/UP000). Following this, the Perseus software was used to visualise and statistically analyse the data (Tyanova, S. *et al.*, 2016).

A total of 2400 different proteins were identified across the twelve samples that consist of 3 biological and technical replicate samples from 4 groups (wild-type, ΔS357, ΔS547 and ΔS612). The filter criteria were set such that at least one group needed to identify a protein in at least two samples. On average, 1915 proteins of the endospore proteome were identified in the wild-type samples, 1973 proteins were identified in ΔS357 samples, 1838 proteins in ΔS547, and 1975 proteins were identified in ΔS612 samples (Figure 54). The lowest number of proteins identified belonged to the second sample of ΔS547, nevertheless, a two-tailed *t*-test between the wild-type and ΔS547 revealed the reduction to be statistically insignificant.

5.2.4 Functional Categorization of proteins recovered from the wild-type

Significantly more proteins were recovered from this study in comparison to previous studies; a total of 1967 proteins were found to be present in at least two replicates of the wild-type compared to a maximum of 929 proteins identified from other studies (Zheng, L. *et al.*, 2016). As proteome recovery was significantly improved, functional analysis was performed to determine a more detailed view of the composition of an endospore.

Classification of proteins was based on the functional classifications of SubtiWiki (Michna, R. H. *et al.*, 2016) and displayed in Figure 49. Unsurprisingly, many proteins belonged to the membrane proteins and sporulation categories. In addition, many proteins present in a dormant endospore are related to those involved in protein

synthesis, modification and degradation category, presumably present to assist in germination and perhaps left over from the final stages of creating the endospore structures. Relatively few proteins are found that are related to homeostasis and RNA synthesis/turnover; a finding that is surprising due to the demand of these proteins upon stimulation of germination.

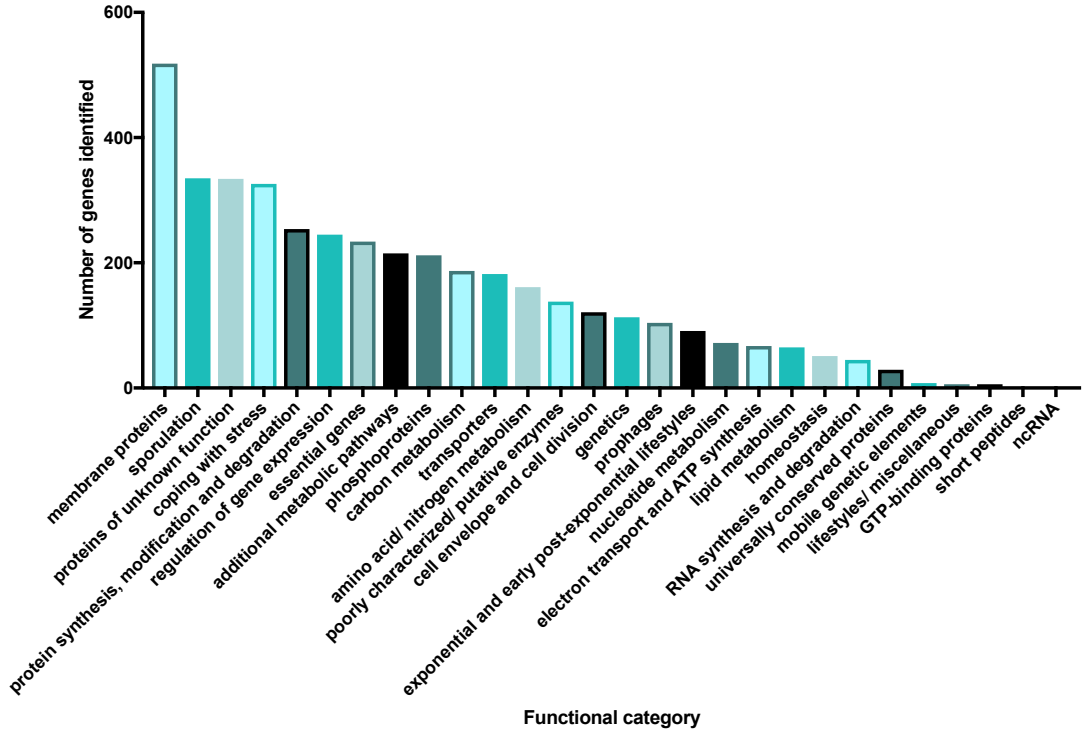


Figure 49 – Functional categories of proteins identified in the wild-type dormant endospore. Categories were assigned based on classification by SubtiWiki. Many proteins present from within a dormant endospore belong to membrane or sporulation categories whilst relatively few are associated with RNA synthesis and turnover

5.2.4.1 This analysis of a dormant endospore vs other proteomics investigations

To determine if this extraction process was successful, the proteins identified from the wild-type endospore from this study were compared to that of other proteomic studies. The most comprehensive study in terms of number of proteins returned is that of the endospore inner membrane (Zheng, L. *et al.*, 2016). A total of 929 proteins were identified in at least two of the biological replicates, with 334 of those being assigned as membrane proteins. Membrane protein classification was based on six bioinformatics tools, which have not been developed with the unique endospore inner membrane in mind, in contrast to this study which has used tools that are more predominantly based upon *B. subtilis* that include more detailed information on its sporulation lifestyle. Nevertheless, this study identified 518 membrane proteins, 184 more than Zheng *et al.* (2016) (Figure 49).

It is impossible to not also collect cytoplasmic and core proteins when isolating the inner membrane; the 929 proteins identified by Zheng *et al.* are therefore components of the proteome from the inner membrane through to the core. Zheng, L. *et al.* (2016) found 46 proteins that had not been previously isolated in other proteomic studies carried out on *B. subtilis* endospores.

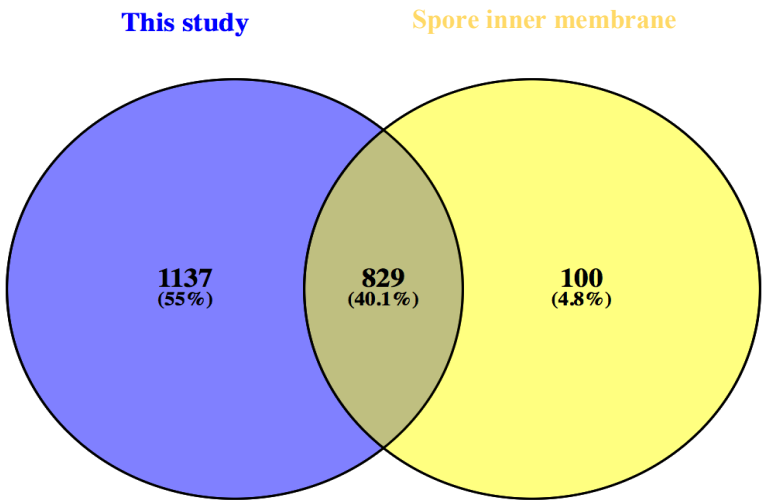


Figure 50 – The proteome of the dormant endospore identified in this study vs the proteome of the inner membrane (Zheng, L. *et al.*, 2016)

Figure 50 presents a comparison of the proteins identified in at least two replicates in both this study and that of Zheng *et al.* There is a significant overlap between the two studies, with 829 proteins shared and 100 proteins not identified in this study. 1137 proteins are unique to the dormant endospore are presented in this study, presumably being the proteins in which Zheng *et al.* would have removed during washing, as their study focused on the analysis of the inner membrane.

In addition, if a comparison is performed with the inner membrane raw data, without being filtered for being present in at least 2 replicates, this points to another 255 shared between the two studies, with an additional 137 proteins being unique to their dataset. This would therefore reduce the unique proteins in this study to 882. The proteins lacking in one replicate when compared to another may be due to the reproducibility of the capture method used, as opposed to the proteins not actually being present. As such, the proteins identified that are unique to either dataset are not by chance and could potentially be shared across samples with increased sampling.

The Zheng, *et al*. 2016 study identified many proteins that have been found to be localised in structures outside of the inner membrane and core; many proteins were identified in the coat layer. An example of proteins that are shared between our two studies that are not of the inner membrane are CotE, C, B and A which can be found in the outer coat with LipC and CotS. Additionally, SspB and SspA, which are found in the endospore core. This perhaps suggests that the extraction methods used by Zheng, *et al*. allowed for cross-contamination of proteins from other layers.

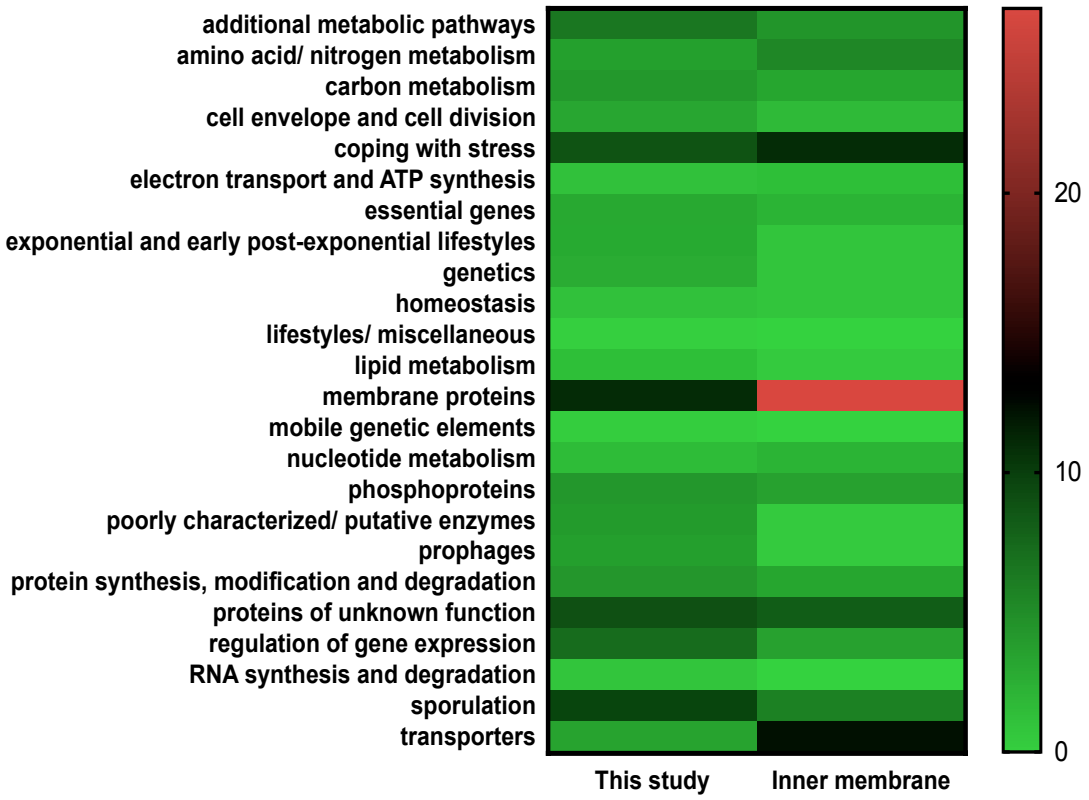


Figure 51 – Functional categorisation based on SubtiWiki of proteins uniquely found in this study or in the Zheng *et al.* 2016 inner membrane study as a percentage of the total number of unique proteins per study. Membrane proteins are clearly enriched in both studies, but particularly the inner membrane study.

Functional characterisation of the proteins found to be unique in either study indicates that the only statistically significant category to be enriched is the membrane proteins (Figure 51), with 26% percent of the unique proteins to the study of the inner membrane as opposed to 11% of the unique proteins to this study. Of the 9.6% proteins unique to this study that are present in the sporulation category, 36 are classified as coat proteins whilst no coat proteins were found to be unique to the inner membrane study. However, the proteins that are shared between the two studies indicate at least three proteins were identified that are associated with the coat; CotE, CotY and YisY. Membrane proteins are the most enriched category that were identified by both studies (Figure 52), which are typically the most technically challenging proteins to isolate due to the low solubility properties.

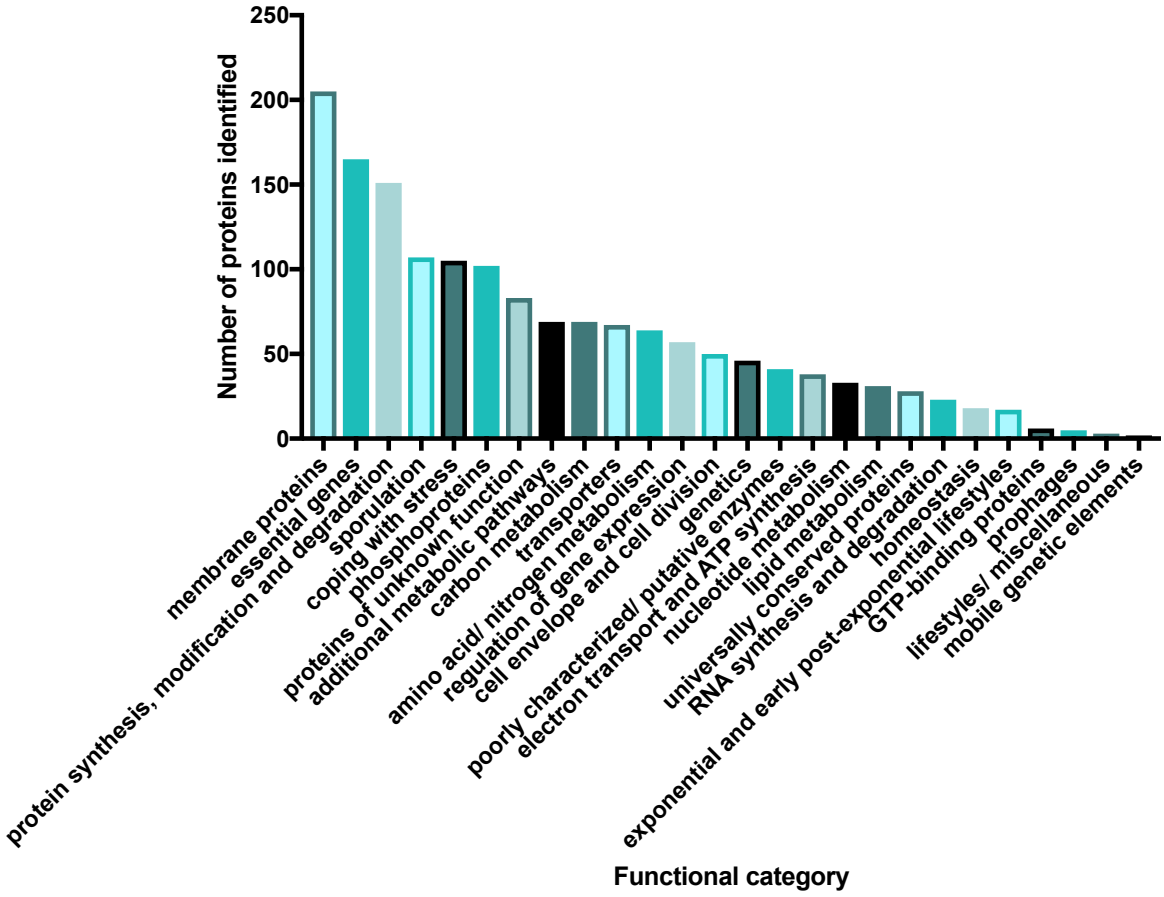


Figure 52 – Functional categories based on SubtiWiki of the proteins shared between Zheng *et al. 2016* and this study. Membrane proteins are enriched as is protein synthesis.

The next best characterised layer of the endospore is the coat that has had numerous studies dedicated to its protein composition. The endospore coat has been split up into two fractions, insoluble and soluble. It is thought to comprise of over 70 proteins and is the most studied structure of an endospore (Henriques, A. O. & Moran, C. P., Jr., 2007). High cysteine content and many disulphide crosslinks make up the insoluble fraction, but attempts have been made to study this specifically (Abhyankar, W. *et al.*, 2011). Abhyankar, W. *et al.* (2011) utilised a MOPS-buffered defined medium to induce sporulation, but endospore formation still followed the natural depletion of nutrients. Both a lab strain (PB2) and a food isolate (A163) were studied. The lab strain yielded 55 endospore coat proteins, in which 21 novel putative endospore coat proteins were identified.

This study identified 43 coat proteins, 12 fewer than Abhyanker, *et al*. Discrepancies between the two are likely due to a few factors, namely the different strains and experimental conditions that were used to induce sporulation. However, many studies that used gel based proteomic methods had found minimal coat-associated proteins. Indeed, many of the most cysteine rich components of the endospore crust are encoded by a cluster of *cot* genes (VWXYZ) (Zhang, J. *et al.*, 1993; Imamura, D. *et al.*, 2011); these were not identified by this study, with only CotZ and CotY being identified. The gel-free method also failed to identify CotVW, indicating room for improvements for both coat protein extraction methods. Of the 27 Cot proteins previously identified (Lai, E. M. *et al.*, 2003; Kuwana, R. *et al.*, 2002), 18 were also identified by this study. Additional to this, the paralog of CotS, CotI, was also identified.

In conclusion, the dormant endospore is complex and rich in structures that have previously been difficult to isolate. This study has significantly improved endospore protein yield through the optimisation and combination of previously determined protein solubilisation methods. Whilst it will always be impossible to capture all proteins present in the endospore, as the proteins are so divergent, comparison with previous studies indicate this new method is robust and able to capture many of the unique proteins across the diverse layers of the endospore.

5.2.5 Wild-type vs sRNA deletion mutants

It was hypothesised that deletion mutants of sRNAs may display distinct protein profiles that differ both from the wild-type and from each other. Endospore proteins from the wild type and the sRNA deletion mutants, S357, S547 and S612 (Figure 56 and 57) were extracted and analysed in the same manner as discussed in the previous sections. As such, Pearson correlation coefficients were calculated from all data to show if there is any similarity between the different groups of samples. Pearson correlation coefficients from all samples were generally good, being largely over 0.9 (Figure 55) and therefore displaying high similarity. As none of the strains produced distinct bands of correlation, this disproves the hypothesis. However, the second sample of ΔS547 was shown to be different to all the other samples, with the lowest correlation score of 0.83. This was further evidenced by testing of endospore extraction, where ΔS547 extraction was on occasion difficult to extract in comparison to wild-type yields (Figure 53) and may have resulted in some disparity between samples (Figure 54). As such, protein extraction was repeated until comparable protein yields was achieved.

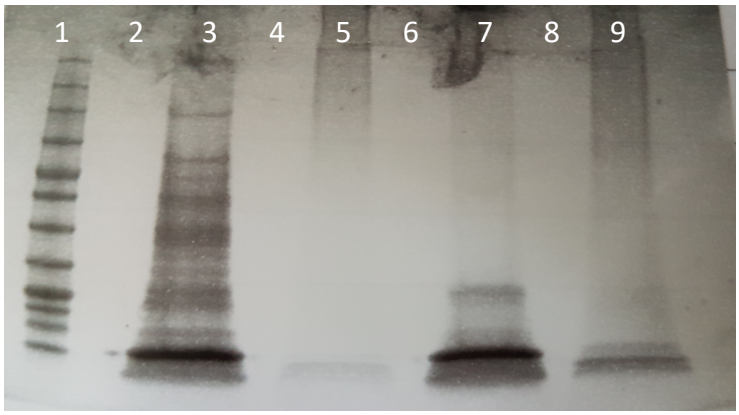


Figure 53 – Extraction of ΔS547 endospore proteins was inconsistent. A ladder was run (lane 1) a number of blank lanes were also included for clarity (lanes 2, 4, 6, and 8). Extraction of wild type (lane 3 and 5) and ΔS547 (lane 7 and 9) endospores using previously optimised method. Protein extraction yielded two separate fractions, total protein without coat (lane 3 and 7) and endospore coat proteins (lane 5 and 9).

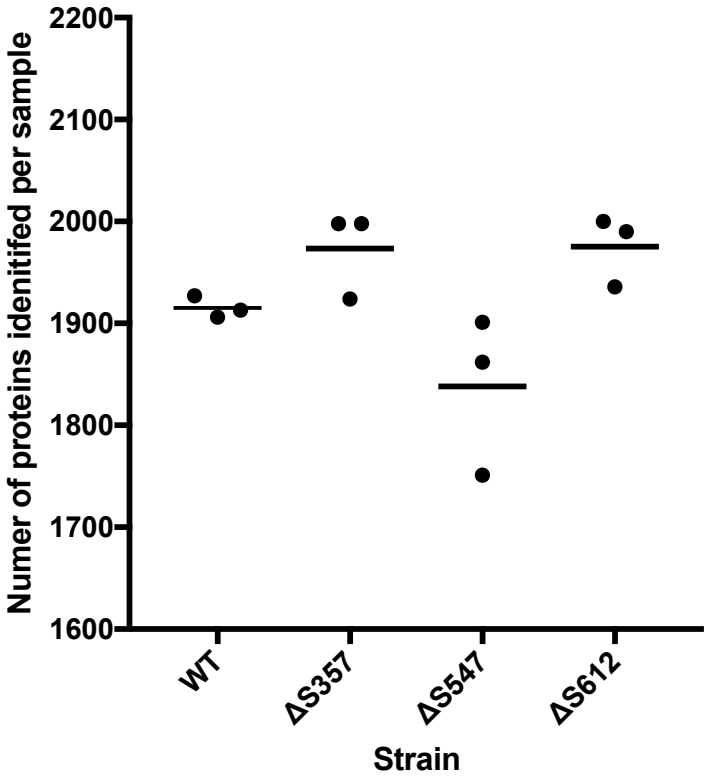


Figure 54 – Total number of proteins identified in each of the triplicate samples per strain. Each circle represents an individual sample and each line represents the mean number of proteins.

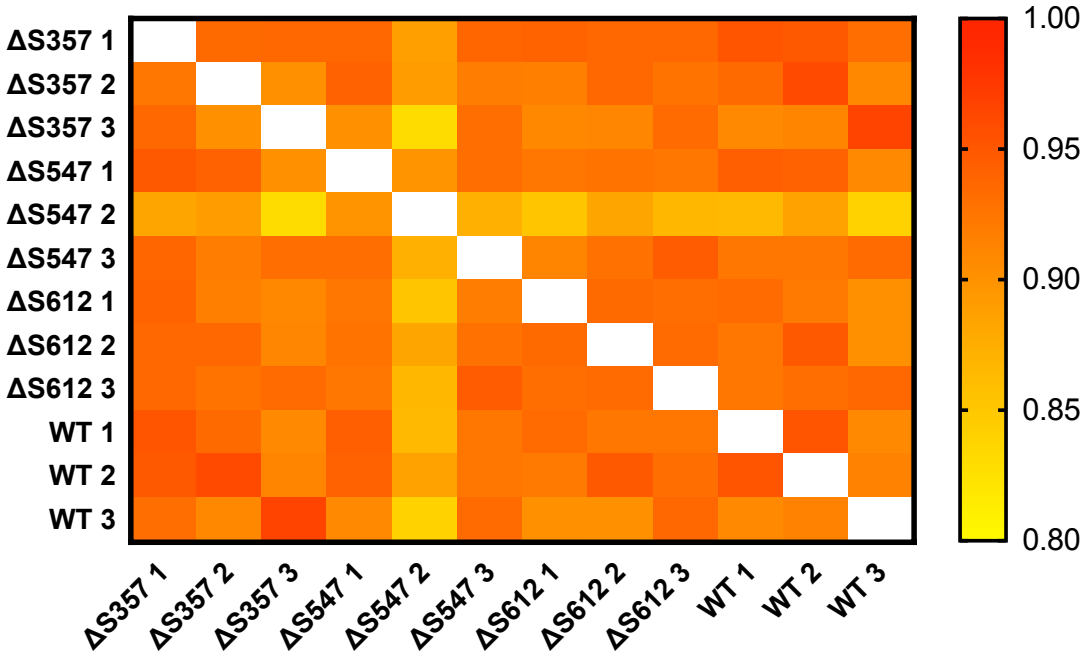


Figure 55 – Pearson correlation coefficients show high correlation between strains. The more yellow, the lower the correlation. One sample of S547 displayed lower correlation to all other samples.

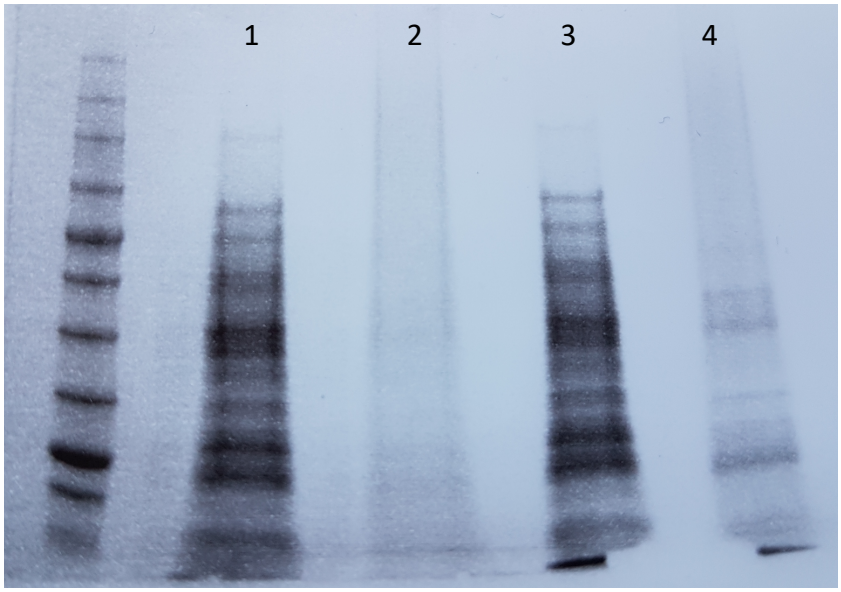


Figure 56 – Extraction of endospore proteins for wild type (lanes 1 and 2) and ΔS357 (lanes 3 and 4) for proteomic analysis. Lanes 2 and 4 are extracted proteins from the de-coating solution and therefore coat proteins.

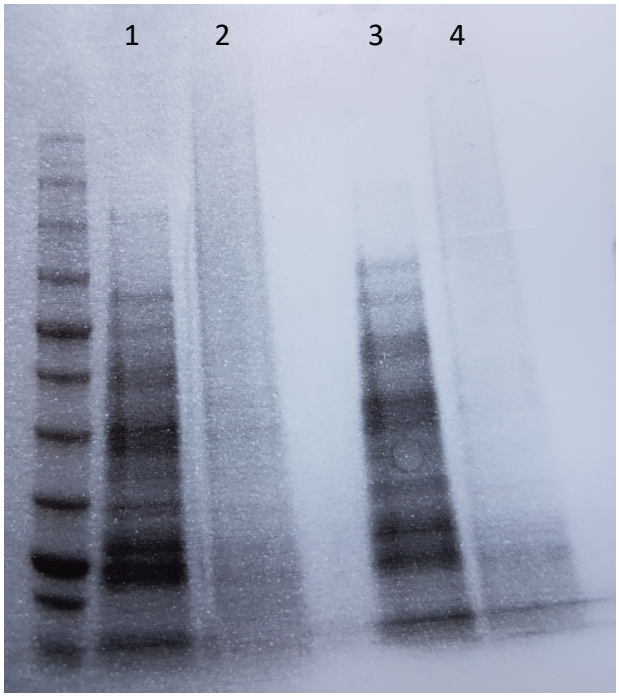


Figure 57 – Extraction of endospore proteins for ΔS547 (lanes 1 and 2) ΔS612 (lanes 3 and 4) for proteomic analysis. Lanes 2 and 4 are extracted proteins from the de-coating solution and therefore coat proteins.

Contaminants, such as human keratin were filtered out of the analysis using the Perseus software. LFQ intensities, calculated based on protein abundance in quantitative protein analysis, were log transformed to a base of 2. Log transformation of LFQ intensities creates NaN errors on values that are 0 which effects downstream analysis. Values of 0 can result for several reasons; either the protein was not present, or that the intensity was under the noise level and therefore undetectable by the instrument. This was processed in two ways. The first being accepting a protein to be faithfully present based on data gathered in multiple replicates. In addition, the second approach taken here was to first reject any proteins which did not have at least two biological replicates in one group of samples, as this allows for a scenario where a protein may be absent from at least one strain. From here, it is possible to use imputation to replace the missing values with values based on the protein distribution (Figure 58). This allows for the rejection of random proteins and corrects for limits of detection, improving the statistical power. This function was inbuilt into the perseus software. The proteins identified in this study showed a normal distribution, important for downstream analysis, after logarithmically transforming the LFQ intensities. Imputation of missing values was carried out based on this normal distribution (Figure 58).

A total of 1590 proteins are shared between all strains tested, found in at least two replicates of each strain (Figure 59). Less than 15% of total proteins identified are unique to any one strain. ΔS612 has the most unique proteins with 113, closely followed by ΔS357 with 91 proteins being unique to those samples. The wild-type has only 71 unique proteins and ΔS547 shared the most proteins, with 51 being present in only those samples.

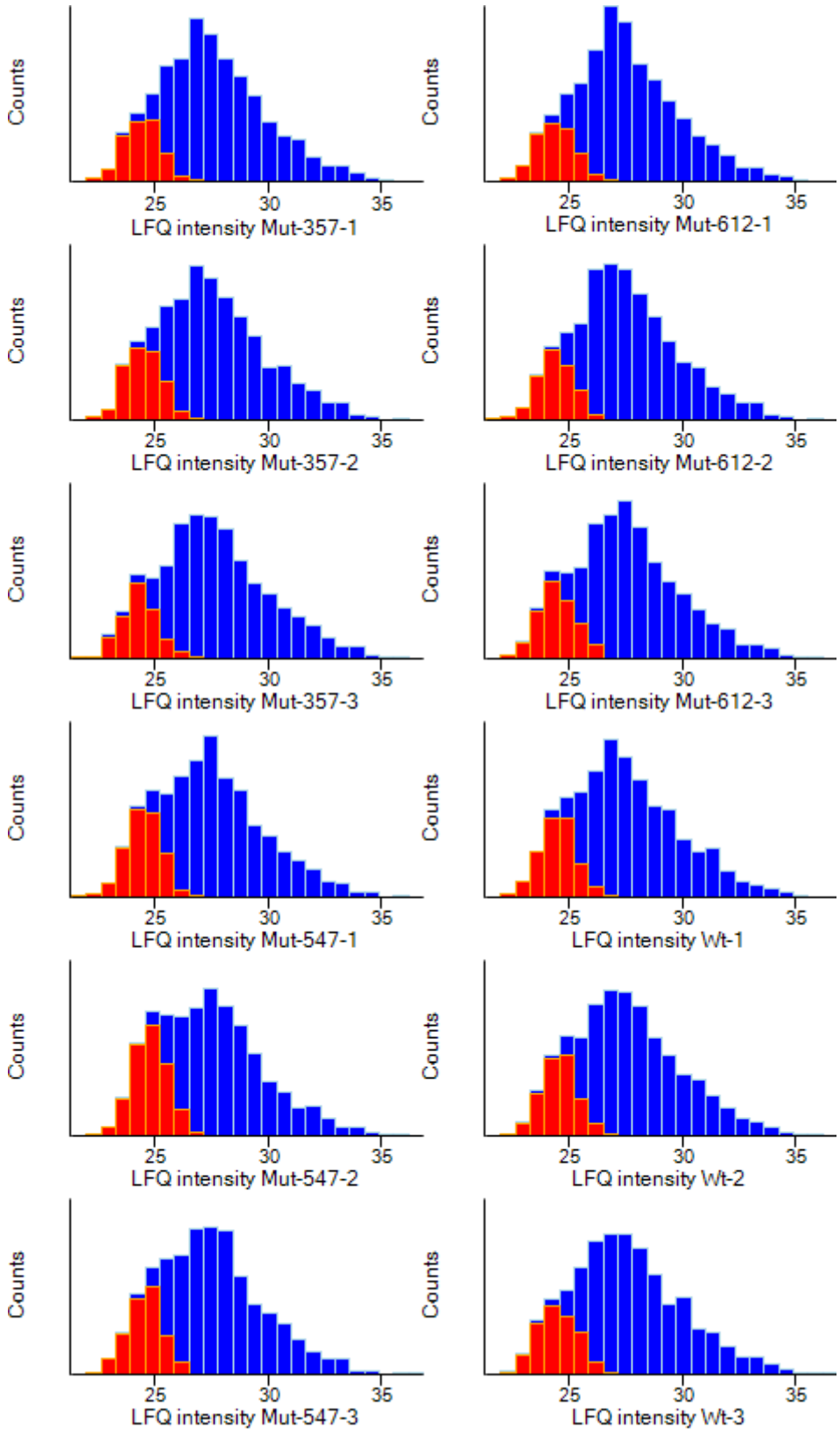


Figure 58 – Distribution of log transformed intensities for proteins identified in each sample. Missing values were replaced by imputation (in red) and were based on the normal distribution of values. Represented are the log transformed intensities for 3 separate biological replicates of each strain.

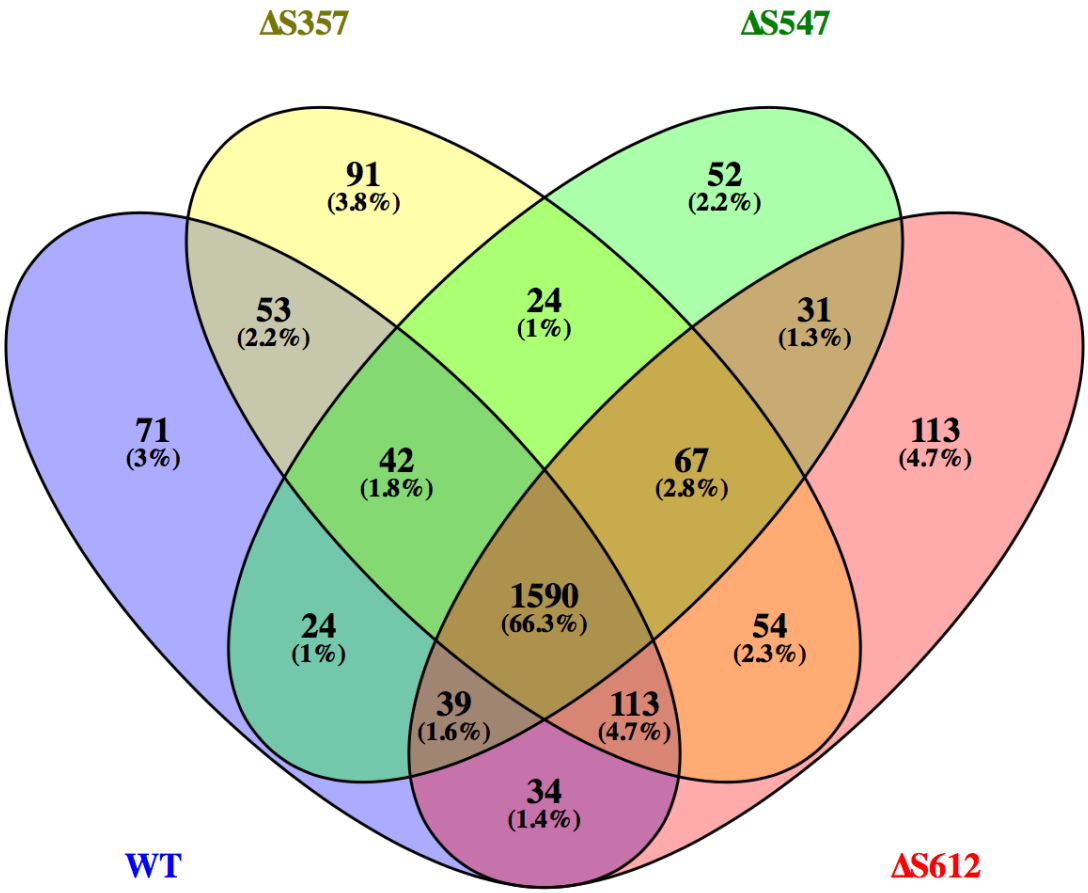


Figure 59 – Comparison of the proteins identified in each strain. Many proteins are conserved across the WT and mutants, with 66.3% of the proteins identified constituting this group. A small proportion of the proteins from each sample could also be identified as unique to that strain alone.

For all subsequent analysis between wild-type and the sRNA deletion mutants using this data set, the presence or absence of a protein is accepted based upon at least 2 of the 3 replicates.

5.2.5.1 Wild-type vs ΔS357

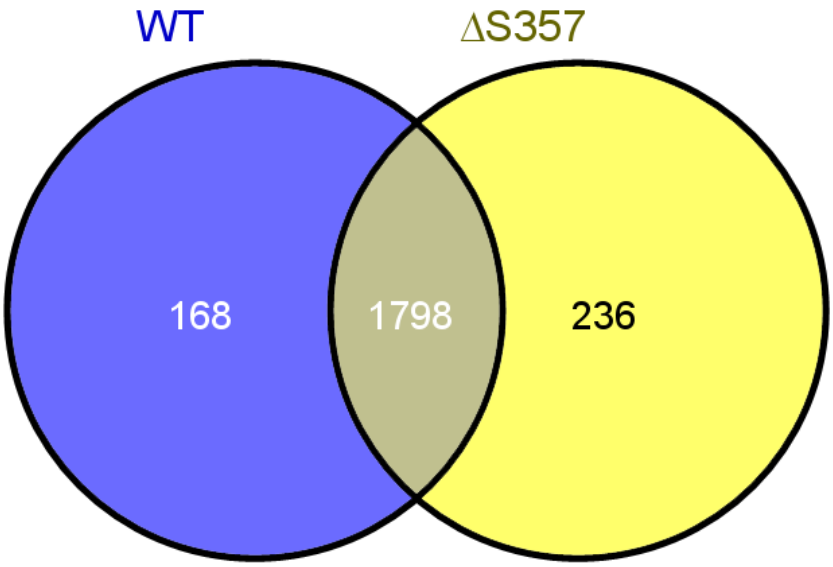


Figure 60 – Comparision of the proteins that are present in at least two replicates in the wild-type and ΔS357 endospores showing significant overlap of protein composition. However, unique proteins can be found in each sample.

A comparison between wild-type and ΔS357 samples found that 1798 of the proteins were shared. ΔS357 had a total of 236 proteins unique to its samples and the wild-type had 168 unique proteins (Figure 60). Using Bionic Visualizations' Proteomaps (Liebermeister, W. *et al.*, 2014), there are some functional differences between the unique proteins (Figure 61). The most prominent changes between the wild-type and ΔS357 strains involve diverse processes, with ΔS357 having increased proteins in transport and lacking in transcription factors. In addition to other functional categories that are also much changed, such as carbohydrate metabolism and other enzymes. Yet it remains to be seen whether these proteins allow for more efficient germination in the wild-type compared to the mutant.

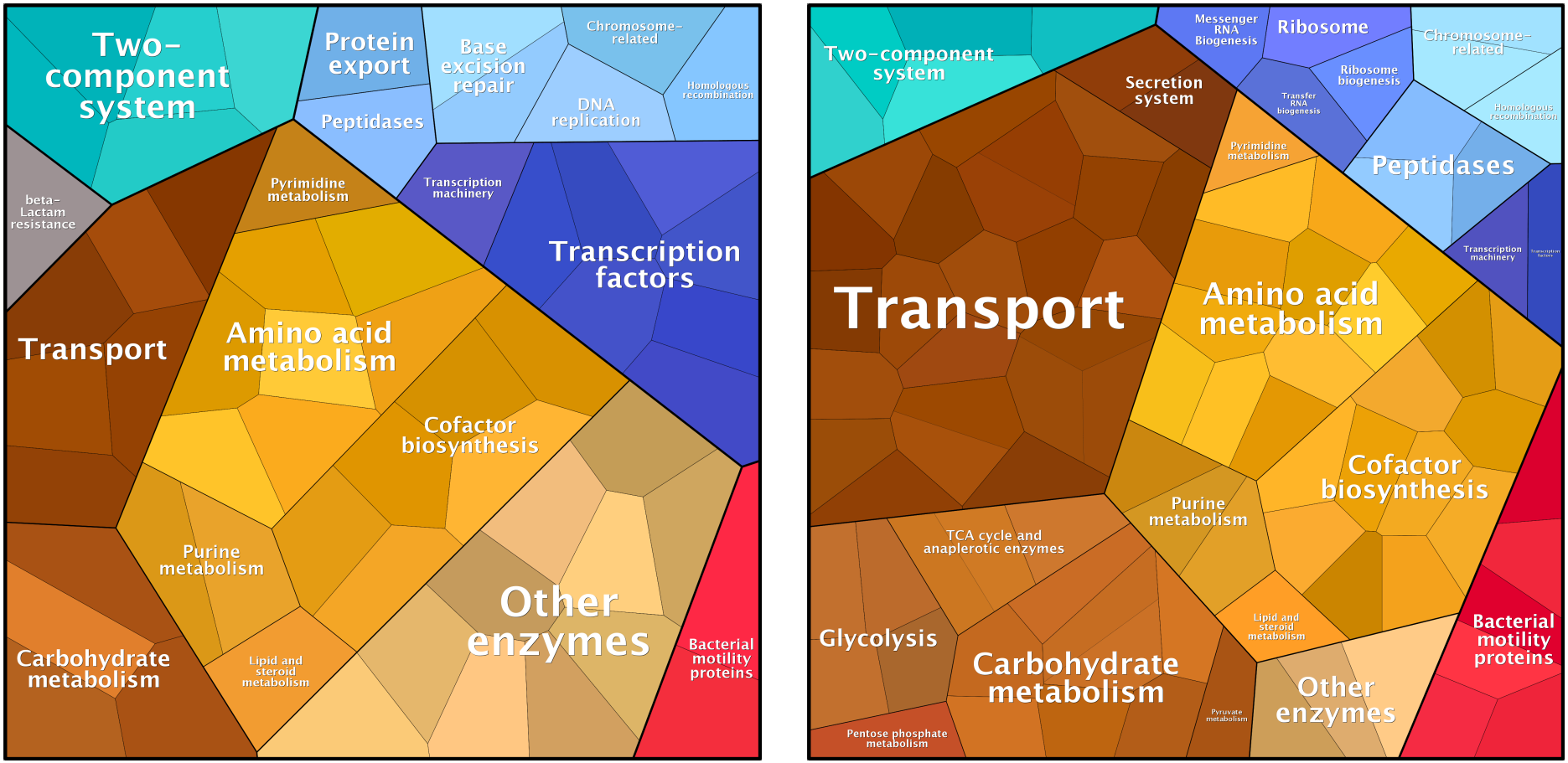


Figure 61 – Functional characterisation of the proteins unique to wild-type endospores (left) and ΔS357 (right) utilising Bionic Visualizations' Proteomaps. Each protein is signified as a polygon and the abundance of the protein determines the size. Differences can be seen for transcription factors, transport, carbohydrate metabolism and other enzymes.

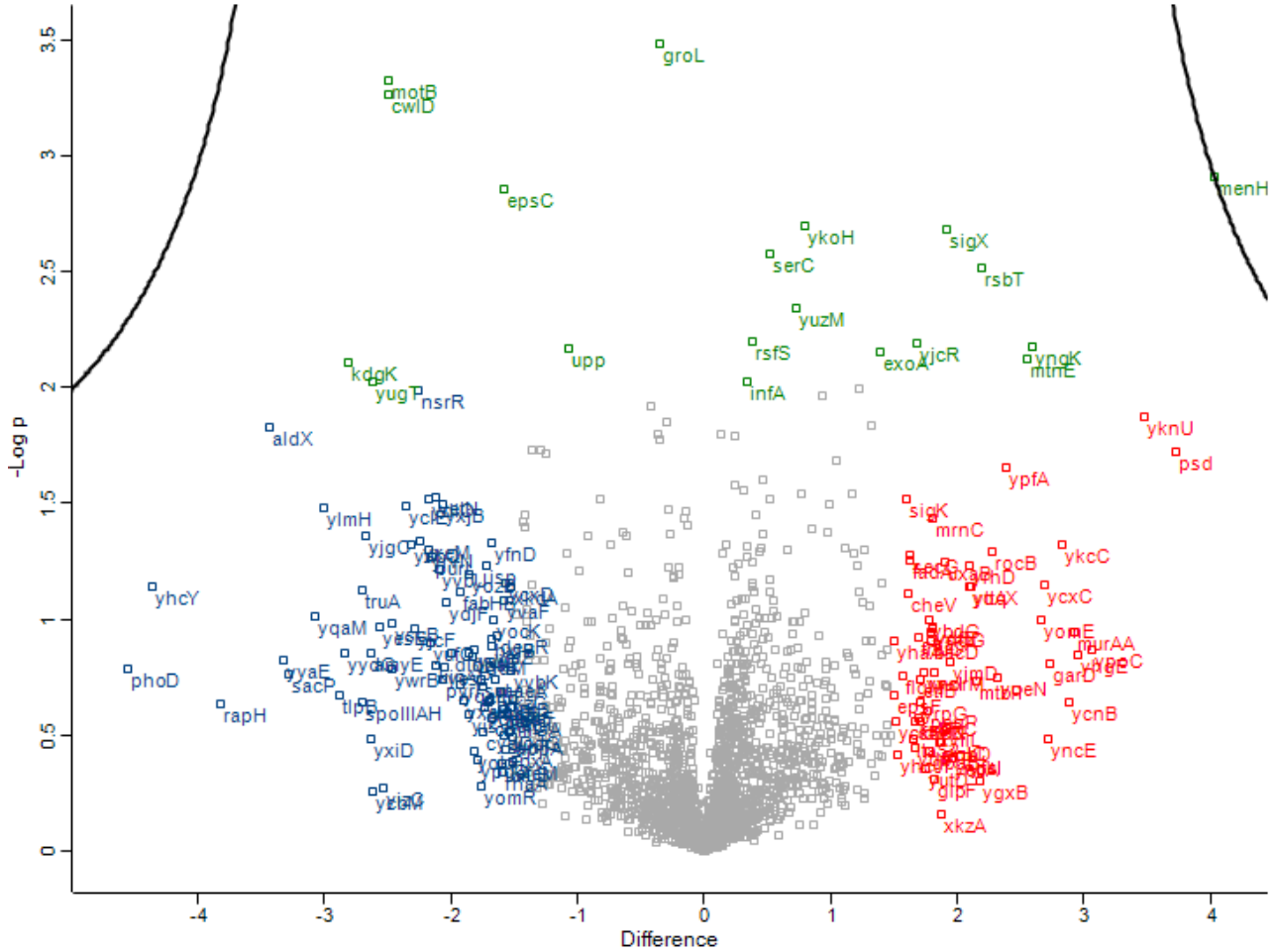


Figure 62 – Volcano plot Illustration of proteomic data using wild-type vs ΔS357 post imputation. Volcano plot was generated by comparing statistical significance of each identified protein with the change in prescence from WT. The majority of proteins showed no difference from WT. Coloured in green are proteins with higher statistical confidence (-Log q value of 2 or above) and those in red and blue are those which have high fold changes (> 1.5-fold change), but low confidence.

In addition to a general overview of the functional proteome, it is possible to look more in depth at individual proteins. As such, a differential analysis was performed (Figure 62). The proteins CwlD and MotB are the closest to being significant, but this is reduced post imputation and is not significant. Nonetheless, they are statistically confident of having a small fold change of increase in the sRNA mutant. There is a lack of significant findings in general – perhaps unsurprising due to the fine-tuning mechanisms sRNAs can play.

CwlD is an amidase involved in the synthesis of endospore cortex peptidoglycan (Gilmore, M. E. *et al.*, 2004). It is expressed during sporulation, under the control of SigE and SigG. Curiously, ΔS357 has a higher amount of CwlD. This perhaps indicates that the germination defect found in chapter four may be due to an alteration of the peptidoglycan composition.

MotB is a membrane bound motility protein that in complex with MotA is essential for motility (Chan, J. M. *et al.*, 2014). MotB uses an H^+ coupling ion to provide energy for motor force. The ion flow is yet to be determined, but it has been suggested that the MotB C-terminus is fixed to the peptidoglycan layer, causing the MotAB complex to open the proton channel (Takahashi, Y. *et al.*, 2014). Transporters have been found to be important in germination in *Bacillus* species. One of the earliest features in germination is the release of divalent cations, K^+, H^+ and Na^+, and this is hypothesized to be the result of transporters, such as the antiporter *grmA* in *B. megaterium* (Tani, K. *et al.*, 1996) and homolog *gerN* in *B. cereus*. This was found to interrupt normal germination to varying germinants (Southworth, T. W. *et al.*, 2001; Thackray, P. D. *et al.*, 2001). In spite of advances in other *Bacillus* species this mechanism in *B. subtilis* is yet to be determined (Paredes-Sabja, D. *et al.*, 2011). However, this study presents that ΔS357 was found to have more of MotB than wild-type and thus may mean another phenotype is yet to be discovered, other than the germination phenotype. MotA, the other protein in the same complex was found to be unchanged from wild-type endospores to ΔS357.

MenH is the closest to being significant post-imputation and was upregulated in the mutant. MenH is involved in the biosynthesis of menaquinone, directing the final methylation step which is essential in the citric acid cycle (Koike-Takeshita, A. *et al.*, 1997). MenH is required for oxidation of succinate to fumarate which is also involved in the electron transport chain. Δ*menH* has been shown to reduce sporulation efficiency to 12.7 of wild-type cells but have no cytological phenotype (Meeske, A. J. *et al.*, 2016). This is despite appearing to be down-regulated in sporulation as per the tiling array study (Nicolas, P. *et al.*, 2012). Electron transfer capacity during sporulation, dormancy and germination have been implicated to be related to menaquinone levels in *B. cereus* (Escamilla, J. E. *et al.*, 1988) and previously the *gerC* locus, of which has since been renamed, that contains a homolog of *menH* was reported to have a germination phenotype to alanine, but this result seems to be under contention as it may be an indirect effect (Leatherbarrow, A. J. *et al.*, 1998). MenH was not identified in ΔS357, but is present in wild-type endospores.

Additional proteins that fail to make significance, but are of interest due to being close to significance and a higher fold change than 1.5, are displayed in Table 15.

Table 15 – Additional proteins close to significance for ΔS357. Proteins selected based on a fold change of over 1.5 Expression during sporulation and germination indicated as per study by Nicolas, *et al.*

Proteins for which transcription levels are unchanged from basal levels of expression during sporulation and germination ×

Proteins which appear to be transcriptionally upregulated (↑) and down regulated (↓) during sporulation or germination.

Protein name	Function	Expression during sporulation	Expression during germination
KdgK	Utilisation of galacturonic acid	×	×
YugT	Metabolism/unknown	↑	×
RsbT	Control of SigB activity. Part of stressome	↓	×
YngK	Unknown	↑	×
MtnE	Methionine salvage	↓	↓

5.2.5.2 Wild-type vs ΔS547

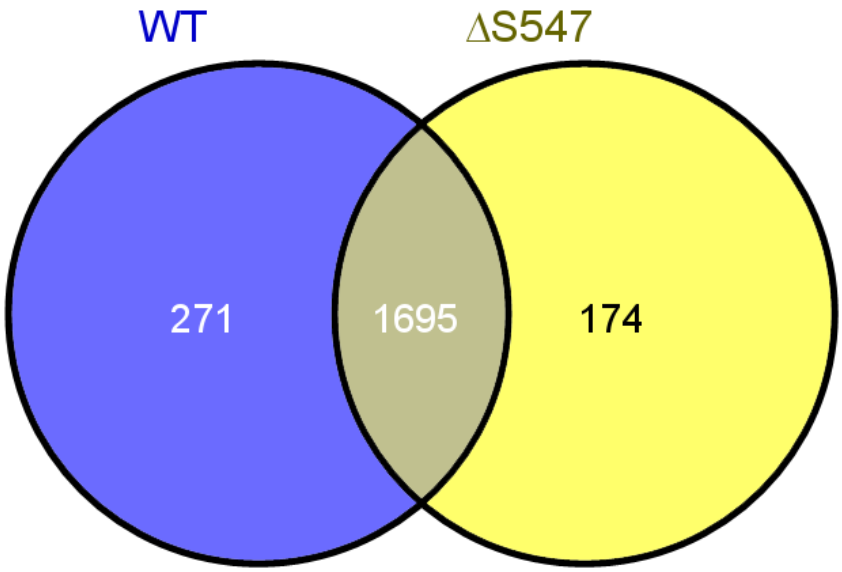


Figure 63 – Comparison of proteins that are present in at least two replicates in the wild-type and in ΔS547 endospores.

Wild-type and ΔS547 endospores share 1695 proteins (Figure 63). The wild-type endospores have a total of 271 unique proteins, with 174 proteins that are only present in those endospores lacking S547. When all strains are considered, this total is reduced to 52 proteins (Figure 59). Wild-type and ΔS547 endospores have a similar spread of functional categories that are unique to each stain and as such are considered insignificant, apart from other enzymes and transcription factors which have no obvious connection to previously identified phenotypes (Figure 64).

Upon differential analysis, only one protein can be considered significantly different, Upp (Figure 65). Upp is a uracil phosphoribosyltransferase, one of two in *B. subtilis* that is involved in uracil monophosphate (UMP) synthesis (Martinussen, J. *et al.*, 1995). Upp is downregulated in sporulation conditions (Nicolas, P. *et al.*, 2012). The *upp* gene is a part of the deletion cassette that was used to create ΔS547 and therefore is likely to be upregulated in ΔS547 for this reason. In green are those proteins with high statistical confidence, albeit not significant after correction of false discovery rate. CwlD is differentially expressed as seen in the ΔS357 study but not as strongly, as is MotB. No clear indication of cause of germination deficiency could be found in other genes close to statistical significance (Table 16).

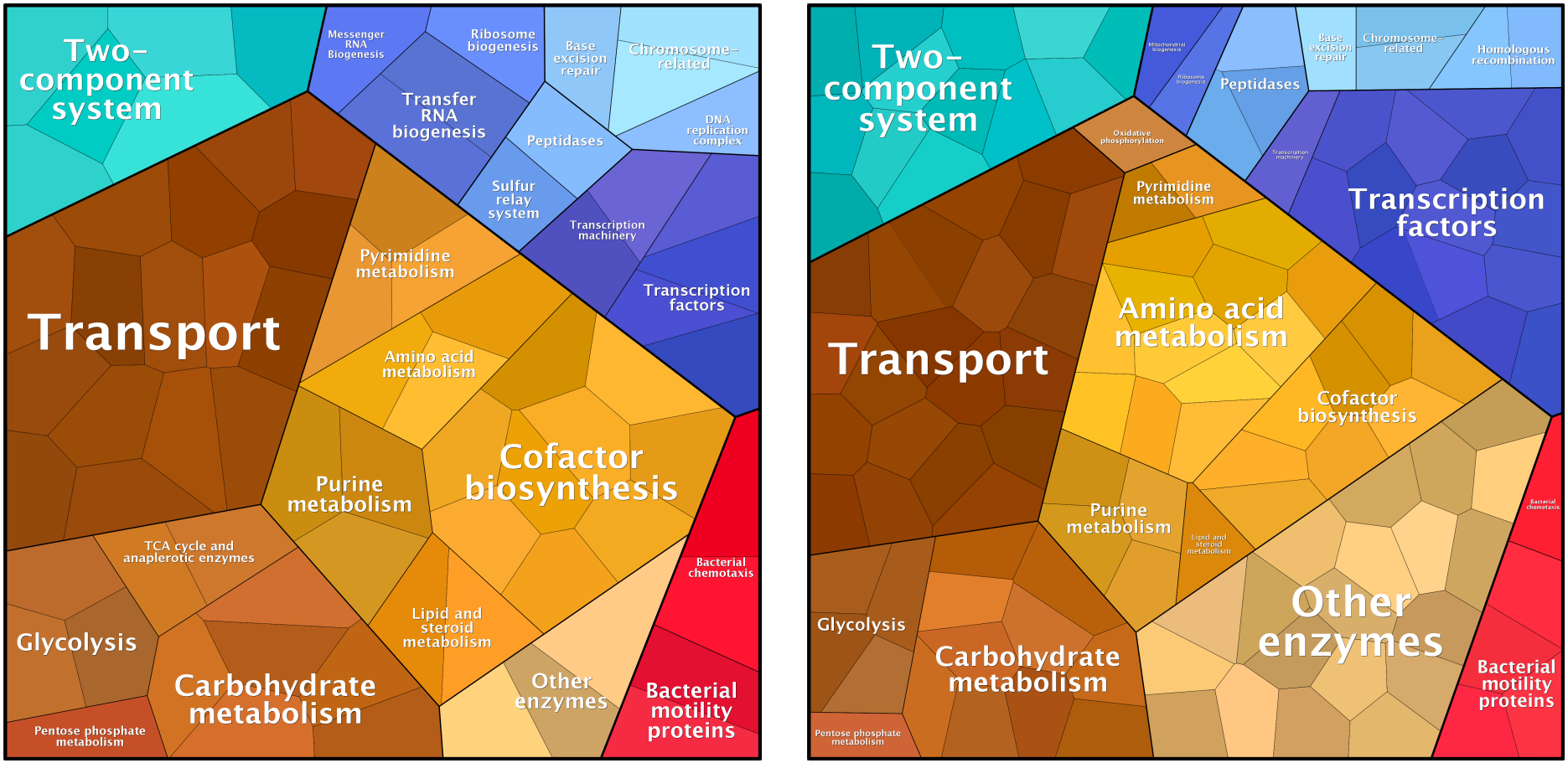


Figure 64 – Functional characterisation of the proteins unique to the wild-type (B) and ΔS547 (A) strains using Bionic Visualizations' Proteomaps. Each protein is signified as a polygon and the abundance of the protein determines the size. Reduction in transcription factors and other enzymes categories can be seen for ΔS547.

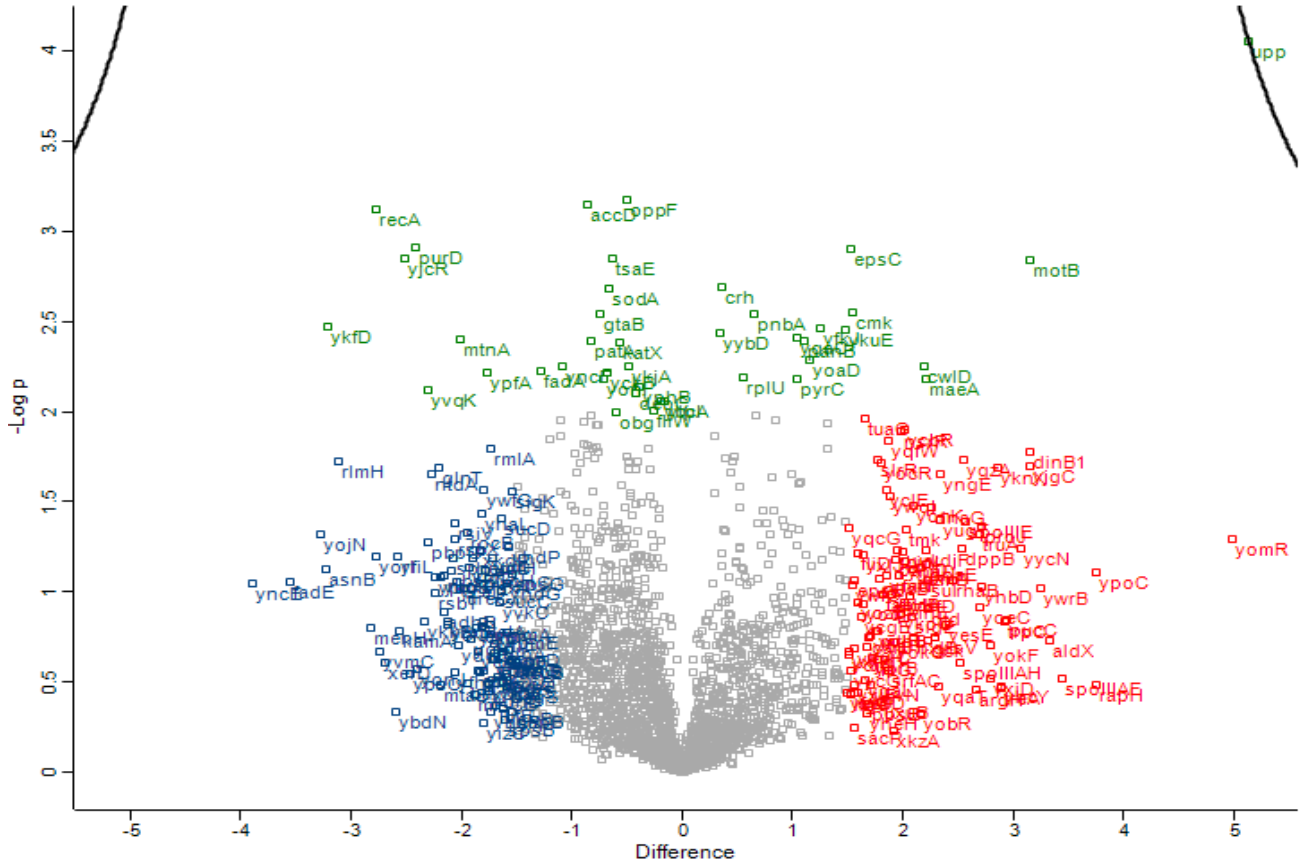


Figure 65 – Volcano plot illustration of proteomic data of wild-type protein vs ΔS547 post imputation. Volcano plot was generated by comparing statistical significance of each identified protein with the change in prescence from WT. Coloured in green are proteins with higher statistical confidence (-Log q value of 2 or above) and those in red and blue are those which have high fold changes (> 1.5-fold change), but low confidence.

Table 16 – Additional genes close to significance for ΔS547 with a higher fold change than 1.5

Genes which are unchanged from basal levels of expression ×

Genes which appear to be actively upregulated (↑) and down regulated (↓).

Gene name	function	Expression during sporulation	Expression during germination
recA	DNA recombination/repair	↓	×
purD	Purine biosynthesis	↓	↓
yjcR	Unknown	↑	×
ykfD	Similar to oligopeptide ABC transporter	↑	×
yvqK	Unknown	×	×
cwlD	Endospore cortex synthesis	↑	×
maeA	Malate utilization	×	×
motB	Motility protein	×	×

5.2.5.3 Wild-type vs ΔS612

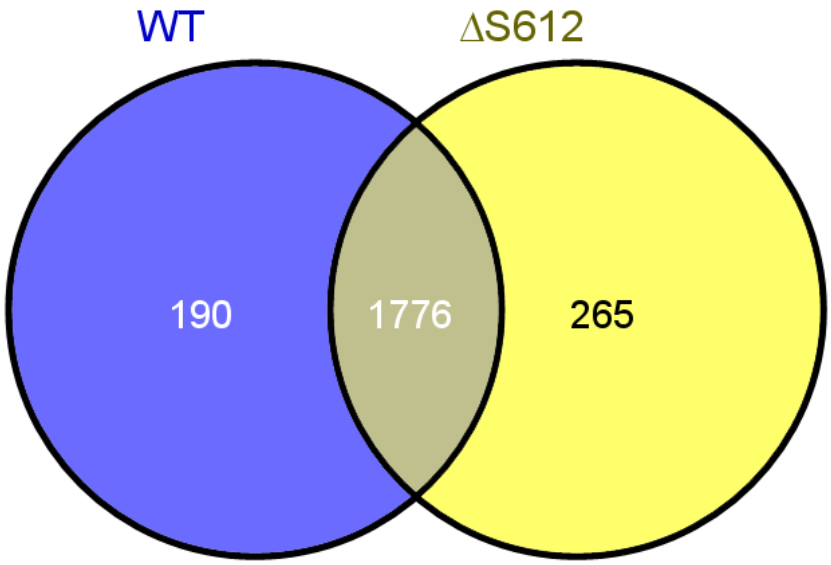


Figure 66 – Comparison of proteins that are present in at least two replicates in the wild-type and in ΔS612 endospores.

There are 190 proteins that can be considered unique to wild-type endospores and 265 that are unique to endospores of ΔS612 (Figure 66). Functional characterisation utilising Bionic visualisations (Figure 67) however, does not reveal a particularly strong pattern of differences to this other than with putative enzymes of unknown function. Differential protein analysis has displayed no significantly changed protein intensities (Figure 68).

Closest to being statistically significant is YoaD which makes an unknown product. YoaD is hypothesised to be involved in methionine biosynthesis and is believed to likely be a phosphoglycerate dehydrogenase. When methionine starvation is experienced the S-box riboswitch system, known to be involved with biosynthesis of methionine, can change conformation relieving the anti-antiterminator present (Grundy, F. J. & Henkin, T. M., 1998).

In addition, whilst not statistically significant, RapH is totally absent in wild-type endospores and highly present in ΔS612 endospores (T-test difference of 6.09). RapH is involved in the initiation of sporulation; specifically, it dephosphorylates Spo0F and also inhibits the binding of the competence regulator, ComA, to its DNA targets

(Smits, W. K. *et al.*, 2007). *rapH* transcription is active during early sporulation but is strongly downregulated during the final hours (Nicolas, P. *et al.*, 2012). Other differentially regulated genes not quite statistically significant can also not be linked to phenotypic changes found previously (Table 17).

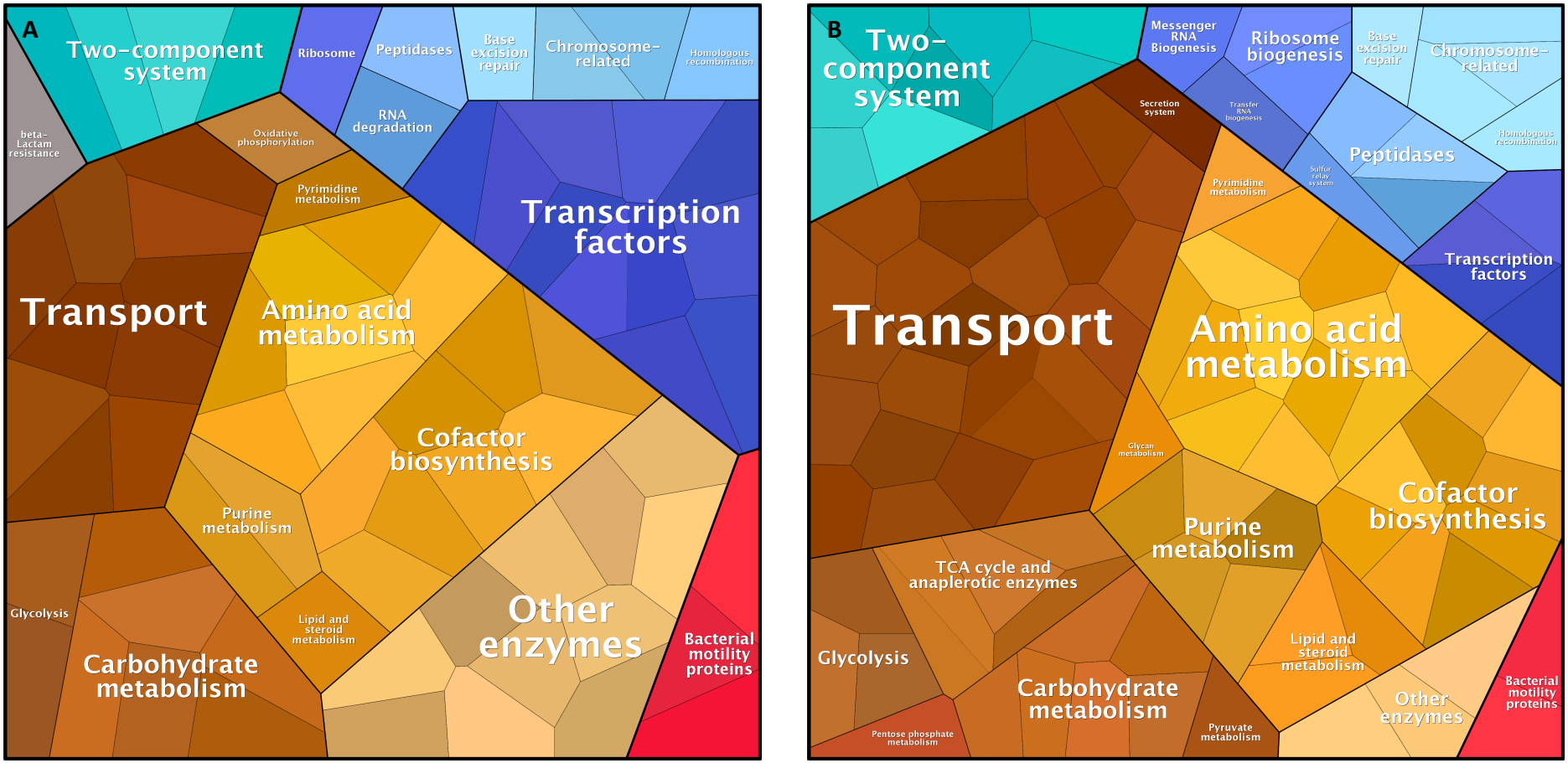


Figure 67 – Functional characterisation of the proteins unique to (A) wild-type and (B) ΔS612 strains utilising Bionic Visualizations' Proteomaps. Each protein is signified as a polygon and the abundance of the protein determines the size.

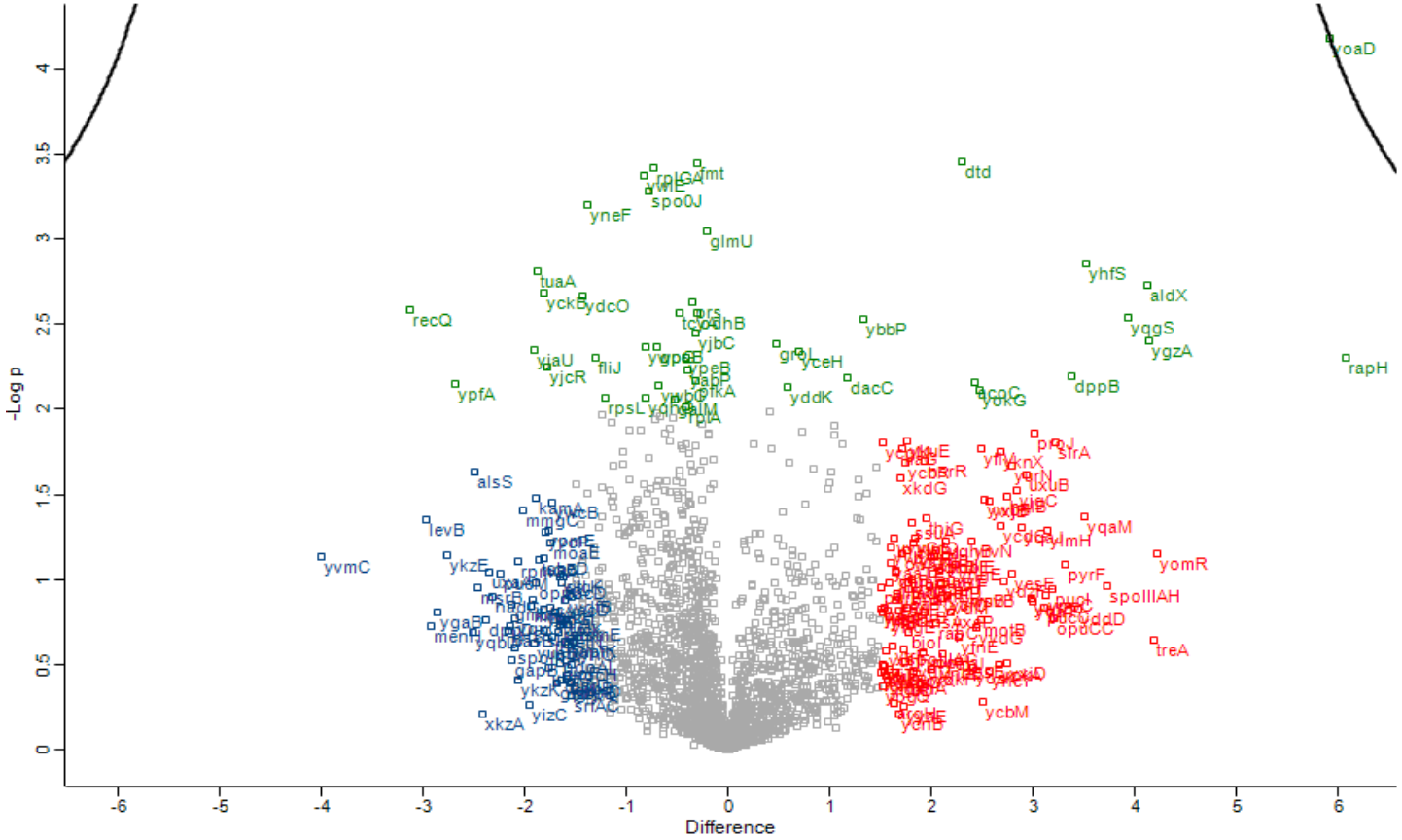


Figure 68 – Volcano plot illustration of proteomic data of wild-type protein vs ΔS612 post imputation. Volcano plot was generated by comparing statistical significance of each identified protein with the change in prescence from WT. Coloured in green are proteins with higher statistical confidence (-Log q value of 2 or above) and those in red and blue are those which have high fold changes (> 1.5-fold change), but low confidence.

Table 17 – additional genes close to significance for ΔS612 with a fold change higher than 1.5

Genes which are unchanged from basal levels of expression ×

Genes which appear to be actively upregulated (↑) and down regulated (↓).

Gene name	Function	Expression during sporulation	Expression during germination
recQ	ATP-dependent DNA helicase	×	×
ypfA	Control of MotA activity	×	×
dtd	Prevention of misincorporation of D-amino acids	↓	×
yhfS	Unknown	↓	↓
aldX	Aldehyde dehydrogenase	×	×
yqgS	Biosynthesis of lipoteichtoic acid	×	×
ygzA	Unknown	↑	↑
rapH	Control of sporulation initiation	×	↑
dppB	Uptake of dipeptides	↓	↑
yokG	Unknown	↓	↓
acoC	Acetoin utilization	×	↑

5.2.6 Other methods for identifying targets of an sRNA

5.2.6.1 Target predictions *in silico*

Target predictions of sRNAs can be used as a tool to guide the correct experimental conditions to study and have been previously successfully utilised (Tjaden, B. *et al.*, 2006; Wright, P. R. *et al.*, 2013; Sharma, C. M. *et al.*, 2011). However, target predictions must be treated with caution as large numbers of false-positives are typically reported. As such, two methods of target prediction will be tested here. The Target prediction was performed utilising output from CopraRNA and this is reduced to putative targets that were unique to one sample.

5.2.6.1.1 CopraRNA vs unique proteins

This study asked if the target predictions can be used in conjunction with our proteomics data. CopraRNA, a tool from Freiburg RNA tools, was chosen to predict sRNA targets (Wright, P. R. *et al.*, 2014; Wright, P. R. *et al.*, 2013). CopraRNA utilises homologous sRNA sequences to compute whole genome target predictions from each species to determine if a given homolog is present in to enrich for conservation. 75 nucleotides upstream and downstream of the start codon of potential targets was used, as many sRNAs illicit their functions post-transcriptionally via binding around the Shine Dalgarno sequence and other regulatory sites. CopraRNA produces a summary table of the 100 top hits that was computed to be the most likely targets from a given sequence. A summary table for each S357, S547 and S612 (Table 18, Table 19 and Table 20, respectively) was produced which has been reduced from the 100 top hits CopraRNA produced to those that are uniquely present in the proteomics study between the mutant and wild-type.

Three targets can be predicted for S357 that are absent in ΔS357 proteome and present in the wild-type, suggesting that S357 may have a role in stabilising these transcripts. A further nine are predicted the other way around, being present in ΔS357 and absent in wild-type, suggesting S357 may destabilise or promote the degradation of these target transcripts. Many genes have no obvious association with the known phenotypes of S357 set out in the previous chapters, or have no

obvious link to germination and/or sporulation. Discussed below are targets that may be linked with germination and/or sporulation.

pucI is a putative target transcript that is present in the mutant proteome, but not in the wild-type, an allantoin transport protein (Ma, P. *et al.*, 2016). Interestingly, *pucI* was found to be strongly induced during germination under osmotic stress (Nagler, K. *et al.*, 2016). *pucI* is upregulated in the later stages of sporulation (Nicolas, P. *et al.*, 2012) but is induced as germination continues regardless of osmotic stress (from the data reanalysis in chapter one, (Nagler, K. *et al.*, 2016)).

yttP encodes a DNA binding protein and is a regulator of *ftsZ*, which mediates the switch from medial to polar FtsZ ring placement during sporulation (Wagner-Herman, J. K. *et al.*, 2012). In addition to this, YttP binds motifs close to the *oriC* locus on the chromosome to determine the precise part of the chromosome that is initially captured in the forespore (Miller, A. K. *et al.*, 2016). *yttP* is expressed highly throughout sporulation (Nicolas, P. *et al.*, 2012).

spsD is a mother cell specific protein thought to be involved in the endospore surface synthesis. The *sps* operon has previously been shown to have a reduced germination efficacy in comparison to wild-type (Cangiano, G. *et al.*, 2014), like that of ΔS357 in previous chapters.

yuiC is a sporulation protein or a stationary phase survival protein of unknown function (Quay, D. H. *et al.*, 2015). *yuiC* is expressed in the forespore early in sporulation (Wang, S. T. *et al.*, 2006).

Table 19 displays 12 putative targets for S547 based on the presence and absence of proteins in a dormant endospore. Interestingly, the mother cell-specific RNA polymerase sigma factor, name, was predicted to be a putative target, which directs gene expression early during sporulation. In addition, *yngE* is expressed early in the mother cell (SigE regulated) and is linked with metabolism, specifically leucine

utilisation (Steil, L. *et al.*, 2005). Of the putative targets that are also identified in the proteomics study, most have no clear link to the phenotypes identified previously. This can also be said for that of S612 (Table 20). The putative targets that can be identified in the proteomics study are not obviously linked to the phenotypic studies previously identified.

Table 18 – Target prediction of S357 using CopraRNA.

Rank	CopraRNA p-value	Gene Name	Energy kcal/mol	IntaRNA p-value	Position mRNA	Position ncRNA	Annotation	Position Seed – mRNA	Position Seed – ncRNA	Hybridization Energy kcal/mol	Present in S357?	Present in WT?	Altered expression during Sporulation (Nicolas, P. *et al.*, 2012)	Altered expression during Germination (Nagler, K. *et al.*, 2016)
27	0.01	*pucI*	-9.64	0.06	130 -- 139	221 -- 230	Allantoin permease	133 -- 139	221 -- 227	-17.4	✓	×		
29	0.01	*yttP*	-7.99	0.15	82 -- 90	213 -- 221	TetR family transcriptional regulator	84 -- 90	213 -- 219	-16.3	✓	×		
35	0.01	*yjgC*	-11.9	0.02	61 -- 73	98 -- 109	Oxidoreductase. Formate dehydrogenase – ethanol and salt stress (Hoper, D. *et al.*, 2005)	67 -- 73	98 -- 104	-14.8	✓	×	🡻	
39	0.01	*sacA*	-8.36	0.12	114 -- 124	98 -- 110	Sucrose-6-phosphate hydrolase (Fouet, A. *et al.*, 1986)	118 -- 124	98 -- 104	-11.3	×	✓	🡻	🡹 (Nicolas, P. *et al.*, 2012)
47	0.01	*yclK*	-5.82	0.41	56 -- 69	97 -- 111	Sensor histidine kinase induced in oxygen limitation(Hartig, E. *et al.*, 2004)	57 -- 63	104 -- 110	-14.8	✓	×		
66	0.02	*yabB*	-9.9	0.05	111 -- 129	81 -- 97	Unknown	123 -- 129	81 -- 87	-18.4	×	✓	🡻	

73	0.02	*spsD*	-5.56	0.45	143 -- 150	84 -- 91	Endospore coat polysaccharide biosynthesis protein SpsD	144 -- 150	84 -- 90	-9.6	✓	×		
82	0.02	*yuiC*	-5.94	0.39	107 -- 124	165 -- 181	Evidence 4: Homologs of previously reported genes of unknown function	118 -- 124	165 -- 171	-15	✓	×		
83	0.02	*ispD*	-9.35	0.07	67 -- 94	218 -- 244	2-D-methyl-D-erythritol 4-phosphate cytidylyltransferase. essential gene involved in the MEP pathway of isoprenoid biosynthesis	67 -- 73	238 -- 244	-22	✓	×	↑	↑ after 30 mins
84	0.03	*moaE*	-10.24	0.04	122 -- 149	91 -- 122	Molybdopterin synthase catalytic subunit (Rudolph, M. J. *et al.*, 2003)	135 -- 141	101 -- 107	-23.3	×	✓	↓	
92	0.03	*fliF*	-6.51	0.31	40 -- 78	81 -- 118	Flagellar M-ring protein. Flagella basal-body membrane anchor (Guttenplan, S. B. *et al.*, 2013)	41 -- 47	111 -- 117	-26.3	✓	×	↓	↓ under osmotic stress
98	0.03	*comGA*	-5.64	0.44	93 -- 100	215 -- 222	ComG operon protein 1. traffic ATPase which binds and transports transforming DNA during competence (Chung, Y. S. & Dubnau, D., 1998)	94 -- 100	215 -- 221	-13.4	✓	×	↑	

Table 19 – Target prediction of S547 using CopraRNA

Rank	CopraRNA p-value	Gene Name	Energy kcal/mol	IntaRNA p-value	Position mRNA	Position ncRNA	Annotation	Position Seed – mRNA	Position Seed – ncRNA	Hybridization Energy kcal/mol	Present in S547	Present in WT	Altered expression during Sporulation (Nicolas, P. *et al.*, 2012)	Altered expression during Germination (Nagler, K. *et al.*, 2016)
15	0.01	*rapC*	-12.27	0.06	60 - - 68	77 - - 85	Control of ComA activity, a response transcriptional regulator for genetic competence, RapC interacts with ComA to inhibit its DNA binding (Core, L. & Perego, M., 2003)	62 - - 68	77 - - 83	-17.7	×	✓	↑	
64	0.03	*gudB*	-11.12	0.10	31 - - 50	72 - - 88	Cryptic catabolic NAD-specific glutamate dehydrogenase for the utilization of glutamate as a sole carbon source (Commichau, F. M. *et al.*, 2008)	44 - - 50	72 - - 78	-21.6	×	✓	↓	↑ after an hour
67	0.03	*ymfC*	-11.59	0.08	68 - - 75	48 - - 55	GntR family transcriptional regulator	69 - - 75	48 - - 54	-16.1	×	✓		
85	0.03	*ruvA*	-6.04	0.47	90 - - 97	82 - - 89	Holliday junction ATP-dependent DNA helicase (Canas, C. *et al.*, 2014)	91 - - 97	82 - - 88	-9.1	×	✓	↓	
91	0.03	*yndG*	-16.28	0.01	60 - - 73	76 - - 88	Membrane protein	62 - - 68	80 - - 86	-19.5	×	✓	↓	

37	0.02	*pyrC*	-15.05	0.02	60 - - 70	72 - - 82	Dihydroorotase	64 - - 70	72 - - 78	-18.3	✓	×	↓	
59	0.02	*sigE*	-17.4	0.01	62 - - 73	77 - - 88	RNA polymerase sigma-E factor	67 - - 73	77 - - 83	-22.4	✓	×		
66	0.03	*sul*	-8.26	0.26	61 - - 68	81 - - 88	Dihydropteroate synthase	62 - - 68	81 - - 87	-11.5	✓	×	Slightly ↓	↑
71	0.03	*yngE*	-10.05	0.14	65 - - 72	81 - - 88	Carboxylase YngE	66 - - 72	81 - - 87	-12.8	✓	×		

Table 20 – CopraRNA target prediction for S612

Rank	CopraRNA p-value	Gene Name	Energy kcal/mol	IntaRNA p-value	Position mRNA	Position ncRNA	Annotation	Position Seed – mRNA	Position Seed – ncRNA	Hybridization Energy kcal/mol	S612	WT	Altered expression during Sporulation (Nicolas. P. *et al.*,	Altered expression during Germination
21	0.00	*ycxC*	-9.15	0.02	7 -- 21	1 -- 16	Transporter	11 -- 17	6 -- 12	-18.9	×	✓	↑	
33	0.01	*trpE*	-7.04	0.10	29 -- 40	1 -- 10	Biosynthesis of tryptophan (Gollnick, P. *et al.*, 2005)	34 -- 40	1 -- 7	-12.7	×	✓	↓	↑
38	0.01	*ylmD*	-10.48	0.01	34 -- 75	42 -- 81	Peptidoglycan editing factor (Parveen, S. & Reddy, M., 2017)	69 -- 75	42 -- 48	-31.3	×	✓	↓	↑
92	0.02	*ycbP*	-3.98	0.54	80 -- 86	3 -- 9	Unknown general stress protein (Hoper, D. *et al.*, 2005)	80 -- 86	3 -- 9	-10.4	×	✓	↑	↑
16	0.00	*amyE*	-11.38	0.00	122 -- 148	36 -- 62	Alpha-amylase – Starch degradation (Gupta, M. & Rao, K. K., 2014)	124 -- 130	54 -- 60	-21.2	✓	×	↑	↑
35	0.01	*rapH*	-8.05	0.05	129 -- 139	58 -- 68	Response regulator aspartate phosphatase H. Control of sporulation initiation and ComA activity (Smits, W. K. *et al.*, 2007)	133 -- 139	58 -- 64	-12.2	✓	×	↑	↑
55	0.01	*yycS*	-9.57	0.01	51 -- 77	69 -- 99	Unknown	71 -- 77	69 -- 75	-21.7	✓	×	↑	

5.2.6.2 First steps for the future: RNA-RNA interactome

In vivo high-throughput methods have been investigated recently for the identification of the direct targets of sRNAs. These methods revolve around RNA-RNA interaction studies in which an sRNA is cross-linked with its target and followed by identification of the linked RNA. One molecule that has been used for this is 4'-Aminomethyltrioxsalen hydrochloride (AMT). AMT is a psoralen that can intercalate into double-stranded RNA and, upon introduction of UV light at a long wavelength (365 nm), can introduce covalent bonds between pyrimidines. This crosslinking is reversible by exposure to short-wavelength UV (254 nm) to allow for processing. As such, psoralen derivatives such as AMT are useful molecules in the study of RNA-RNA interactions (Wassarman, D. A., 1993).

One of the first methods to directly identify sRNA-mRNA interactions was RNA walk (Lustig, Y. *et al.*, 2010). RNA walk utilises AMT to link RNAs together and this was subsequently mapped by RT-PCR. Using RNA walk, interactions between RNA of trypanosome signal recognition particle (SRP) and the ribosome were catalogued, identifying four regions of interaction. This method relies on affinity selection, whereby the sRNA was used to specify capture and thus the sRNA interaction itself may preclude it from such. In this case, the target would have to be used and thus prevents RNA walk from being a global method for unbiased interaction studies.

Two similar methods utilising AMT allow high-throughput global captures of sRNA-mRNA interactions and both utilise proteins as scaffolds. The first, termed RIL-seq (RNA interaction by ligation and sequencing), identified Hfq-bound sRNAs and their targets (Melamed, S. *et al.*, 2016). This method uses protein-RNA crosslinking *in-vivo*, in which cell lysates are then co-immunoprecipitated utilising FLAG-tagged Hfq. The RNA not crosslinked to Hfq is then digested by RNases and the resulting ends are ligated together before Hfq is digested with proteinase K. This results in a long RNA fragment comprised of the two RNA molecules which were interacting with the Hfq protein. The RNA is then isolated and sequenced. Computationally, this is then separated into chimeric sequences, where the two RNA molecules that have been ligated together are two separate RNA species, and single fragments, in which the

RNA sequenced is a single RNA species. The likelihood of a RNA being part of a chimeric fragment is then computed, deciding if a RNA pair is overrepresented than would be expected by chance. Secondly, CLASH was implemented utilising RNase E as its scaffold protein (Waters, S. A. *et al.*, 2017). RNA and FLAG-tagged RNase E is similarly UV crosslinked together and co-immunoprecipitated (co-IP). The RNA is again trimmed via RNase digestion. The RNA is then ligated and linkers are added to be compatible with further downstream sequencing. RNA-RNase E complexes are then size selected using SDS-PAGE. The RNA is then recovered and again subjected to high throughput sequencing. Both methods were successful in identifying numerous sRNA-mRNA interactions; in log phase cells 633 were found with RIL-seq and 782 utilising CLASH. In addition, the seed region of well characterised sRNAs between the two studies was found to be similar, indicating both methods successfully identified real interacting regions.

Whilst cross-linking and sequencing of hybrid RNA-RNA interaction molecules has been successful, both rely on specific proteins to facilitate capture and therefore are unable to secure the entirety of RNA-RNA interactomes. Several studies in eukaryotes and one in bacteria have since been published that do not use a protein to isolate interacting regions. Notably is the study by Sharma, E. *et al.* (2016) in which LIGation of interacting RNA and high-throughput sequencing (LIGR-seq) was used. This was carried out in eukaryotic cells, where AMT was used to cross-link interacting RNA together. This was then purified and RNase treated. The resulting interacting RNA was then ligated together, uncross-linked and sequenced. This study successfully identified many non-coding RNA and mRNA interactions.

In bacteria, a modified CLASH protocol was performed on *E. coli* to similarly capture RNA-RNA interactions independent of protein associations (Liu, T. *et al.*, 2017). In this protocol, AMT was again utilised to cross-link RNA molecules together. RNA was extracted, RNase digestion was performed and DNA was depleted. Then 20 nucleotide oligo-deoxy-ribonucleotides were added and further digested by RNase H. Size selection takes place next, where RNAs that are between 40 and 100 nucleotides were recovered to then be ligated. Cross-linking was then reversed and

the resulting RNA was sequenced. rRNA was not depleted, and as such many of the interacting RNAs involved rRNAs. In addition, no known sRNA-mRNA interactions were found and therefore this modified CLASH protocol is not sufficient as it stands for sRNA-mRNA interaction studies.

RNA interactome studies could be particularly advantageous in heterogeneous populations such as sporulation. Previous attempts to use RNAseq during sporulation has proven problematic, with sequenced libraries having large variation between replicates due to the noise of the population (Eijlander, RT *et al.*, 2016). The protocols for RNA interactome studies involve degrading any RNA molecules that are not in a duplex and therefore RNA from dead cells would be removed.

As such, a preliminary study here focused on capturing the RNA-RNA interactome to identify sRNA-mRNA targets during sporulation in *B. subtilis*, utilising a protein-free capture method developed by Sharma *et al.* to ensure that AMT could successfully cross the mother-cell into the forespore.

5.2.6.2.1 AMT crosslinking enables RNA interactions to be identified

AMT was used to crosslink RNAs at four hours after the initiation of sporulation. The RNA was processed to ligate interacting RNAs which were then sequenced. The data analysis of the reads resulting from this experiment is not yet fully robust and this initial screen was not in enough depth for thorough analysis. However, reads were analysed utilising the scripts available from Melamed, S. *et al.* (2016) to see if this method could potentially be used in the future after further optimisation. Briefly, reads were mapped to the *B. subtilis* genome and the resulting bam file was used to probe 25 nucleotides at each end of each read. These were then used to map back to the genome. If both the start and end of the read was mapped to the same gene this was classified as a single fragment, and if they were mapped to different genes the read was classified as a chimera.

One test sample taken at hour 4 of sporulation resulted in 2,620 chimeras post tRNA-depletion with AMT added, and 2,229 were found in the absence AMT (a sample

used as a control to show chimera enrichment with the presence of AMT). Of those reads that could be assigned a SubtiWiki class, 17.37% of + AMT interactions and 14.15% -AMT interactions are allocated to the sporulation class. In addition, reads were successfully mapped to many S segments, several of which mapped to S357. When mapped to the genome in Artemis, it is clear that there could be a hotspot of chimeric reads within S357 with the addition of AMT (Figure 69). Red indicates the reads from the control sample, without the addition of AMT, and blue is the test condition with AMT. The example with S357 shows a clear clustering of reads around the same location, indicating a potential location of preferential binding. In addition, this location corresponds to regions of conservation identified in chapter three.

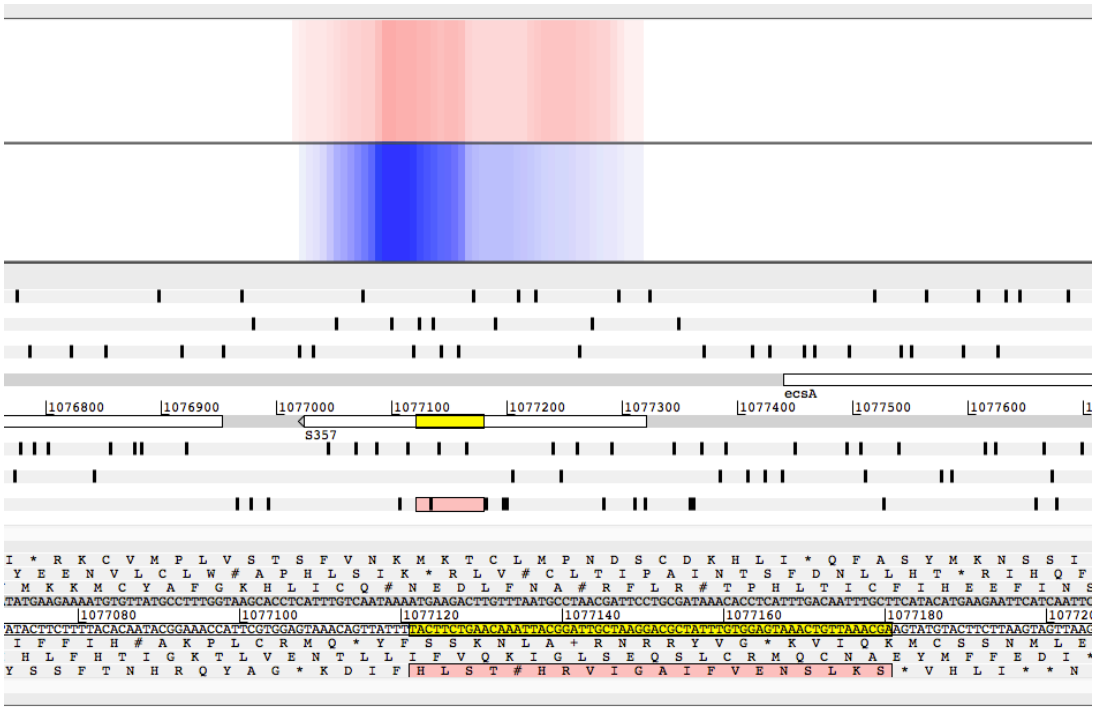


Figure 69 – Heat map of reads mapping to S357. Red is in the absence of AMT and blue is in the presence of AMT. Darker colours represent a high number of reads while fainter colours show lower numbers.

Only one potential target for S357 was identified, most likely since the samples were not sequenced with an adequate depth. This target was *spoVID* which encodes for a protein which is important for the assembly of the endospore coat (Costa, T. *et al.*, 2006). This is tenuous due to the lack of chimeric reads available. Targets for S547

and S612 were not found with more than one chimeric sequence available which is likely again due to the same lacking of depth of coverage.

5.3 Discussion

RNAseq has become an informative tool in biological studies that is used in the pursuit of understanding regulatory variation. However, the quality of the starting material to be studied is vital as methods of detection are becoming more sensitive. As the process of sporulation is heterogeneous, the RNA pool to be recovered from a given time will be of questionable quality (Figure 47) and therefore is not ideal in the pursuit of understanding gene expression on a fine level. Nevertheless, techniques are being developed that may be of use in the future.

Full characterisation of endospores has long been attempted (Goldman, R. C. & Tipper, D. J., 1978). Bacterial endospores are of particular interest as understanding their composition is a vital step in understanding the processes of sporulation and germination, and is informative for the improvement of sterilisation techniques as pathogenic species are of concern to industry. The comprehensive list of proteins recovered during sporulation that has been achieved by this study provides a marked improvement on what has been recovered previously. Whole endospore proteomics before this study identified protein levels in the early hundreds in comparison to the thousands recovered here (Kuwana, R. *et al.*, 2002).

Whilst the method of protein extraction from endospores developed here is a great tool for understanding the composition of a dormant endospore, it gave no results that showed statistically significant differences between the wild-type and sRNA deletion mutants and thus it is unsurprising that it was not able to identify discrepancies which could explain earlier phenotypes found in chapter four upon deletion of an sRNA. A multitude of explanations for this could be likely. Firstly, this could suggest that the deficiency leading to the phenotypic changes identified previously isn't in the proteins present, but some other factor such as the RNA content. Alternatively, the defect caused is small and therefore this method is not

sensitive enough as sporulation is a heterogeneous process and more replicates may be needed to pull out what will end up being statistically significant changes that may then correlate up to phenotypes. A significant alteration of protein composition could also be missing due to redundancy (as potentially shown in figures 14, 17, and 20), highlighting the importance of identifying a specific interacting partner. sRNAs are often referred to be fine-tuning mechanisms in gene regulation (Lenz, D. H. *et al.*, 2004; Mars, R. A. *et al.*, 2015b) and thus may require studies to be more targeted to deduce the specific role that they play, rather than using global transcriptomic/proteomic approaches. In addition, there are many steps that could happen in between translation and producing an active product, such as post-translational modifications and correct localisation. This study did not separate out the individual endospore structures and as such incorrect localisation could not be identified nor was methylation of proteins studied. This study did not investigate these and it remains to be seen whether the sRNAs studied here have roles in these events.

The proteomics data may not have given any statistically significant results due to the condition used. End-point endospores were used for all samples, done so as the mutants displayed germination defects. There is evidence that the proteins needed for germination (such as certain nutrient receptors etc.) are built into the dormant endospore in the sporulation process. However, perhaps proteomics in germination conditions may hold the key to sRNA targets/functions.

Whilst identification of sRNAs has been facilitated by differential gene expression approaches, identification of their targets and characterising their interactions has been more challenging. One method of predicting targets of an sRNA is computationally. This approach is typically based on a sequence matching approach at the points where the sRNA binds with its targets to illicit its effects. This, however, is hampered by the imperfect nature of binding, often leading to many false-positives. Algorithms have become more sophisticated, incorporating other elements such as conservation of a sequence (Kery, M. B. *et al.*, 2014; Wright, P. R. *et al.*, 2014; Wright, P. R. *et al.*, 2013). Whilst methods have improved, it is still

plagued with too many false-positives to warrant experimental study of each predicted target.

Target prediction is now superseded by efforts to directly identify RNA-RNA interactomes *in vivo*. Whilst RNAseq was not viable, this study tested if this method could produce suitable reads in an initial experiment that would then warrant a more in-depth study. The first question that needed to be answered relates to the ability of AMT to enter the forespore compartment, and secondly the quality of RNA had to be sufficiently high to give intact reads. The first concerns appear to be unwarranted for studies on or before the 4th hour of sporulation as both mother cell and forespore transcripts could be identified as being chimeric, such as the forespore-specific germination genes transcripts *gerBA* and *gerBC* that were found to interact with genes such as *coxA*, an endospore cortex protein. This is a very early indication that the likelihood of AMT entering the forespore is high and the quality of reads was good, but further tests and more extensive sequencing will be required for depth of coverage. A similar number of chimeras were identified for both with and without AMT, indicating a level of crosslinking may be occurring without the assistance of AMT. This could be due to the UV treatment and as such a non-UV control could be advantageous in the future.

Whilst the proteins discussed throughout this study have links to sporulation and germination, it remains to be seen how proteins with these specific functions can be attributed to previously seen phenotypes and thus it is unlikely that these proteins are the targets of these sRNAs. In conclusion, further studies with dormant endospore proteins in the future can be improved upon with the new endospore protein extraction protocol devised here. However, this method was not effective for the pursuit of sRNA targets. The future of sRNA studies may hang on the success of RNA-RNA interaction studies. In addition, if direct binding partners can be confirmed, further characterisation can become tailored to the interaction.

6 Chapter 6: Discussion

B. subtilis has been shown to produce a compendium of RNA species that did not appear in the original annotation of the genome (Nicolas, P. *et al.*, 2012; Kunst, F. *et al.*, 1997; Irnov, I. *et al.*, 2010; Rasmussen, S. *et al.*, 2009). The most in-depth study on sRNAs in *B. subtilis* was carried out by Nicolas *et al.* in 2012. This study documented the transcriptional landscape of the wild-type laboratory strain, *B. subtilis* 168 trp+, using tiling arrays. A total of 104 different conditions were analysed, which identified 1583 previously unannotated RNA species (Nicolas, P. *et al.*, 2012).

The initial part of this study, described in chapter three, aimed to identify potentially functional sRNAs active in sporulation in *B. subtilis* via a bioinformatics analysis to determine which sRNAs warranted further experimental study. This study was split into two parts, the first utilising transcriptional profiling to identify differential regulation of putative sRNAs and expression during sporulation and then promoter and terminator predictions to further narrow this list. The second objective was to use combinational approaches to reduce false positive indications of putative sRNAs.

This workflow facilitated the capture of putative sRNA sequences to create a list of 85 newly annotated S segments which are likely to be independently expressed during the sporulation time course (Table 5). Further analysis of promoter regions indicated that the 85 putative sRNA sequences possessed enrichment for sigma factor binding motifs which are sporulation specific. Based on previous research on protein coding genes involved in sporulation, this provides an excellent indication for a sporulation-specific sRNAs. The workflow successfully narrowed down potential candidates to 24, but experimental evidence is essential for confirmation. Regardless, it was possible to take those identified in chapter three and tailor experiments to further characterise putative sRNAs involved in the various processes of endospore formation.

DNA binding motifs were predicted for the transcriptional regulator CodY for six putative sRNAs. The CodY regulon is comprised of many genes important in the transition from exponential growth to stationary phase and also sporulation. For example, a member of the CodY regulon is Spo0A, which is a major transcription

factor itself, essential for the initiation of sporulation (Chastanet, A. *et al.*, 2010). Differential gene expression reanalysis of data produced by Brinsmade, S. R. *et al.* (2014), as their study did not look for sRNAs, revealed that two of the predicted CodY binding sites for sporulation-related sRNAs were indeed regulated by CodY. In addition to this, there are a further 56 S segments with a statistically significant log2 fold change, suggesting an expanded CodY regulon.

This study provided a proof-of-principle for the identification of sRNAs and their regulators. Meeske, A. J. *et al.* (2016) utilised Tn-seq to identify essential genes involved in endospore formation of *B. subtilis*. 28 putative sRNAs were confirmed to have a large ratio change from control conditions to sporulation conditions in a transposon sequencing library, 11 of which were featured in the selection from the workflow. Furthermore, Nagler, K. *et al.* (2016) studied the transcriptional landscape of endospores during outgrowth via RNAseq. Reanalysis of these data signified that genes in the 10 most highly expressed genes in a dormant endospore include four S segments, one of which was selected in the workflow (S357).

Chapter four first confirmed sporulation specific expression of eleven sRNAs identified by the workflow in chapter three. Four sRNAs S357, S547, S612 and S849 that were seen to be expressed in the forespore were then chosen to study in further detail. This study successfully identified three putative sRNAs to have phenotypes in sporulation or germination (S357, S547 and S612).

Chapter five was dedicated to target finding utilising omics techniques. RNAseq was deemed unreliable, as the RNA samples obtained from cultures undergoing sporulation was not homogenous. This led to the development of a much-improved endospore proteome extraction method, yielding over 1000 proteins as opposed to previous studies which only identified proteins in the early hundreds. Whilst unsuccessful for the search for the sRNA targets, this method could be used for other studies for the characterisation of dormant endospores.

Chapter five also touched upon methodologies for future use. AMT RNA crosslinking was implemented and identified RNA-RNA interactions during the sporulation process. The direct method of detecting sRNA-RNA interactions in native conditions is one way forward for the identification of sRNAs and their targets, which should be followed by phenotypic assays tailored to specific targets.

Experimental technologies are constantly improving and because of this RNA secondary structure predictions are now being paired with high-throughput sequencing. Identification of RNA interaction groups by mutational profiling (RING-MaP). RING-MaP utilises chemical probing with dimethyl sulphate, termed mutational profiling (MaP), and the resultant cDNA is then read by massively parallel sequencing (Krokhotin, A. *et al.*, 2017). This combination can identify base-pairing interactions, and therefore resolving the secondary structure. This could then guide predictions of accessible regions of sRNAs which could be seed regions for base-paring with their targets.

The identification of a RNA binding protein in sporulation would facilitate sRNA target finding. Methods such as gradient profiling by sequencing (GRAD-seq) has previously been utilised to identify RNA-protein duplexes (Smirnov, A. *et al.*, 2016). GRAD-seq successfully pinpointed a conserved protein, ProQ, to associate with a large group of highly structured sRNAs. This approach could be utilised during sporulation to identify a Hfq-like binding partner, which could then be used in methods such as RIL-seq and CLASH to identify sRNA targets.

Overall, this study highlighted the benefit of compiling information of putative sRNAs and the use of specialised tools prior to stepping into the lab. In addition, it provided the essential foundations for the characterisation of three sRNAs, S357, S547 and S612. Most significantly, method development of protein extraction of dormant endospores was successful, providing a significant improvement on a method that has been very ineffectual up to this point. Confirmation of sRNA-target interactions is behind identification of an sRNA, where thousands of putative sRNAs have been ascertained but as little as 492 have validated target interactions, fewer still with

binding regions being characterised (Wang, J. *et al.*, 2016). However, with new high-throughput methods of directly determining sRNA-target interactions further characterisation of the three sRNAs will continue with the identification of their binding partners and hence their mechanisms of actions.

References

Abe, A.; Ogawa, S.; Kohno, T. & Watabe, K. (1993) Purification of Bacillus subtilis spore coat protein by electrophoretic elution procedure and determination of NH2-terminal amino acid sequences. *Microbiol Immunol*, 37 (10): 809-812.

Abe, T.; Sakaki, K.; Fujihara, A.; Ujiie, H.; Ushida, C.; Himeno, H.; Sato, T. & Muto, A. (2008) tmRNA-dependent trans-translation is required for sporulation in Bacillus subtilis. *Mol Microbiol*, 69 (6): 1491-1498.

Abhyankar, W.; Beek, A. T.; Dekker, H.; Kort, R.; Brul, S. & de Koster, C. G. (2011) Gel-free proteomic identification of the Bacillus subtilis insoluble spore coat protein fraction. *Proteomics*, 11 (23): 4541-4550.

Aiba, H.; Matsuyama, S.; Mizuno, T. & Mizushima, S. (1987) Function of micF as an antisense RNA in osmoregulatory expression of the ompF gene in Escherichia coli. *J Bacteriol*, 169 (7): 3007-3012.

Alen, C. & Sonenshein, A. L. (1999) Bacillus subtilis aconitase is an RNA-binding protein. *Proc Natl Acad Sci U S A*, 96 (18): 10412-10417.

Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W. & Lipman, D. J. (1990) Basic local alignment search tool. *J Mol Biol*, 215 (3): 403-410.

Argaman, L.; Hershberg, R.; Vogel, J.; Bejerano, G.; Wagner, E. G.; Margalit, H. & Altuvia, S. (2001) Novel small RNA-encoding genes in the intergenic regions of Escherichia coli. *Curr Biol*, 11 (12): 941-950.

Arnold, T. E.; Yu, J. & Belasco, J. G. (1998) mRNA stabilization by the ompA 5' untranslated region: two protective elements hinder distinct pathways for mRNA degradation. *Rna*, 4 (3): 319-330.

Atrih, A. & Foster, S. J. (1999) The role of peptidoglycan structure and structural dynamics during endospore dormancy and germination. *Antonie Van Leeuwenhoek*, 75 (4): 299-307.

Atrih, A.; Zollner, P.; Allmaier, G. & Foster, S. J. (1996) Structural analysis of Bacillus subtilis 168 endospore peptidoglycan and its role during differentiation. *J Bacteriol*, 178 (21): 6173-6183.

Backofen, R. & Hess, W. R. (2010) Computational prediction of sRNAs and their targets in bacteria. *RNA Biol*, 7 (1): 33-42.

Bagyan, I. & Setlow, P. (2002) Localization of the cortex lytic enzyme CwlJ in spores of Bacillus subtilis. *J Bacteriol*, 184 (4): 1219-1224.

Baker, C. S.; Morozov, I.; Suzuki, K.; Romeo, T. & Babitzke, P. (2002) CsrA regulates glycogen biosynthesis by preventing translation of glgC in Escherichia coli. *Mol Microbiol*, 44 (6): 1599-1610.

Bandyra, K. J.; Said, N.; Pfeiffer, V.; Gorna, M. W.; Vogel, J. & Luisi, B. F. (2012) The seed region of a small RNA drives the controlled destruction of the target mRNA by the endoribonuclease RNase E. *Mol Cell*, 47 (6): 943-953.

Beaman, T. C.; Greenamyre, J. T.; Corner, T. R.; Pankratz, H. S. & Gerhardt, P. (1982) Bacterial spore heat resistance correlated with water content, wet density, and protoplast/sporoplast volume ratio. *J Bacteriol*, 150 (2): 870-877.

Belda, E.; Sekowska, A.; Le Fevre, F.; Morgat, A.; Mornico, D.; Ouzounis, C.; Vallenet, D.; Medigue, C. & Danchin, A. (2013) An updated metabolic view of the Bacillus subtilis 168 genome. *Microbiology*, 159 (Pt 4): 757-770.

Ben-Yehuda, S.; Fujita, M.; Liu, X. S.; Gorbatyuk, B.; Skoko, D.; Yan, J.; Marko, J. F.; Liu, J. S.; Eichenberger, P.; Rudner, D. Z. & Losick, R. (2005) Defining a centromere-like element in Bacillus subtilis by Identifying the binding sites for the chromosome-anchoring protein RacA. *Mol Cell*, 17 (6): 773-782.

Ben-Yehuda, S.; Rudner, D. Z. & Losick, R. (2003) RacA, a bacterial protein that anchors chromosomes to the cell poles. *Science*, 299 (5606): 532-536.

Bennett, G. M. & Moran, N. A. (2013) Small, smaller, smallest: the origins and evolution of ancient dual symbioses in a Phloem-feeding insect. *Genome Biol Evol*, 5 (9): 1675-1688.

Bernhart, S. H.; Hofacker, I. L.; Will, S.; Gruber, A. R. & Stadler, P. F. (2008) RNAalifold: improved consensus structure prediction for RNA alignments. *BMC Bioinformatics*, 9 474.

Bidnenko, V.; Nicolas, P.; Grylak-Mielnicka, A.; Delumeau, O.; Auger, S.; Aucouturier, A.; Guerin, C.; Repoila, F.; Bardowski, J.; Aymerich, S. & Bidnenko, E. (2017) Termination factor Rho: From the control of pervasive transcription to cell fate determination in Bacillus subtilis. *PLoS Genet*, 13 (7): e1006909.

Bilusic, I.; Popitsch, N.; Rescheneder, P.; Schroeder, R. & Lybecker, M. (2014) Revisiting the coding potential of the E. coli genome through Hfq co-immunoprecipitation. *RNA Biol*, 11 (5): 641-654.

Block, K. F.; Hammond, M. C. & Breaker, R. R. (2010) Evidence for widespread gene control function by the ydaO riboswitch candidate. *J Bacteriol*, 192 (15): 3983-3989.

Bohn, C.; Rigoulay, C. & Bouloc, P. (2007) No detectable effect of RNA-binding protein Hfq absence in Staphylococcus aureus. *BMC Microbiol*, 7 10.

Bohn, C.; Rigoulay, C.; Chabelskaya, S.; Sharma, C. M.; Marchais, A.; Skorski, P.; Borezee-Durant, E.; Barbet, R.; Jacquet, E.; Jacq, A.; Gautheret, D.; Felden, B.; Vogel, J. & Bouloc, P. (2010) Experimental discovery of small RNAs in Staphylococcus aureus reveals a riboregulator of central metabolism. *Nucleic Acids Res*, 38 (19): 6620-6636.

Botella, E.; Fogg, M.; Jules, M.; Piersma, S.; Doherty, G.; Hansen, A.; Denham, E. L.; Le Chat, L.; Veiga, P.; Bailey, K.; Lewis, P. J.; van Dijl, J. M.; Aymerich, S.; Wilkinson, A. J. & Devine, K. M. (2010) pBaSysBioII: an integrative plasmid generating gfp transcriptional fusions for high-throughput analysis of gene expression in Bacillus subtilis. *Microbiology*, 156 (Pt 6): 1600-1608.

Brinsmade, S. R.; Alexander, E. L.; Livny, J.; Stettner, A. I.; Segre, D.; Rhee, K. Y. & Sonenshein, A. L. (2014) Hierarchical expression of genes controlled by the Bacillus subtilis global regulatory protein CodY. *Proc Natl Acad Sci U S A*, 111 (22): 8227-8232.

Brown, L. & Elliott, T. (1997) Mutations that increase expression of the rpoS gene and decrease its dependence on hfq function in Salmonella typhimurium. *J Bacteriol*, 179 (3): 656-662.

Buescher, J. M.; Liebermeister, W.; Jules, M.; Uhr, M.; Muntel, J.; Botella, E.; Hessling, B.; Kleijn, R. J.; Le Chat, L.; Lecointe, F.; Mader, U.; Nicolas, P.; Piersma, S.; Rugheimer, F.; Becher, D.; Bessieres, P.; Bidnenko, E.; Denham, E. L.; Dervyn, E.; Devine, K. M.; Doherty, G.; Drulhe, S.; Felicori, L.; Fogg, M. J.; Goelzer, A.; Hansen, A.; Harwood, C. R.; Hecker, M.; Hubner, S.; Hultschig, C.; Jarmer, H.; Klipp, E.; Leduc, A.; Lewis, P.; Molina, F.; Noirot, P.; Peres, S.; Pigeonneau, N.; Pohl, S.; Rasmussen, S.; Rinn, B.; Schaffer, M.; Schnidder, J.; Schwikowski, B.; Van Dijl, J. M.; Veiga, P.; Walsh, S.; Wilkinson, A. J.; Stelling, J.; Aymerich, S. & Sauer, U. (2012) Global network reorganization during dynamic adaptations of Bacillus subtilis metabolism. *Science*, 335 (6072): 1099-1103.

Busch, A.; Richter, A. S. & Backofen, R. (2008) IntaRNA: efficient prediction of bacterial sRNA targets incorporating target site accessibility and seed regions. *Bioinformatics*, 24 (24): 2849-2856.

Cabrera-Martinez, R. M.; Tovar-Rojo, F.; Vepachedu, V. R. & Setlow, P. (2003) Effects of overexpression of nutrient receptors on germination of spores of Bacillus subtilis. *J Bacteriol*, 185 (8): 2457-2464.

Camp, A. H. & Losick, R. (2008) A novel pathway of intercellular signalling in Bacillus subtilis involves a protein with similarity to a component of type III secretion channels. *Mol Microbiol*, 69 (2): 402-417.

Camp, A. H. & Losick, R. (2009) A feeding tube model for activation of a cell-specific transcription factor during sporulation in Bacillus subtilis. *Genes Dev*, 23 (8): 1014-1024.

Canas, C.; Suzuki, Y.; Marchisone, C.; Carrasco, B.; Freire-Beneitez, V.; Takeyasu, K.; Alonso, J. C. & Ayora, S. (2014) Interaction of branch migration translocases with the Holliday junction-resolving enzyme and their implications in Holliday junction resolution. *J Biol Chem*, 289 (25): 17634-17646.

Cangiano, G.; Sirec, T.; Panarella, C.; Isticato, R.; Baccigalupi, L.; De Felice, M. & Ricca, E. (2014) The sps Gene Products Affect the Germination, Hydrophobicity, and Protein Adsorption of Bacillus subtilis Spores. *Appl Environ Microbiol*, 80 (23): 7293-7302.

Cao, M.; Kobel, P. A.; Morshedi, M. M.; Wu, M. F.; Paddon, C. & Helmann, J. D. (2002a) Defining the Bacillus subtilis sigma(W) regulon: a comparative analysis of promoter consensus search, run-off transcription/macroarray analysis (ROMA), and transcriptional profiling approaches. *J Mol Biol*, 316 (3): 443-457.

Cao, M.; Wang, T.; Ye, R. & Helmann, J. D. (2002b) Antibiotics that inhibit cell wall biosynthesis induce expression of the Bacillus subtilis sigma(W) and sigma(M) regulons. *Mol Microbiol*, 45 (5): 1267-1276.

Chalfie, M.; Tu, Y.; Euskirchen, G.; Ward, W. W. & Prasher, D. C. (1994) Green fluorescent protein as a marker for gene expression. *Science*, 263 (5148): 802-805.

Chan, J. M.; Guttenplan, S. B. & Kearns, D. B. (2014) Defects in the flagellar motor increase synthesis of poly-gamma-glutamate in Bacillus subtilis. *J Bacteriol*, 196 (4): 740-753.

Chastanet, A.; Vitkup, D.; Yuan, G. C.; Norman, T. M.; Liu, J. S. & Losick, R. M. (2010) Broadly heterogeneous activation of the master regulator for sporulation in Bacillus subtilis. *Proc Natl Acad Sci U S A*, 107 (18): 8486-8491.

Chen, S.; Lesnik, E. A.; Hall, T. A.; Sampath, R.; Griffey, R. H.; Ecker, D. J. & Blyn, L. B. (2002) A bioinformatics based approach to discover small RNA genes in the Escherichia coli genome. *Biosystems*, 65 (2-3): 157-177.

Chirakkal, H.; O'Rourke, M.; Atrih, A.; Foster, S. J. & Moir, A. (2002) Analysis of spore cortex lytic enzymes and related proteins in Bacillus subtilis endospore germination. *Microbiology*, 148 (Pt 8): 2383-2392.

Christiansen, J. K.; Nielsen, J. S.; Ebersbach, T.; Valentin-Hansen, P.; Sogaard-Andersen, L. & Kallipolitis, B. H. (2006) Identification of small Hfq-binding RNAs in Listeria monocytogenes. *Rna*, 12 (7): 1383-1396.

Chumsakul, O.; Takahashi, H.; Oshima, T.; Hishimoto, T.; Kanaya, S.; Ogasawara, N. & Ishikawa, S. (2011) Genome-wide binding profiles of the Bacillus subtilis transition state regulator AbrB and its homolog Abh reveals their interactive role in transcriptional regulation. *Nucleic Acids Res*, 39 (2): 414-428.

Chung, Y. S. & Dubnau, D. (1998) All seven comG open reading frames are required for DNA binding during transformation of competent Bacillus subtilis. *J Bacteriol*, 180 (1): 41-45.

Commichau, F. M.; Gunka, K.; Landmann, J. J. & Stulke, J. (2008) Glutamate metabolism in Bacillus subtilis: gene expression and enzyme activities evolved to avoid futile cycles and to allow rapid responses to perturbations of the system. *J Bacteriol*, 190 (10): 3557-3564.

Core, L. & Perego, M. (2003) TPR-mediated interaction of RapC with ComA inhibits response regulator-DNA binding for competence development in Bacillus subtilis. *Mol Microbiol*, 49 (6): 1509-1522.

Costa, T.; Isidro, A. L.; Moran, C. P., Jr. & Henriques, A. O. (2006) Interaction between coat morphogenetic proteins SafA and SpoVID. *J Bacteriol*, 188 (22): 7731-7741.

Crick, F. (1970) Central dogma of molecular biology. *Nature*, 227 (5258): 561-563.

Crooks, G. E.; Hon, G.; Chandonia, J. M. & Brenner, S. E. (2004) WebLogo: a sequence logo generator. *Genome Res*, 14 (6): 1188-1190.

Crow, A.; Liu, Y.; Moller, M. C.; Le Brun, N. E. & Hederstedt, L. (2009) Structure and functional properties of Bacillus subtilis endospore biogenesis factor StoA. *J Biol Chem*, 284 (15): 10056-10066.

Dambach, M.; Irnov, I. & Winkler, W. C. (2013) Association of RNAs with Bacillus subtilis Hfq. *PLoS One*, 8 (2): e55156.

Dar, D.; Shamir, M.; Mellin, J. R.; Koutero, M.; Stern-Ginossar, N.; Cossart, P. & Sorek, R. (2016) Term-seq reveals abundant ribo-regulation of antibiotics resistance in bacteria. *Science*, 352 (6282): aad9822.

Darfeuille, F.; Unoson, C.; Vogel, J. & Wagner, E. G. (2007) An antisense RNA inhibits translation by competing with standby ribosomes. *Mol Cell*, 26 (3): 381-392.

Das, A.; Court, D. & Adhya, S. (1976) Isolation and characterization of conditional lethal mutants of Escherichia coli defective in transcription termination factor rho. *Proc Natl Acad Sci U S A*, 73 (6): 1959-1963.

de Hoon, M. J.; Eichenberger, P. & Vitkup, D. (2010) Hierarchical evolution of the bacterial sporulation network. *Curr Biol*, 20 (17): R735-745.

de Jong, I. G.; Veening, J. W. & Kuipers, O. P. (2010) Heterochronic phosphorelay gene expression as a source of heterogeneity in Bacillus subtilis spore formation. *J Bacteriol*, 192 (8): 2053-2067.

Deatherage Kaiser, B. L.; Wunschel, D. S.; Sydor, M. A.; Warner, M. G.; Wahl, K. L. & Hutchison, J. R. (2015) Improved proteomic analysis following trichloroacetic acid extraction of Bacillus anthracis spore proteins. *J Microbiol Methods*, 118 18-24.

Dembek, M.; Barquist, L.; Boinett, C. J.; Cain, A. K.; Mayho, M.; Lawley, T. D.; Fairweather, N. F. & Fagan, R. P. (2015) High-throughput analysis of gene essentiality and sporulation in Clostridium difficile. *MBio*, 6 (2): e02383.

Ding, Y.; Davis, B. M. & Waldor, M. K. (2004) Hfq is essential for Vibrio cholerae virulence and downregulates sigma expression. *Mol Microbiol*, 53 (1): 345-354.

Doherty, G. P.; Bailey, K. & Lewis, P. J. (2010) Stage-specific fluorescence intensity of GFP and mCherry during sporulation In Bacillus Subtilis. *BMC Res Notes*, 3 303.

Donovan, W.; Zheng, L. B.; Sandman, K. & Losick, R. (1987) Genes encoding spore coat polypeptides from Bacillus subtilis. *J Mol Biol*, 196 (1): 1-10.

Doona, C. J.; Ghosh, S.; Feeherry, F. F.; Ramirez-Peralta, A.; Huang, Y.; Chen, H. & Setlow, P. (2014) High pressure germination of Bacillus subtilis spores with alterations in levels and types of germination proteins. *J Appl Microbiol*, 117 (3): 711-720.

Dubey, A. K.; Baker, C. S.; Romeo, T. & Babitzke, P. (2005) RNA sequence and secondary structure participate in high-affinity CsrA-RNA interaction. *Rna*, 11 (10): 1579-1587.

Dugar, G.; Herbig, A.; Forstner, K. U.; Heidrich, N.; Reinhardt, R.; Nieselt, K. & Sharma, C. M. (2013) High-resolution transcriptome maps reveal strain-specific regulatory features of multiple Campylobacter jejuni isolates. *PLoS Genet*, 9 (5): e1003495.

Duncan, L.; Alper, S. & Losick, R. (1996) SpoIIAA governs the release of the cell-type specific transcription factor sigma F from its anti-sigma factor SpoIIAB. *J Mol Biol*, 260 (2): 147-164.

Durand, S.; Braun, F.; Lioliou, E.; Romilly, C.; Helfer, A. C.; Kuhn, L.; Quittot, N.; Nicolas, P.; Romby, P. & Condon, C. (2015) A nitric oxide regulated small RNA controls expression of genes involved in redox homeostasis in Bacillus subtilis. *PLoS Genet*, 11 (2): e1004957.

Durand, S.; Gilet, L. & Condon, C. (2012) The essential function of B. subtilis RNase III is to silence foreign toxin genes. *PLoS Genet*, 8 (12): e1003181.

Ebert, M. S.; Neilson, J. R. & Sharp, P. A. (2007) MicroRNA sponges: competitive inhibitors of small RNAs in mammalian cells. *Nat Methods*, 4 (9): 721-726.

Ebert, M. S. & Sharp, P. A. (2010) Emerging roles for natural microRNA sponges. *Curr Biol*, 20 (19): R858-861.

Edwards, T. E.; Klein, D. J. & Ferre-D'Amare, A. R. (2007) Riboswitches: small-molecule recognition by gene regulatory RNAs. *Curr Opin Struct Biol*, 17 (3): 273-279.

Eggenhofer, F.; Tafer, H.; Stadler, P. F. & Hofacker, I. L. (2011) RNApredator: fast accessibility-based prediction of sRNA targets. *Nucleic Acids Res*, 39 (Web Server issue): W149-154.

Eiamphungporn, W. & Helmann, J. D. (2009) Extracytoplasmic function sigma factors regulate expression of the Bacillus subtilis yabE gene via a cis-acting antisense RNA. *J Bacteriol*, 191 (3): 1101-1105.

Eichenberger, P.; Fawcett, P. & Losick, R. (2001) A three-protein inhibitor of polar septation during sporulation in Bacillus subtilis. *Mol Microbiol*, 42 (5): 1147-1162.

Eichenberger, P.; Fujita, M.; Jensen, S. T.; Conlon, E. M.; Rudner, D. Z.; Wang, S. T.; Ferguson, C.; Haga, K.; Sato, T.; Liu, J. S. & Losick, R. (2004) The program of gene transcription for a single differentiating cell type during sporulation in Bacillus subtilis. *PLoS Biol*, 2 (10): e328.

Eichenberger, P.; Jensen, S. T.; Conlon, E. M.; van Ooij, C.; Silvaggi, J.; Gonzalez-Pastor, J. E.; Fujita, M.; Ben-Yehuda, S.; Stragier, P.; Liu, J. S. & Losick, R. (2003) The sigmaE regulon and the identification of additional sporulation genes in Bacillus subtilis. *J Mol Biol*, 327 (5): 945-972.

Eijlander, R. T.; de Jong, A.; Krawczyk, A. O.; Holsappel, S. & Kuipers, O. P. (2013) SporeWeb: an interactive journey through the complete sporulation cycle of Bacillus subtilis. *Nucleic Acids Res*,

Eijlander, RT; Holsappel, S; de Jong, A; Ghosh, A; Christie, G & Kuipers, OP (2016) SpoVT: From Fine-Tuning Regulator in Bacillus subtilis to Essential Sporulation. *Front Microbiol.*, 7 - 1607.

Erlendsson, L. S.; Moller, M. & Hederstedt, L. (2004) Bacillus subtilis StoA Is a thiol-disulfide oxidoreductase important for spore cortex synthesis. *J Bacteriol*, 186 (18): 6230-6238.

Escamilla, J. E.; Barquera, B.; Ramirez, R.; Garcia-Horsman, A. & del Arenal, P. (1988) Role of menaquinone in inactivation and activation of the Bacillus cereus forespore respiratory system. *J Bacteriol*, 170 (12): 5908-5912.

Fabret, C.; Ehrlich, S. D. & Noirot, P. (2002) A new mutation delivery system for genome-scale approaches in Bacillus subtilis. *Mol Microbiol*, 46 (1): 25-36.

Fimlaid, K. A. & Shen, A. (2015) Diverse mechanisms regulate sporulation sigma factor activity in the Firmicutes. *Curr Opin Microbiol*, 24 88-95.

Finn, R. D.; Tate, J.; Mistry, J.; Coggill, P. C.; Sammut, S. J.; Hotz, H. R.; Ceric, G.; Forslund, K.; Eddy, S. R.; Sonnhammer, E. L. & Bateman, A. (2008) The Pfam protein families database. *Nucleic Acids Res*, 36 (Database issue): D281-288.

Fleischmann, R. D.; Adams, M. D.; White, O.; Clayton, R. A.; Kirkness, E. F.; Kerlavage, A. R.; Bult, C. J.; Tomb, J. F.; Dougherty, B. A.; Merrick, J. M. & et al. (1995) Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. *Science*, 269 (5223): 496-512.

Fortnagel, P. & Freese, E. (1968) Analysis of sporulation mutants. II. Mutants blocked in the citric acid cycle. *J Bacteriol*, 95 (4): 1431-1438.

Fouet, A.; Klier, A. & Rapoport, G. (1986) Nucleotide sequence of the sucrase gene of Bacillus subtilis. *Gene*, 45 (2): 221-225.

Franze de Fernandez, M. T.; Hayward, W. S. & August, J. T. (1972) Bacterial proteins required for replication of phage Q ribonucleic acid. Pruification and properties of host factor I, a ribonucleic acid-binding protein. *J Biol Chem*, 247 (3): 824-831.

Fraser, C. M.; Gocayne, J. D.; White, O.; Adams, M. D.; Clayton, R. A.; Fleischmann, R. D.; Bult, C. J.; Kerlavage, A. R.; Sutton, G.; Kelley, J. M.; Fritchman, R. D.; Weidman, J. F.; Small, K. V.; Sandusky, M.; Fuhrmann, J.; Nguyen, D.; Utterback, T. R.; Saudek, D. M.; Phillips, C. A.; Merrick, J. M.; Tomb, J. F.; Dougherty, B. A.; Bott, K. F.; Hu, P. C.; Lucier, T. S.; Peterson, S. N.; Smith, H. O.; Hutchison, C. A., 3rd & Venter, J. C. (1995) The minimal gene complement of Mycoplasma genitalium. *Science*, 270 (5235): 397-403.

Freiburg Bioinformatics, Group (2017) *Freiburg RNA Tools* [online] Available from: http://rna.informatik.uni-freiburg.de/ (Accessed

Fujita, M. & Losick, R. (2005) Evidence that entry into sporulation in Bacillus subtilis is governed by a gradual increase in the level and activity of the master regulator Spo0A. *Genes Dev*, 19 (18): 2236-2244.

Computer ProgramGalili, T. (2017) heatmaply: Interactive Cluster Heat Maps Using 'plotly'. .

Galperin, M. Y. & Koonin, E. V. (2000) Who's your neighbor? New computational approaches for functional genomics. *Nat Biotechnol*, 18 (6): 609-613.

Galperin, M. Y.; Mekhedov, S. L.; Puigbo, P.; Smirnov, S.; Wolf, Y. I. & Rigden, D. J. (2012) Genomic determinants of sporulation in Bacilli and Clostridia: towards the minimal set of sporulation-specific genes. *Environ Microbiol*, 14 (11): 2870-2890.

Ghosh, S.; Scotland, M. & Setlow, P. (2012) Levels of germination proteins in dormant and superdormant spores of Bacillus subtilis. *J Bacteriol*, 194 (9): 2221-2227.

Ghosh, S. & Setlow, P. (2009) Isolation and characterization of superdormant spores of Bacillus species. *J Bacteriol*, 191 (6): 1787-1797.

Gilmore, M. E.; Bandyopadhyay, D.; Dean, A. M.; Linnstaedt, S. D. & Popham, D. L. (2004) Production of muramic delta-lactam in Bacillus subtilis spore peptidoglycan. *J Bacteriol*, 186 (1): 80-89.

Gimpel, M.; Heidrich, N.; Mader, U.; Krugel, H. & Brantl, S. (2010) A dual-function sRNA from B. subtilis: SR1 acts as a peptide encoding mRNA on the gapA operon. *Mol Microbiol*, 76 (4): 990-1009.

Gimpel, M.; Preis, H.; Barth, E.; Gramzow, L. & Brantl, S. (2012) SR1--a small RNA with two remarkably conserved functions. *Nucleic Acids Res*, 40 (22): 11659-11672.

Goldman, R. C. & Tipper, D. J. (1978) Bacillus subtilis spore coats: complexity and purification of a unique polypeptide component. *J Bacteriol*, 135 (3): 1091-1106.

Goldrick, S. & Setlow, P. (1983) Expression of a Bacillus megaterium sporulation-specific gene during sporulation of Bacillus subtilis. *J Bacteriol*, 155 (3): 1459-1462.

Gollnick, P.; Babitzke, P.; Antson, A. & Yanofsky, C. (2005) Complexity in regulation of tryptophan biosynthesis in Bacillus subtilis. *Annu Rev Genet*, 39 47-68.

Gordon, G. C.; Cameron, J. C. & Pfleger, B. F. (2017) RNA Sequencing Identifies New RNase III Cleavage Sites in Escherichia coli and Reveals Increased Regulation of mRNA. *MBio*, 8 (2):

Gorke, B. & Vogel, J. (2008) Noncoding RNA control of the making and breaking of sugars. *Genes Dev*, 22 (21): 2914-2925.

Gorodkin, J.; Hofacker, I. L.; Torarinsson, E.; Yao, Z.; Havgaard, J. H. & Ruzzo, W. L. (2010) De novo prediction of structured RNAs from genomic sequences. *Trends Biotechnol*, 28 (1): 9-19.

Gottesman, S. & Storz, G. (2011) Bacterial small RNA regulators: versatile roles and rapidly evolving variations. *Cold Spring Harb Perspect Biol*, 3 (12):

Griffiths, K. K.; Zhang, J.; Cowan, A. E.; Yu, J. & Setlow, P. (2011) Germination proteins in the inner membrane of dormant Bacillus subtilis spores colocalize in a discrete cluster. *Mol Microbiol*, 81 (4): 1061-1077.

Gruber, A. R.; Martin, G.; Keller, W. & Zavolan, M. (2014) Means to an end: mechanisms of alternative polyadenylation of messenger RNA precursors. *Wiley Interdiscip Rev RNA*, 5 (2): 183-196.

Grundy, F. J. & Henkin, T. M. (1998) The S box regulon: a new global transcription termination control system for methionine and cysteine biosynthesis genes in gram-positive bacteria. *Mol Microbiol*, 30 (4): 737-749.

Guerrier-Takada, C.; Gardiner, K.; Marsh, T.; Pace, N. & Altman, S. (1983) The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell*, 35 (3 Pt 2): 849-857.

Gundlach, J.; Herzberg, C.; Kaever, V.; Gunka, K.; Hoffmann, T.; Weiss, M.; Gibhardt, J.; Thurmer, A.; Hertel, D.; Daniel, R.; Bremer, E.; Commichau, F. M. & Stulke, J. (2017) Control of potassium homeostasis is an essential function of the second messenger cyclic di-AMP in Bacillus subtilis. *Sci Signal*, 10 (475):

Gupta, M. & Rao, K. K. (2014) Phosphorylation of DegU is essential for activation of amyE expression in Bacillus subtilis. *J Biosci*, 39 (5): 747-752.

Guttenplan, S. B.; Shaw, S. & Kearns, D. B. (2013) The cell biology of peritrichous flagella in Bacillus subtilis. *Mol Microbiol*, 87 (1): 211-229.

Haldenwang, W. G. (1995) The sigma factors of Bacillus subtilis. *Microbiol Rev*, 59 (1): 1-30.

Hammerle, H.; Amman, F.; Vecerek, B.; Stulke, J.; Hofacker, I. & Blasi, U. (2014) Impact of Hfq on the Bacillus subtilis Transcriptome. *PLoS One*, 9 (6): e98661.

Hamoen, L. W.; Venema, G. & Kuipers, O. P. (2003) Controlling competence in Bacillus subtilis: shared use of regulators. *Microbiology*, 149 (Pt 1): 9-17.

Handke, L. D.; Shivers, R. P. & Sonenshein, A. L. (2008) Interaction of Bacillus subtilis CodY with GTP. *J Bacteriol*, 190 (3): 798-806.

Hartig, E.; Geng, H.; Hartmann, A.; Hubacek, A.; Munch, R.; Ye, R. W.; Jahn, D. & Nakano, M. M. (2004) Bacillus subtilis ResD induces expression of the potential regulatory genes yclJK upon oxygen limitation. *J Bacteriol*, 186 (19): 6477-6484.

Hartzog, G. A. & Martens, J. A. (2009) ncRNA transcription makes its mark. *Embo j*, 28 (12): 1679-1680.

Harwood, C. R., and S. M. Cutting. (1990) *Molecular biological methods for Bacillus.* New York, N.Y.: John Wiley & Sons.

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

Heidrich, N.; Chinali, A.; Gerth, U. & Brantl, S. (2006) The small untranslated RNA SR1 from the Bacillus subtilis genome is involved in the regulation of arginine catabolism. *Mol Microbiol*, 62 (2): 520-536.

Henriques, A. O. & Moran, C. P., Jr. (2007) Structure, assembly, and function of the spore surface layers. *Annu Rev Microbiol*, 61 555-588.

Higgins, D. & Dworkin, J. (2012) Recent progress in Bacillus subtilis sporulation. *FEMS Microbiol Rev*, 36 (1): 131-148.

Higgs, R. E.; Knierman, M. D.; Gelfanova, V.; Butler, J. P. & Hale, J. E. (2005) Comprehensive label-free method for the relative quantification of proteins from biological samples. *J Proteome Res*, 4 (4): 1442-1450.

Hilbert, D. W. & Piggot, P. J. (2004) Compartmentalization of gene expression during Bacillus subtilis spore formation. *Microbiol Mol Biol Rev*, 68 (2): 234-262.

Himes, P.; McBryant, S. J. & Kroos, L. (2010) Two regions of Bacillus subtilis transcription factor SpoIIID allow a monomer to bind DNA. *J Bacteriol*, 192 (6): 1596-1606.

Hindley, J. (1967) Fractionation of 32P-labelled ribonucleic acids on polyacrylamide gels and their characterization by fingerprinting. *J Mol Biol*, 30 (1): 125-136.

Hofacker, I. L.; Fekete, M. & Stadler, P. F. (2002) Secondary structure prediction for aligned RNA sequences. *J Mol Biol*, 319 (5): 1059-1066.

Hoper, D.; Volker, U. & Hecker, M. (2005) Comprehensive characterization of the contribution of individual SigB-dependent general stress genes to stress resistance of Bacillus subtilis. *J Bacteriol*, 187 (8): 2810-2826.

Hudson, K. D.; Corfe, B. M.; Kemp, E. H.; Feavers, I. M.; Coote, P. J. & Moir, A. (2001) Localization of GerAA and GerAC germination proteins in the Bacillus subtilis spore. *J Bacteriol*, 183 (14): 4317-4322.

Imamura, D.; Kuwana, R.; Takamatsu, H. & Watabe, K. (2011) Proteins involved in formation of the outermost layer of Bacillus subtilis spores. *J Bacteriol*, 193 (16): 4075-4080.

Online DatabaseInc., Plotly Technologies (2015) Collaborative data science.

Irnov, I.; Sharma, C. M.; Vogel, J. & Winkler, W. C. (2010) Identification of regulatory RNAs in Bacillus subtilis. *Nucleic Acids Res*, 38 (19): 6637-6651.

Isaacs, F. J.; Dwyer, D. J.; Ding, C.; Pervouchine, D. D.; Cantor, C. R. & Collins, J. J. (2004) Engineered riboregulators enable post-transcriptional control of gene expression. *Nat Biotechnol*, 22 (7): 841-847.

Ishikawa, S.; Yamane, K. & Sekiguchi, J. (1998) Regulation and characterization of a newly deduced cell wall hydrolase gene (cwlJ) which affects germination of Bacillus subtilis spores. *J Bacteriol*, 180 (6): 1375-1380.

Isticato, R.; Pelosi, A.; De Felice, M. & Ricca, E. (2010) CotE binds to CotC and CotU and mediates their interaction during spore coat formation in Bacillus subtilis. *J Bacteriol*, 192 (4): 949-954.

Iwanicki, A.; Hinc, K.; Ronowicz, A.; Piotrowski, A.; Woloszyk, A. & Obuchowski, M. (2013) A genome-wide transcriptional profiling of sporulating Bacillus subtilis strain lacking PrpE protein phosphatase. *Mol Genet Genomics*, 288 (10): 469-481.

Kawamoto, H.; Koide, Y.; Morita, T. & Aiba, H. (2006) Base-pairing requirement for RNA silencing by a bacterial small RNA and acceleration of duplex formation by Hfq. *Mol Microbiol*, 61 (4): 1013-1022.

Kawano, M.; Aravind, L. & Storz, G. (2007) An antisense RNA controls synthesis of an SOS-induced toxin evolved from an antitoxin. *Mol Microbiol*, 64 (3): 738-754.

Kawano, M.; Reynolds, A. A.; Miranda-Rios, J. & Storz, G. (2005) Detection of 5'- and 3'-UTR-derived small RNAs and cis-encoded antisense RNAs in Escherichia coli. *Nucleic Acids Res*, 33 (3): 1040-1050.

Kay, E.; Humair, B.; Denervaud, V.; Riedel, K.; Spahr, S.; Eberl, L.; Valverde, C. & Haas, D. (2006) Two GacA-dependent small RNAs modulate the quorum-sensing response in Pseudomonas aeruginosa. *J Bacteriol*, 188 (16): 6026-6033.

Keijser, B. J.; Ter Beek, A.; Rauwerda, H.; Schuren, F.; Montijn, R.; van der Spek, H. & Brul, S. (2007) Analysis of temporal gene expression during Bacillus subtilis spore germination and outgrowth. *J Bacteriol*, 189 (9): 3624-3634.

Kery, M. B.; Feldman, M.; Livny, J. & Tjaden, B. (2014) TargetRNA2: identifying targets of small regulatory RNAs in bacteria. *Nucleic Acids Res*, 42 (Web Server issue): W124-129.

Khan, M. A.; Gopel, Y.; Milewski, S. & Gorke, B. (2016) Two Small RNAs Conserved in Enterobacteriaceae Provide Intrinsic Resistance to Antibiotics Targeting the Cell Wall Biosynthesis Enzyme Glucosamine-6-Phosphate Synthase. *Front Microbiol*, 7 908.

Kiel, J. A.; Boels, J. M.; Beldman, G. & Venema, G. (1994) Glycogen in Bacillus subtilis: molecular characterization of an operon encoding enzymes involved in glycogen biosynthesis and degradation. *Mol Microbiol*, 11 (1): 203-218.

Koike-Takeshita, A.; Koyama, T. & Ogura, K. (1997) Identification of a novel gene cluster participating in menaquinone (vitamin K2) biosynthesis. Cloning and sequence determination of the 2-heptaprenyl-1,4-naphthoquinone

methyltransferase gene of Bacillus stearothermophilus. *J Biol Chem*, 272 (19): 12380-12383.

Koonin, E. V. & Wolf, Y. I. (2008) Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world. *Nucleic Acids Res*, 36 (21): 6688-6719.

Koslover, D. J.; Fazal, F. M.; Mooney, R. A.; Landick, R. & Block, S. M. (2012) Binding and translocation of termination factor rho studied at the single-molecule level. *J Mol Biol*, 423 (5): 664-676.

Krokhotin, A.; Mustoe, A. M.; Weeks, K. M. & Dokholyan, N. V. (2017) Direct identification of base-paired RNA nucleotides by correlated chemical probing. *Rna*, 23 (1): 6-13.

Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A. M.; Alloni, G.; Azevedo, V.; Bertero, M. G.; Bessieres, P.; Bolotin, A.; Borchert, S.; Borriss, R.; Boursier, L.; Brans, A.; Braun, M.; Brignell, S. C.; Bron, S.; Brouillet, S.; Bruschi, C. V.; Caldwell, B.; Capuano, V.; Carter, N. M.; Choi, S. K.; Codani, J. J.; Connerton, I. F.; Danchin, A. & et al. (1997) The complete genome sequence of the gram-positive bacterium Bacillus subtilis. *Nature*, 390 (6657): 249-256.

Kuwana, R.; Kasahara, Y.; Fujibayashi, M.; Takamatsu, H.; Ogasawara, N. & Watabe, K. (2002) Proteomics characterization of novel spore proteins of Bacillus subtilis. *Microbiology*, 148 (Pt 12): 3971-3982.

Laaberki, M. H.; Pfeffer, J.; Clarke, A. J. & Dworkin, J. (2011) O-Acetylation of peptidoglycan is required for proper cell separation and S-layer anchoring in Bacillus anthracis. *J Biol Chem*, 286 (7): 5278-5288.

Lai, E. M.; Phadke, N. D.; Kachman, M. T.; Giorno, R.; Vazquez, S.; Vazquez, J. A.; Maddock, J. R. & Driks, A. (2003) Proteomic analysis of the spore coats of Bacillus subtilis and Bacillus anthracis. *J Bacteriol*, 185 (4): 1443-1454.

Langmead, B. & Salzberg, S. L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods*, 9 (4): 357-359.

Laue, M.; Niederwohrmeier, B. & Bannert, N. (2007) Rapid diagnostic thin section electron microscopy of bacterial endospores. *J Microbiol Methods*, 70 (1): 45-54.

Lease, R. A.; Cusick, M. E. & Belfort, M. (1998) Riboregulation in Escherichia coli: DsrA RNA acts by RNA:RNA interactions at multiple loci. *Proc Natl Acad Sci U S A*, 95 (21): 12456-12461.

Leatherbarrow, A. J.; Yazdi, M. A.; Curson, J. P. & Moir, A. (1998) The gerC locus of Bacillus subtilis, required for menaquinone biosynthesis, is concerned only indirectly with spore germination. *Microbiology*, 144 (Pt 8) 2125-2130.

Lee, C. H.; Wu, T. Y. & Shaw, G. C. (2013) Involvement of OpcR, a GbsR-type transcriptional regulator, in negative regulation of two evolutionarily closely related choline uptake genes in Bacillus subtilis. *Microbiology*, 159 (Pt 10): 2087-2096.

Lenz, D. H.; Miller, M. B.; Zhu, J.; Kulkarni, R. V. & Bassler, B. L. (2005) CsrA and three redundant small RNAs regulate quorum sensing in Vibrio cholerae. *Mol Microbiol*, 58 (4): 1186-1202.

Lenz, D. H.; Mok, K. C.; Lilley, B. N.; Kulkarni, R. V.; Wingreen, N. S. & Bassler, B. L. (2004) The small RNA chaperone Hfq and multiple small RNAs control quorum sensing in Vibrio harveyi and Vibrio cholerae. *Cell*, 118 (1): 69-82.

Levin, P. A. & Losick, R. (1996) Transcription factor Spo0A switches the localization of the cell division protein FtsZ from a medial to a bipolar pattern in Bacillus subtilis. *Genes Dev*, 10 (4): 478-488.

Lewis, J. C.; Snell, N. S. & Burr, H. K. (1960) Water Permeability of Bacterial Spores and the Concept of a Contractile Cortex. *Science*, 132 (3426): 544-545.

Li, G. W.; Burkhardt, D.; Gross, C. & Weissman, J. S. (2014) Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. *Cell*, 157 (3): 624-635.

Licht, A.; Preis, S. & Brantl, S. (2005) Implication of CcpN in the regulation of a novel untranslated RNA (SR1) in Bacillus subtilis. *Mol Microbiol*, 58 (1): 189-206.

Liebermeister, W.; Noor, E.; Flamholz, A.; Davidi, D.; Bernhardt, J. & Milo, R. (2014) Visual account of protein investment in cellular functions. *Proc Natl Acad Sci U S A*, 111 (23): 8488-8493.

Lindgreen, S.; Umu, S. U.; Lai, A. S.; Eldai, H.; Liu, W.; McGimpsey, S.; Wheeler, N. E.; Biggs, P. J.; Thomson, N. R.; Barquist, L.; Poole, A. M. & Gardner, P. P. (2014) Robust identification of noncoding RNA from transcriptomes requires phylogenetically-informed sampling. *PLoS Comput Biol*, 10 (10): e1003907.

Link, T. M.; Valentin-Hansen, P. & Brennan, R. G. (2009) Structure of Escherichia coli Hfq bound to polyriboadenylate RNA. *Proc Natl Acad Sci U S A*, 106 (46): 19292-19297.

Liu, J.; Hanne, J.; Britton, B. M.; Bennett, J.; Kim, D.; Lee, J. B. & Fishel, R. (2016) Cascading MutS and MutL sliding clamps control DNA diffusion to activate mismatch repair. *Nature*, 539 (7630): 583-587.

Liu, M. Y.; Yang, H. & Romeo, T. (1995) The product of the pleiotropic Escherichia coli gene csrA modulates glycogen biosynthesis via effects on mRNA stability. *J Bacteriol*, 177 (10): 2663-2672.

Liu, T.; Zhang, K.; Xu, S.; Wang, Z.; Fu, H.; Tian, B.; Zheng, X. & Li, W. (2017) Detecting RNA-RNA interactions in E. coli using a modified CLASH method. *BMC Genomics*, 18 (1): 343.

Lloyd, C. R.; Park, S.; Fei, J. & Vanderpool, C. K. (2017) The Small Protein SgrT Controls Transport Activity of the Glucose-Specific Phosphotransferase System. *J Bacteriol*, 199 (11):

Lopez, D. & Kolter, R. (2010) Extracellular signals that define distinct and coexisting cell fates in Bacillus subtilis. *FEMS Microbiol Rev*, 34 (2): 134-149.

Lorenz, R.; Bernhart, S. H.; Honer Zu Siederdissen, C.; Tafer, H.; Flamm, C.; Stadler, P. F. & Hofacker, I. L. (2011) ViennaRNA Package 2.0. *Algorithms Mol Biol*, 6 26.

Lott, S.; Mann, M. ; Hess, W. ; Voss, B. & Georg, J. (2017) GLASSgo - Automated and reliable detection of sRNA homologs from a single input sequence. *Nucleic Acids Res*, (under review)

Ludwig, H.; Homuth, G.; Schmalisch, M.; Dyka, F. M.; Hecker, M. & Stulke, J. (2001) Transcription of glycolytic genes and operons in Bacillus subtilis: evidence for the presence of multiple levels of control of the gapA operon. *Mol Microbiol*, 41 (2): 409-422.

Lugtenberg, B.; Peters, R.; Bernheimer, H. & Berendsen, W. (1976) Influence of cultural conditions and mutations on the composition of the outer membrane proteins of Escherichia coli. *Mol Gen Genet*, 147 (3): 251-262.

Lustig, Y.; Wachtel, C.; Safro, M.; Liu, L. & Michaeli, S. (2010) 'RNA walk' a novel approach to study RNA-RNA interactions between a small RNA and its target. *Nucleic Acids Res*, 38 (1): e5.

Lybecker, M.; Bilusic, I. & Raghavan, R. (2014) Pervasive transcription: detecting functional RNAs in bacteria. *Transcription*, 5 (4): e944039.

Ma, P.; Patching, S. G.; Ivanova, E.; Baldwin, J. M.; Sharples, D.; Baldwin, S. A. & Henderson, P. J. (2016) Allantoin transport protein, PucI, from Bacillus subtilis: evolutionary relationships, amplified expression, activity and specificity. *Microbiology*, 162 (5): 823-836.

Mader, U.; Antelmann, H.; Buder, T.; Dahl, M. K.; Hecker, M. & Homuth, G. (2002a) Bacillus subtilis functional genomics: genome-wide analysis of the DegS-DegU regulon by transcriptomics and proteomics. *Mol Genet Genomics*, 268 (4): 455-467.

Mader, U.; Homuth, G.; Scharf, C.; Buttner, K.; Bode, R. & Hecker, M. (2002b) Transcriptome and proteome analysis of Bacillus subtilis gene expression modulated by amino acid availability. *J Bacteriol*, 184 (15): 4288-4295.

Majdalani, N.; Hernandez, D. & Gottesman, S. (2002) Regulation and mode of action of the second small RNA activator of RpoS translation, RprA. *Mol Microbiol*, 46 (3): 813-826.

Maki, K.; Morita, T.; Otaka, H. & Aiba, H. (2010) A minimal base-pairing region of a bacterial small RNA SgrS required for translational repression of ptsG mRNA. *Mol Microbiol*, 76 (3): 782-792.

Mandal, M.; Boese, B.; Barrick, J. E.; Winkler, W. C. & Breaker, R. R. (2003) Riboswitches control fundamental biochemical pathways in Bacillus subtilis and other bacteria. *Cell*, 113 (5): 577-586.

Mann, M. (2006) Functional and quantitative proteomics using SILAC. *Nat Rev Mol Cell Biol*, 7 (12): 952-958.

Mao, L.; Jiang, S.; Wang, B.; Chen, L.; Yao, Q. & Chen, K. (2011) Protein profile of Bacillus subtilis spore. *Curr Microbiol*, 63 (2): 198-205.

Marchais, A.; Duperrier, S.; Durand, S.; Gautheret, D. & Stragier, P. (2011) CsfG, a sporulation-specific, small non-coding RNA highly conserved in endospore formers. *RNA Biol*, 8 (3): 358-364.

Mars, R. A.; Mendonca, K.; Denham, E. L. & van Dijl, J. M. (2015a) The reduction in small ribosomal subunit abundance in ethanol-stressed cells of Bacillus subtilis is mediated by a SigB-dependent antisense RNA. *Biochim Biophys Acta*, 1853 (10 Pt A): 2553-2559.

Mars, R. A.; Nicolas, P.; Ciccolini, M.; Reilman, E.; Reder, A.; Schaffer, M.; Mader, U.; Volker, U.; van Dijl, J. M. & Denham, E. L. (2015b) Small regulatory RNA-induced growth rate heterogeneity of Bacillus subtilis. *PLoS Genet*, 11 (3): e1005046.

Mars, R. A.; Nicolas, P.; Denham, E. L. & van Dijl, J. M. (2016) Regulatory RNAs in Bacillus subtilis: a Gram-Positive Perspective on Bacterial RNA-Mediated Regulation of Gene Expression. *Microbiol Mol Biol Rev*, 80 (4): 1029-1057.

Martinussen, J.; Glaser, P.; Andersen, P. S. & Saxild, H. H. (1995) Two genes encoding uracil phosphoribosyltransferase are present in Bacillus subtilis. *J Bacteriol*, 177 (1): 271-274.

Masse, E. & Gottesman, S. (2002) A small RNA regulates the expression of genes involved in iron metabolism in Escherichia coli. *Proc Natl Acad Sci U S A*, 99 (7): 4620-4625.

Matsuno, K.; Blais, T.; Serio, A. W.; Conway, T.; Henkin, T. M. & Sonenshein, A. L. (1999) Metabolic imbalance and sporulation in an isocitrate dehydrogenase mutant of Bacillus subtilis. *J Bacteriol*, 181 (11): 3382-3391.

Maughan, H. & Nicholson, W. L. (2004) Stochastic processes influence stationary-phase decisions in Bacillus subtilis. *J Bacteriol*, 186 (7): 2212-2214.

McKenney, P. T.; Driks, A.; Eskandarian, H. A.; Grabowski, P.; Guberman, J.; Wang, K. H.; Gitai, Z. & Eichenberger, P. (2010) A distance-weighted interaction map reveals a previously uncharacterized layer of the Bacillus subtilis spore coat. *Curr Biol*, 20 (10): 934-938.

Meeske, A. J.; Rodrigues, C. D.; Brady, J.; Lim, H. C.; Bernhardt, T. G. & Rudner, D. Z. (2016) High-Throughput Genetic Screens Identify a Large and Diverse Collection of New Sporulation Genes in Bacillus subtilis. *PLoS Biol*, 14 (1): e1002341.

Melamed, S.; Peer, A.; Faigenbaum-Romm, R.; Gatt, Y. E.; Reiss, N.; Bar, A.; Altuvia, Y.; Argaman, L. & Margalit, H. (2016) Global Mapping of Small RNA-Target Interactions in Bacteria. *Mol Cell*, 63 (5): 884-897.

Melly, E.; Cowan, A. E. & Setlow, P. (2002) Studies on the mechanism of killing of Bacillus subtilis spores by hydrogen peroxide. *J Appl Microbiol*, 93 (2): 316-325.

Meng, Y.; Shao, C.; Wang, H. & Jin, Y. (2012) Target mimics: an embedded layer of microRNA-involved gene regulatory networks in plants. *BMC Genomics*, 13 197.

Michna, R. H.; Zhu, B.; Mader, U. & Stulke, J. (2016) SubtiWiki 2.0--an integrated database for the model organism Bacillus subtilis. *Nucleic Acids Res*, 44 (D1): D654-662.

Mikulecky, P. J.; Kaw, M. K.; Brescia, C. C.; Takach, J. C.; Sledjeski, D. D. & Feig, A. L. (2004) Escherichia coli Hfq has distinct interaction surfaces for DsrA, rpoS and poly(A) RNAs. *Nat Struct Mol Biol*, 11 (12): 1206-1214.

Miller, A. K.; Brown, E. E.; Mercado, B. T. & Herman, J. K. (2016) A DNA-binding protein defines the precise region of chromosome capture during Bacillus sporulation. *Mol Microbiol*, 99 (1): 111-122.

Miyakoshi, M.; Chao, Y. & Vogel, J. (2015) Cross talk between ABC transporter mRNAs via a target mRNA-derived sponge of the GcvB small RNA. *Embo j*, 34 (11): 1478-1492.

Mizuno, T.; Chou, M. Y. & Inouye, M. (1984) A unique mechanism regulating gene expression: translational inhibition by a complementary RNA transcript (micRNA). *Proc Natl Acad Sci U S A*, 81 (7): 1966-1970.

Moeller, R.; Raguse, M.; Reitz, G.; Okayasu, R.; Li, Z.; Klein, S.; Setlow, P. & Nicholson, W. L. (2014) Resistance of Bacillus subtilis spore DNA to lethal ionizing radiation damage relies primarily on spore core components and DNA repair, with minor effects of oxygen radical detoxification. *Appl Environ Microbiol*, 80 (1): 104-109.

Moeller, R.; Setlow, P.; Pedraza-Reyes, M.; Okayasu, R.; Reitz, G. & Nicholson, W. L. (2011) Role of the Nfo and ExoA apurinic/apyrimidinic endonucleases in radiation resistance and radiation-induced mutagenesis of Bacillus subtilis spores. *J Bacteriol*, 193 (11): 2875-2879.

Moir, A. (2006) How do spores germinate? *J Appl Microbiol*, 101 (3): 526-530.

Molle, V.; Fujita, M.; Jensen, S. T.; Eichenberger, P.; Gonzalez-Pastor, J. E.; Liu, J. S. & Losick, R. (2003) The Spo0A regulon of Bacillus subtilis. *Mol Microbiol*, 50 (5): 1683-1701.

Moller, T.; Franch, T.; Hojrup, P.; Keene, D. R.; Bachinger, H. P.; Brennan, R. G. & Valentin-Hansen, P. (2002) Hfq: a bacterial Sm-like protein that mediates RNA-RNA interaction. *Mol Cell*, 9 (1): 23-30.

Montange, R. K. & Batey, R. T. (2008) Riboswitches: emerging themes in RNA structure and function. *Annu Rev Biophys*, 37 117-133.

Morinaga, T.; Kobayashi, K.; Ashida, H.; Fujita, Y. & Yoshida, K. (2010) Transcriptional regulation of the Bacillus subtilis asnH operon and role of the 5'-proximal long sequence triplication in RNA stabilization. *Microbiology*, 156 (Pt 6): 1632-1641.

Morita, T.; Maki, K. & Aiba, H. (2005) RNase E-based ribonucleoprotein complexes: mechanical basis of mRNA destabilization mediated by bacterial noncoding RNAs. *Genes Dev*, 19 (18): 2176-2186.

Mukherjee, S.; Yakhnin, H.; Kysela, D.; Sokoloski, J.; Babitzke, P. & Kearns, D. B. (2011) CsrA-FliW interaction governs flagellin homeostasis and a checkpoint on flagellar morphogenesis in Bacillus subtilis. *Mol Microbiol*, 82 (2): 447-461.

Muntel, J.; Fromion, V.; Goelzer, A.; Maabeta, S.; Mader, U.; Buttner, K.; Hecker, M. & Becher, D. (2014) Comprehensive absolute quantification of the cytosolic proteome of Bacillus subtilis by data independent, parallel fragmentation in liquid chromatography/mass spectrometry (LC/MS(E)). *Mol Cell Proteomics*, 13 (4): 1008-1019.

Nagler, K.; Krawczyk, A. O.; De Jong, A.; Madela, K.; Hoffmann, T.; Laue, M.; Kuipers, O. P.; Bremer, E. & Moeller, R. (2016) Identification of Differentially Expressed Genes during Bacillus subtilis Spore Outgrowth in High-Salinity Environments Using RNA Sequencing. *Front Microbiol*, 7 1564.

Nakano, M. M. & Zuber, P. (1998) Anaerobic growth of a "strict aerobe" (Bacillus subtilis). *Annu Rev Microbiol*, 52 165-190.

Nakashio, S. & Gerhardt, P. (1985) Protoplast dehydration correlated with heat resistance of bacterial spores. *J Bacteriol*, 162 (2): 571-578.

Narula, J.; Devi, S. N.; Fujita, M. & Igoshin, O. A. (2012) Ultrasensitivity of the Bacillus subtilis sporulation decision. *Proc Natl Acad Sci U S A*, 109 (50): E3513-3522.

Naville, M.; Ghuillot-Gaudeffroy, A.; Marchais, A. & Gautheret, D. (2011) ARNold: a web tool for the prediction of Rho-independent transcription terminators. *RNA Biol*, 8 (1): 11-13.

Nicolas, P.; Mader, U.; Dervyn, E.; Rochat, T.; Leduc, A.; Pigeonneau, N.; Bidnenko, E.; Marchadier, E.; Hoebeke, M.; Aymerich, S.; Becher, D.; Bisicchia, P.; Botella, E.; Delumeau, O.; Doherty, G.; Denham, E. L.; Fogg, M. J.; Fromion, V.; Goelzer, A.; Hansen, A.; Hartig, E.; Harwood, C. R.; Homuth, G.; Jarmer, H.; Jules, M.; Klipp, E.; Le Chat, L.; Lecointe, F.; Lewis, P.; Liebermeister, W.; March, A.; Mars, R. A.; Nannapaneni, P.; Noone, D.; Pohl, S.; Rinn, B.; Rugheimer, F.; Sappa, P. K.; Samson, F.; Schaffer, M.; Schwikowski, B.; Steil, L.; Stulke, J.; Wiegert, T.; Devine, K. M.; Wilkinson, A. J.; van Dijl, J. M.; Hecker, M.; Volker, U.; Bessieres, P. & Noirot, P. (2012) Condition-dependent transcriptome reveals high-level regulatory architecture in Bacillus subtilis. *Science*, 335 (6072): 1103-1106.

Nishiguchi, M.; Honda, K.; Amikura, R.; Nakamura, K. & Yamane, K. (1994) Structural requirements of Bacillus subtilis small cytoplasmic RNA for cell growth, sporulation, and extracellular enzyme production. *J Bacteriol*, 176 (1): 157-165.

Ohki, R.; Tateno, K.; Takizawa, T.; Aiso, T. & Murata, M. (2005) Transcriptional termination control of a novel ABC transporter gene involved in antibiotic resistance in Bacillus subtilis. *J Bacteriol*, 187 (17): 5946-5954.

Onyenwoke, R. U.; Brill, J. A.; Farahi, K. & Wiegel, J. (2004) Sporulation genes in members of the low G+C Gram-type-positive phylogenetic branch (Firmicutes). *Arch Microbiol*, 182 (2-3): 182-192.

Opdyke, J. A.; Fozo, E. M.; Hemm, M. R. & Storz, G. (2011) RNase III participates in GadY-dependent cleavage of the gadX-gadW mRNA. *J Mol Biol*, 406 (1): 29-43.

Opdyke, J. A.; Kang, J. G. & Storz, G. (2004) GadY, a small-RNA regulator of acid response genes in Escherichia coli. *J Bacteriol*, 186 (20): 6698-6705.

Otaka, H.; Ishikawa, H.; Morita, T. & Aiba, H. (2011) PolyU tail of rho-independent terminator of bacterial small RNAs is essential for Hfq action. *Proc Natl Acad Sci U S A*, 108 (32): 13059-13064.

Overgaard, M.; Johansen, J.; Moller-Jensen, J. & Valentin-Hansen, P. (2009) Switching off small RNA regulation with trap-mRNA. *Mol Microbiol*, 73 (5): 790-800.

P, Setlow & A, Kornberg (1970) Biochemical studies of bacterial sporulation and germination. *J Biol Chem*, 245 (14): 3645-3652.

Paget, M. S. (2015) Bacterial Sigma Factors and Anti-Sigma Factors: Structure, Function and Distribution. *Biomolecules*, 5 (3): 1245-1265.

Paidhungat, M.; Ragkousi, K. & Setlow, P. (2001) Genetic requirements for induction of germination of spores of Bacillus subtilis by Ca(2+)-dipicolinate. *J Bacteriol*, 183 (16): 4886-4893.

Paidhungat, M.; Setlow, B.; Driks, A. & Setlow, P. (2000) Characterization of spores of Bacillus subtilis which lack dipicolinic acid. *J Bacteriol*, 182 (19): 5505-5512.

Paidhungat, M. & Setlow, P. (2000) Role of ger proteins in nutrient and nonnutrient triggering of spore germination in Bacillus subtilis. *J Bacteriol*, 182 (9): 2513-2519.

Paidhungat, M. & Setlow, P. (2001) Localization of a germinant receptor protein (GerBA) to the inner membrane of Bacillus subtilis spores. *J Bacteriol*, 183 (13): 3982-3990.

Pandey, NK & Aronson, AI (1979) - Properties of the Bacillus subtilis spore coat. *J Bacteriol.*, 137 (3): 1208-1218.

Papenfort, K.; Podkaminski, D.; Hinton, J. C. & Vogel, J. (2012) The ancestral SgrS RNA discriminates horizontally acquired Salmonella mRNAs through a single G-U wobble pair. *Proc Natl Acad Sci U S A*, 109 (13): E757-764.

Papenfort, K.; Sun, Y.; Miyakoshi, M.; Vanderpool, C. K. & Vogel, J. (2013) Small RNA-mediated activation of sugar phosphatase mRNA regulates glucose homeostasis. *Cell*, 153 (2): 426-437.

Papenfort, K. & Vogel, J. (2009) Multiple target regulation by small noncoding RNAs rewires gene expression at the post-transcriptional level. *Res Microbiol*, 160 (4): 278-287.

Paredes-Sabja, D.; Setlow, P. & Sarker, M. R. (2011) Germination of spores of Bacillales and Clostridiales species: mechanisms and proteins involved. *Trends Microbiol*, 19 (2): 85-94.

Parveen, S. & Reddy, M. (2017) Identification of YfiH (PgeF) as a factor contributing to the maintenance of bacterial peptidoglycan composition. *Mol Microbiol*,

Peer, A. & Margalit, H. (2014) Evolutionary patterns of Escherichia coli small RNAs and their regulatory interactions. *Rna*, 20 (7): 994-1003.

Pellegrini, O.; Mathy, N.; Gogos, A.; Shapiro, L. & Condon, C. (2005) The Bacillus subtilis ydcDE operon encodes an endoribonuclease of the MazF/PemK family and its inhibitor. *Mol Microbiol*, 56 (5): 1139-1148.

Perez-Valdespino, A.; Ghosh, S.; Cammett, E. P.; Kong, L.; Li, Y. Q. & Setlow, P. (2013) Isolation and characterization of Bacillus subtilis spores that are superdormant for germination with dodecylamine or Ca2+ -dipicolinic acid. *J Appl Microbiol*, 114 (4): 1109-1119.

Plomp, M.; Carroll, A. M.; Setlow, P. & Malkin, A. J. (2014) Architecture and assembly of the Bacillus subtilis spore coat. *PLoS One*, 9 (9): e108560.

Popham, D. L.; Gilmore, M. E. & Setlow, P. (1999) Roles of low-molecular-weight penicillin-binding proteins in Bacillus subtilis spore peptidoglycan synthesis and spore properties. *J Bacteriol*, 181 (1): 126-132.

Preis, H.; Eckart, R. A.; Gudipati, R. K.; Heidrich, N. & Brantl, S. (2009) CodY activates transcription of a small RNA in Bacillus subtilis. *J Bacteriol*, 191 (17): 5446-5457.

Quay, D. H.; Cole, A. R.; Cryar, A.; Thalassinos, K.; Williams, M. A.; Bhakta, S. & Keep, N. H. (2015) Structure of the stationary phase survival protein YuiC from B.subtilis. *BMC Struct Biol*, 15 12.

Quirk, P. G.; Dunkley, E. A., Jr.; Lee, P. & Krulwich, T. A. (1993) Identification of a putative Bacillus subtilis rho gene. *J Bacteriol*, 175 (3): 647-654.

Rajkowitsch, L. & Schroeder, R. (2007) Coupling RNA annealing and strand displacement: a FRET-based microplate reader assay for RNA chaperone activity. *Biotechniques*, 43 (3): 304, 306, 308 passim.

Rasmussen, S.; Nielsen, H. B. & Jarmer, H. (2009) The transcriptionally active regions in the genome of Bacillus subtilis. *Mol Microbiol*, 73 (6): 1043-1057.

Ratnayake-Lecamwasam, M.; Serror, P.; Wong, K. W. & Sonenshein, A. L. (2001) Bacillus subtilis CodY represses early-stationary-phase genes by sensing GTP levels. *Genes Dev*, 15 (9): 1093-1103.

Ray-Soni, A.; Bellecourt, M. J. & Landick, R. (2016) Mechanisms of Bacterial Transcription Termination: All Good Things Must End. *Annu Rev Biochem*, 85 319-347.

Reeves, A.; Martinez, L. & Haldenwang, W. (2010) Expression of, and in vivo stressosome formation by, single members of the RsbR protein family in Bacillus subtilis. *Microbiology*, 156 (Pt 4): 990-998.

Reuss, D. R.; Commichau, F. M.; Gundlach, J.; Zhu, B. & Stulke, J. (2016) The Blueprint of a Minimal Cell: MiniBacillus. *Microbiol Mol Biol Rev*, 80 (4): 955-987.

Richards, G. R. & Vanderpool, C. K. (2011) Molecular call and response: the physiology of bacterial small RNAs. *Biochim Biophys Acta*, 1809 (10): 525-531.

Rivas, E. & Eddy, S. R. (2000) Secondary structure alone is generally not statistically significant for the detection of noncoding RNAs. *Bioinformatics*, 16 (7): 583-605.

Rivas, E. & Eddy, S. R. (2001) Noncoding RNA gene detection using comparative sequence analysis. *BMC Bioinformatics*, 2 8.

Rivas, E.; Klein, R. J.; Jones, T. A. & Eddy, S. R. (2001) Computational identification of noncoding RNAs in E. coli by comparative genomics. *Curr Biol*, 11 (17): 1369-1373.

Roberfroid, S.; Vanderleyden, J. & Steenackers, H. (2016) Gene expression variability in clonal populations: Causes and consequences. *Crit Rev Microbiol*, 42 (6): 969-984.

Romero, D. (2013) Bacterial determinants of the social behavior of Bacillus subtilis. *Res Microbiol*, 164 (7): 788-798.

Ross, C. & Abel-Santos, E. (2010) The Ger receptor family from sporulating bacteria. *Curr Issues Mol Biol*, 12 (3): 147-158.

Rudolph, M. J.; Wuebbens, M. M.; Turque, O.; Rajagopalan, K. V. & Schindelin, H. (2003) Structural studies of molybdopterin synthase provide insights into its catalytic mechanism. *J Biol Chem*, 278 (16): 14514-14522.

Saito, S.; Kakeshita, H. & Nakamura, K. (2009) Novel small RNA-encoding genes in the intergenic regions of Bacillus subtilis. *Gene*, 428 (1-2): 2-8.

Saks, M. E.; Sampson, J. R. & Abelson, J. N. (1994) The transfer RNA identity problem: a search for rules. *Science*, 263 (5144): 191-197.

Santos, C. L.; Tavares, F.; Thioulouse, J. & Normand, P. (2009) A phylogenomic analysis of bacterial helix-turn-helix transcription factors. *FEMS Microbiol Rev*, 33 (2): 411-429.

Sauer, E.; Schmidt, S. & Weichenrieder, O. (2012) Small RNA binding to the lateral surface of Hfq hexamers and structural rearrangements upon mRNA target recognition. *Proc Natl Acad Sci U S A*, 109 (24): 9396-9401.

Sauter, C.; Basquin, J. & Suck, D. (2003) Sm-like proteins in Eubacteria: the crystal structure of the Hfq protein from Escherichia coli. *Nucleic Acids Res*, 31 (14): 4091-4098.

Savill, N. J.; Hoyle, D. C. & Higgs, P. G. (2001) RNA sequence evolution with secondary structure constraints: comparison of substitution rate models using maximum-likelihood methods. *Genetics*, 157 (1): 399-411.

Savinov, A.; Perez, C. F. & Block, S. M. (2014) Single-molecule studies of riboswitch folding. *Biochim Biophys Acta*, 1839 (10): 1030-1045.

Schmalisch, M.; Maiques, E.; Nikolov, L.; Camp, A. H.; Chevreux, B.; Muffler, A.; Rodriguez, S.; Perkins, J. & Losick, R. (2010) Small genes under sporulation control in the Bacillus subtilis genome. *J Bacteriol*, 192 (20): 5402-5412.

Schmidt, R.; Margolis, P.; Duncan, L.; Coppolecchia, R.; Moran, C. P., Jr. & Losick, R. (1990) Control of developmental transcription factor sigma F by sporulation regulatory proteins SpoIIAA and SpoIIAB in Bacillus subtilis. *Proc Natl Acad Sci U S A*, 87 (23): 9221-9225.

Schu, D. J.; Zhang, A.; Gottesman, S. & Storz, G. (2015) Alternative Hfq-sRNA interaction modes dictate alternative mRNA recognition. *Embo j*, 34 (20): 2557-2573.

Schultz, D.; Wolynes, P. G.; Ben Jacob, E. & Onuchic, J. N. (2009) Deciding fate in adverse times: sporulation and competence in Bacillus subtilis. *Proc Natl Acad Sci U S A*, 106 (50): 21027-21034.

Segev, E.; Rosenberg, A.; Mamou, G.; Sinai, L. & Ben-Yehuda, S. (2013) Molecular kinetics of reviving bacterial spores. *J Bacteriol*, 195 (9): 1875-1882.

Segev, E.; Smith, Y. & Ben-Yehuda, S. (2012) RNA dynamics in aging bacterial spores. *Cell*, 148 (1-2): 139-149.

Serio, A. W.; Pechter, K. B. & Sonenshein, A. L. (2006) Bacillus subtilis aconitase is required for efficient late-sporulation gene expression. *J Bacteriol*, 188 (17): 6396-6405.

Serrano, M.; Zilhao, R.; Ricca, E.; Ozin, A. J.; Moran, C. P., Jr. & Henriques, A. O. (1999) A Bacillus subtilis secreted protein with a role in endospore coat assembly and function. *J Bacteriol*, 181 (12): 3632-3643.

Setlow, B.; Cowan, A. E. & Setlow, P. (2003) Germination of spores of Bacillus subtilis with dodecylamine. *J Appl Microbiol*, 95 (3): 637-648.

Setlow, P. (2001) Resistance of spores of Bacillus species to ultraviolet light. *Environ Mol Mutagen*, 38 (2-3): 97-104.

Setlow, P. (2006) Spores of Bacillus subtilis: their resistance to and killing by radiation, heat and chemicals. *J Appl Microbiol*, 101 (3): 514-525.

Setlow, P. (2007) I will survive: DNA protection in bacterial spores. *Trends Microbiol*, 15 (4): 172-180.

Setlow, P. (2014) Germination of spores of Bacillus species: what we know and do not know. *J Bacteriol*, 196 (7): 1297-1305.

Sharma, C. M.; Hoffmann, S.; Darfeuille, F.; Reignier, J.; Findeiss, S.; Sittka, A.; Chabas, S.; Reiche, K.; Hackermuller, J.; Reinhardt, R.; Stadler, P. F. & Vogel, J. (2010) The

primary transcriptome of the major human pathogen Helicobacter pylori. *Nature*, 464 (7286): 250-255.

Sharma, C. M.; Papenfort, K.; Pernitzsch, S. R.; Mollenkopf, H. J.; Hinton, J. C. & Vogel, J. (2011) Pervasive post-transcriptional control of genes involved in amino acid metabolism by the Hfq-dependent GcvB small RNA. *Mol Microbiol*, 81 (5): 1144-1165.

Sharma, E.; Sterne-Weiler, T.; O'Hanlon, D. & Blencowe, B. J. (2016) Global Mapping of Human RNA-RNA Interactions. *Mol Cell*, 62 (4): 618-626.

Sierro, N.; Makita, Y.; de Hoon, M. & Nakai, K. (2008) DBTBS: a database of transcriptional regulation in Bacillus subtilis containing upstream intergenic conservation information. *Nucleic Acids Res*, 36 (Database issue): D93-96.

Silvaggi, J. M.; Perkins, J. B. & Losick, R. (2005) Small untranslated RNA antitoxin in Bacillus subtilis. *J Bacteriol*, 187 (19): 6641-6650.

Silvaggi, J. M.; Perkins, J. B. & Losick, R. (2006) Genes for small, noncoding RNAs under sporulation control in Bacillus subtilis. *J Bacteriol*, 188 (2): 532-541.

Simons, R. W. & Kleckner, N. (1983) Translational control of IS10 transposition. *Cell*, 34 (2): 683-691.

Sinai, L.; Rosenberg, A.; Smith, Y.; Segev, E. & Ben-Yehuda, S. (2015) The molecular timeline of a reviving bacterial spore. *Mol Cell*, 57 (4): 695-707.

Sittka, A.; Pfeiffer, V.; Tedin, K. & Vogel, J. (2007) The RNA chaperone Hfq is essential for the virulence of Salmonella typhimurium. *Mol Microbiol*, 63 (1): 193-217.

Sledjeski, D. D.; Whitman, C. & Zhang, A. (2001) Hfq is necessary for regulation by the untranslated RNA DsrA. *J Bacteriol*, 183 (6): 1997-2005.

Smirnov, A.; Forstner, K. U.; Holmqvist, E.; Otto, A.; Gunster, R.; Becher, D.; Reinhardt, R. & Vogel, J. (2016) Grad-seq guides the discovery of ProQ as a major small RNA-binding protein. *Proc Natl Acad Sci U S A*, 113 (41): 11591-11596.

Smits, W. K.; Bongiorni, C.; Veening, J. W.; Hamoen, L. W.; Kuipers, O. P. & Perego, M. (2007) Temporal separation of distinct differentiation pathways by a dual specificity Rap-Phr system in Bacillus subtilis. *Mol Microbiol*, 65 (1): 103-120.

Soper, T.; Mandin, P.; Majdalani, N.; Gottesman, S. & Woodson, S. A. (2010) Positive regulation by small RNAs and the role of Hfq. *Proc Natl Acad Sci U S A*, 107 (21): 9602-9607.

Southworth, T. W.; Guffanti, A. A.; Moir, A. & Krulwich, T. A. (2001) GerN, an Endospore Germination Protein of Bacillus cereus, Is an Na(+)/H(+)-K(+) Antiporter. *J Bacteriol*, 183 (20): 5896-5903.

Steil, L.; Serrano, M.; Henriques, A. O. & Volker, U. (2005) Genome-wide analysis of temporally regulated and compartment-specific gene expression in sporulating cells of Bacillus subtilis. *Microbiology*, 151 (Pt 2): 399-420.

Sterlini, J. M. & Mandelstam, J. (1969) Commitment to sporulation in Bacillus subtilis and its relationship to development of actinomycin resistance. *Biochem J*, 113 (1): 29-37.

Stewart, G. C. (2015) The Exosporium Layer of Bacterial Spores: a Connection to the Environment and the Infected Host. *Microbiol Mol Biol Rev*, 79 (4): 437-457.

Stewart, K. A.; Yi, X.; Ghosh, S. & Setlow, P. (2012) Germination protein levels and rates of germination of spores of Bacillus subtilis with overexpressed or deleted genes encoding germination proteins. *J Bacteriol*, 194 (12): 3156-3164.

Stougaard, P.; Molin, S. & Nordstrom, K. (1981) RNAs involved in copy-number control and incompatibility of plasmid R1. *Proc Natl Acad Sci U S A*, 78 (10): 6008-6012.

Stragier, P. & Losick, R. (1990) Cascades of sigma factors revisited. *Mol Microbiol*, 4 (11): 1801-1806.

Stragier, P. & Losick, R. (1996) Molecular genetics of sporulation in Bacillus subtilis. *Annu Rev Genet*, 30 297-241.

Suel, G. M.; Garcia-Ojalvo, J.; Liberman, L. M. & Elowitz, M. B. (2006) An excitable gene regulatory circuit induces transient cellular differentiation. *Nature*, 440 (7083): 545-550.

Sun, X.; Zhulin, I. & Wartell, R. M. (2002) Predicted structure and phyletic distribution of the RNA-binding protein Hfq. *Nucleic Acids Res*, 30 (17): 3662-3671.

Sunde, E. P.; Setlow, P.; Hederstedt, L. & Halle, B. (2009) The physical state of water in bacterial spores. *Proc Natl Acad Sci U S A*, 106 (46): 19334-19339.

Swerdlow, B. M.; Setlow, B. & Setlow, P. (1981) Levels of H+ and other monovalent cations in dormant and germinating spores of Bacillus megaterium. *J Bacteriol*, 148 (1): 20-29.

Tafer, H. & Hofacker, I. L. (2008) RNAplex: a fast tool for RNA-RNA interaction search. *Bioinformatics*, 24 (22): 2657-2663.

Takahashi, Y.; Koyama, K. & Ito, M. (2014) Suppressor mutants from MotB-D24E and MotS-D30E in the flagellar stator complex of Bacillus subtilis. *J Gen Appl Microbiol*, 60 (4): 131-139.

Takamatsu, H.; Kodama, T.; Imamura, A.; Asai, K.; Kobayashi, K.; Nakayama, T.; Ogasawara, N. & Watabe, K. (2000) The Bacillus subtilis yabG gene is transcribed by SigK RNA polymerase during sporulation, and yabG mutant spores have altered coat protein composition. *J Bacteriol*, 182 (7): 1883-1888.

Tanaka, K.; Henry, C. S.; Zinner, J. F.; Jolivet, E.; Cohoon, M. P.; Xia, F.; Bidnenko, V.; Ehrlich, S. D.; Stevens, R. L. & Noirot, P. (2013) Building the repertoire of dispensable chromosome regions in Bacillus subtilis entails major refinement of cognate large-scale metabolic model. *Nucleic Acids Res*, 41 (1): 687-699.

Tani, K.; Watanabe, T.; Matsuda, H.; Nasu, M. & Kondo, M. (1996) Cloning and sequencing of the spore germination gene of Bacillus megaterium ATCC 12872: similarities to the NaH-antiporter gene of Enterococcus hirae. *Microbiol Immunol*, 40 (2): 99-105.

Tavares, M. B.; Souza, R. D.; Luiz, W. B.; Cavalcante, R. C.; Casaroli, C.; Martins, E. G.; Ferreira, R. C. & Ferreira, L. C. (2013) Bacillus subtilis endospores at high purity and recovery yields: optimization of growth conditions and purification method. *Curr Microbiol*, 66 (3): 279-285.

Thackray, P. D.; Behravan, J.; Southworth, T. W. & Moir, A. (2001) GerN, an Antiporter Homologue Important in Germination of Bacillus cereus Endospores. *J Bacteriol*, 183 (2): 476-482.

Thomas, P.; Sekhar, A. C.; Upreti, R.; Mujawar, M. M. & Pasha, S. S. (2015) Optimization of single plate-serial dilution spotting (SP-SDS) with sample anchoring as an assured method for bacterial and yeast cfu enumeration and single colony isolation from diverse samples. *Biotechnol Rep (Amst)*, 8 45-55.

Thompson, B. M.; Binkley, J. M. & Stewart, G. C. (2011) Current physical and SDS extraction methods do not efficiently remove exosporium proteins from Bacillus anthracis spores. *J Microbiol Methods*, 85 (2): 143-148.

Timmermans, J. & Van Melderen, L. (2010) Post-transcriptional global regulation by CsrA in bacteria. *Cell Mol Life Sci*, 67 (17): 2897-2908.

Tjaden, B. (2008) TargetRNA: a tool for predicting targets of small RNA action in bacteria. *Nucleic Acids Res*, 36 (Web Server issue): W109-113.

Tjaden, B.; Goodwin, S. S.; Opdyke, J. A.; Guillier, M.; Fu, D. X.; Gottesman, S. & Storz, G. (2006) Target prediction for small, noncoding RNAs in bacteria. *Nucleic Acids Res*, 34 (9): 2791-2802.

Tomizawa, J.; Itoh, T.; Selzer, G. & Som, T. (1981) Inhibition of ColE1 RNA primer formation by a plasmid-specified small RNA. *Proc Natl Acad Sci U S A*, 78 (3): 1421-1425.

Traag, B. A.; Pugliese, A.; Eisen, J. A. & Losick, R. (2013) Gene conservation among endospore-forming bacteria reveals additional sporulation genes in Bacillus subtilis. *J Bacteriol*, 195 (2): 253-260.

Trausch, J. J.; Ceres, P.; Reyes, F. E. & Batey, R. T. (2011) The structure of a tetrahydrofolate-sensing riboswitch reveals two ligand binding sites in a single aptamer. *Structure*, 19 (10): 1413-1423.

Tsui, H. C.; Leung, H. C. & Winkler, M. E. (1994) Characterization of broadly pleiotropic phenotypes caused by an hfq insertion mutation in Escherichia coli K-12. *Mol Microbiol*, 13 (1): 35-49.

Tyanova, S.; Temu, T.; Sinitcyn, P.; Carlson, A.; Hein, M. Y.; Geiger, T.; Mann, M. & Cox, J. (2016) The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat Methods*, 13 (9): 731-740.

Udekwu, K. I.; Darfeuille, F.; Vogel, J.; Reimegard, J.; Holmqvist, E. & Wagner, E. G. (2005) Hfq-dependent regulation of OmpA synthesis is mediated by an antisense RNA. *Genes Dev*, 19 (19): 2355-2366.

Urban, J. H. & Vogel, J. (2007) Translational control and target recognition by Escherichia coli small RNAs in vivo. *Nucleic Acids Res*, 35 (3): 1018-1037.

van Opijnen, T.; Bodi, K. L. & Camilli, A. (2009) Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. *Nat Methods*, 6 (10): 767-772.

van Opijnen, T. & Camilli, A. (2013) Transposon insertion sequencing: a new tool for systems-level analysis of microorganisms. *Nat Rev Microbiol*, 11 (7): 435-442.

van Opijnen, T.; Lazinski, D. W. & Camilli, A. (2015) Genome-Wide Fitness and Genetic Interactions Determined by Tn-seq, a High-Throughput Massively Parallel Sequencing Method for Microorganisms. *Curr Protoc Microbiol*, 36 1e.3.1-24.

Vanderpool, C. K. (2011) Combined experimental and computational strategies define an expansive regulon for GcvB small RNA. *Mol Microbiol*, 81 (5): 1129-1132.

Vanderpool, C. K. & Gottesman, S. (2004) Involvement of a novel transcriptional activator and small RNA in post-transcriptional regulation of the glucose phosphoenolpyruvate phosphotransferase system. *Mol Microbiol*, 54 (4): 1076-1089.

Vassinova, N. & Kozyrev, D. (2000) A method for direct cloning of fur-regulated genes: identification of seven new fur-regulated loci in Escherichia coli. *Microbiology*, 146 Pt 12 3171-3182.

Vecerek, B.; Rajkowitsch, L.; Sonnleitner, E.; Schroeder, R. & Blasi, U. (2008) The C-terminal domain of Escherichia coli Hfq is required for regulation. *Nucleic Acids Res*, 36 (1): 133-143.

Vishnoi, M.; Narula, J.; Devi, S. N.; Dao, H. A.; Igoshin, O. A. & Fujita, M. (2013) Triggering sporulation in Bacillus subtilis with artificial two-component systems reveals the importance of proper Spo0A activation dynamics. *Mol Microbiol*, 90 (1): 181-194.

Vogel, J.; Argaman, L.; Wagner, E. G. & Altuvia, S. (2004) The small RNA IstR inhibits synthesis of an SOS-induced toxic peptide. *Curr Biol*, 14 (24): 2271-2276.

Vogel, J.; Bartels, V.; Tang, T. H.; Churakov, G.; Slagter-Jager, J. G.; Huttenhofer, A. & Wagner, E. G. (2003) RNomics in Escherichia coli detects new sRNA species and indicates parallel transcriptional output in bacteria. *Nucleic Acids Res*, 31 (22): 6435-6443.

Vogel, J. & Sharma, C. M. (2005) How to find small non-coding RNAs in bacteria. *Biol Chem*, 386 (12): 1219-1238.

Voigt, B.; Antelmann, H.; Albrecht, D.; Ehrenreich, A.; Maurer, K. H.; Evers, S.; Gottschalk, G.; van Dijl, J. M.; Schweder, T. & Hecker, M. (2009) Cell physiology and protein secretion of Bacillus licheniformis compared to Bacillus subtilis. *J Mol Microbiol Biotechnol*, 16 (1-2): 53-68.

Vytvytska, O.; Moll, I.; Kaberdin, V. R.; von Gabain, A. & Blasi, U. (2000) Hfq (HF1) stimulates ompA mRNA decay by interfering with ribosome binding. *Genes Dev*, 14 (9): 1109-1118.

Wade, J. T. & Grainger, D. C. (2014) Pervasive transcription: illuminating the dark matter of bacterial transcriptomes. *Nat Rev Microbiol*, 12 (9): 647-653.

Wadler, C. S. & Vanderpool, C. K. (2007) A dual function for a bacterial small RNA: SgrS performs base pairing-dependent regulation and encodes a functional polypeptide. *Proc Natl Acad Sci U S A*, 104 (51): 20454-20459.

Wagner-Herman, J. K.; Bernard, R.; Dunne, R.; Bisson-Filho, A. W.; Kumar, K.; Nguyen, T.; Mulcahy, L.; Koullias, J.; Gueiros-Filho, F. J. & Rudner, D. Z. (2012) RefZ facilitates the switch from medial to polar division during spore formation in Bacillus subtilis. *J Bacteriol*, 194 (17): 4608-4618.

Waller, L. N.; Fox, N.; Fox, K. F.; Fox, A. & Price, R. L. (2004) Ruthenium red staining for ultrastructural visualization of a glycoprotein layer surrounding the spore of Bacillus anthracis and Bacillus subtilis. *J Microbiol Methods*, 58 (1): 23-30.

Walsh, B. W.; Bolz, S. A.; Wessel, S. R.; Schroeder, J. W.; Keck, J. L. & Simmons, L. A. (2014) RecD2 helicase limits replication fork stress in Bacillus subtilis. *J Bacteriol*, 196 (7): 1359-1368.

Wang, G.; Yi, X.; Li, Y. Q. & Setlow, P. (2011) Germination of individual Bacillus subtilis spores with alterations in the GerD and SpoVA proteins, which are important in spore germination. *J Bacteriol*, 193 (9): 2301-2311.

Wang, J.; Liu, T.; Zhao, B.; Lu, Q.; Wang, Z.; Cao, Y. & Li, W. (2016) sRNATarBase 3.0: an updated database for sRNA-target interactions in bacteria. *Nucleic Acids Res*, 44 (D1): D248-253.

Wang, S. T.; Setlow, B.; Conlon, E. M.; Lyon, J. L.; Imamura, D.; Sato, T.; Setlow, P.; Losick, R. & Eichenberger, P. (2006) The forespore line of gene expression in Bacillus subtilis. *J Mol Biol*, 358 (1): 16-37.

Wang, Y. (2002) The function of OmpA in Escherichia coli. *Biochem Biophys Res Commun*, 292 (2): 396-401.

Wang, Z.; Gerstein, M. & Snyder, M. (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet*, 10 (1): 57-63.

Wassarman, D. A. (1993) Psoralen crosslinking of small RNAs in vitro. *Mol Biol Rep*, 17 (2): 143-151.

Wassarman, K. M. (2007) 6S RNA: a small RNA regulator of transcription. *Curr Opin Microbiol*, 10 (2): 164-168.

Wassarman, K. M. & Saecker, R. M. (2006) Synthesis-mediated release of a small RNA inhibitor of RNA polymerase. *Science*, 314 (5805): 1601-1603.

Waters, L. S. & Storz, G. (2009) Regulatory RNAs in bacteria. *Cell*, 136 (4): 615-628.

Waters, S. A.; McAteer, S. P.; Kudla, G.; Pang, I.; Deshpande, N. P.; Amos, T. G.; Leong, K. W.; Wilkins, M. R.; Strugnell, R.; Gally, D. L.; Tollervey, D. & Tree, J. J. (2017) Small RNA interactome of pathogenic E. coli revealed through crosslinking of RNase E. *Embo j*, 36 (3): 374-387.

Watson, P. Y. & Fedor, M. J. (2011) The glmS riboswitch integrates signals from activating and inhibitory metabolites in vivo. *Nat Struct Mol Biol*, 18 (3): 359-363.

Weng, M.; Nagy, P. L. & Zalkin, H. (1995) Identification of the Bacillus subtilis pur operon repressor. *Proc Natl Acad Sci U S A*, 92 (16): 7455-7459.

Whalen, M. B. & Piggot, P. J. (1997) Gain-of-function mutation of sapB that affects formation of alkaline phosphatase by Bacillus subtilis in sporulation conditions. *Microbiology*, 143 (Pt 2) 577-583.

Whitten, A. E.; Jacques, D. A.; Hammouda, B.; Hanley, T.; King, G. F.; Guss, J. M.; Trewhella, J. & Langley, D. B. (2007) The structure of the KinA-Sda complex suggests an allosteric mechanism of histidine kinase inhibition. *J Mol Biol*, 368 (2): 407-420.

Wilson, K. S. & von Hippel, P. H. (1995) Transcription termination at intrinsic terminators: the role of the RNA hairpin. *Proc Natl Acad Sci U S A*, 92 (19): 8793-8797.

Wright, P. R.; Georg, J.; Mann, M.; Sorescu, D. A.; Richter, A. S.; Lott, S.; Kleinkauf, R.; Hess, W. R. & Backofen, R. (2014) CopraRNA and IntaRNA: predicting small RNA targets, networks and interaction domains. *Nucleic Acids Res*, 42 (Web Server issue): W119-123.

Wright, P. R.; Richter, A. S.; Papenfort, K.; Mann, M.; Vogel, J.; Hess, W. R.; Backofen, R. & Georg, J. (2013) Comparative genomics boosts target prediction for bacterial small RNAs. *Proc Natl Acad Sci U S A*, 110 (37): E3487-3496.

Xenopoulos, P. & Piggot, P. J. (2011) Regulation of growth of the mother cell and chromosome replication during sporulation of Bacillus subtilis. *J Bacteriol*, 193 (12): 3117-3126.

Yang, C.; Hockenberry, A. J.; Jewett, M. C. & Amaral, L. A. (2016) Depletion of Shine-Dalgarno Sequences within Bacterial Coding Regions Is Expression Dependent. *G3 (Bethesda)*,

Yao, S. & Bechhofer, D. H. (2009) Processing and stability of inducibly expressed rpsO mRNA derivatives in Bacillus subtilis. *J Bacteriol*, 191 (18): 5680-5689.

Yao, S.; Blaustein, J. B. & Bechhofer, D. H. (2007) Processing of Bacillus subtilis small cytoplasmic RNA: evidence for an additional endonuclease cleavage site. *Nucleic Acids Res*, 35 (13): 4464-4473.

Yeh, H. Y.; Chen, T. C.; Liou, K. M.; Hsu, H. T.; Chung, K. M.; Hsu, L. L. & Chang, B. Y. (2011) The core-independent promoter-specific interaction of primary sigma factor. *Nucleic Acids Res*, 39 (3): 913-925.

Yi, X. & Setlow, P. (2010) Studies of the commitment step in the germination of spores of bacillus species. *J Bacteriol*, 192 (13): 3424-3433.

Yoshida, K.; Seki, S.; Fujimura, M.; Miwa, Y. & Fujita, Y. (1995) Cloning and sequencing of a 36-kb region of the Bacillus subtilis genome between the gnt and iol operons. *DNA Res*, 2 (2): 61-69.

Young, S. B. & Setlow, P. (2003) Mechanisms of killing of Bacillus subtilis spores by hypochlorite and chlorine dioxide. *J Appl Microbiol*, 95 (1): 54-67.

Zhang, A.; Wassarman, K. M.; Ortega, J.; Steven, A. C. & Storz, G. (2002) The Sm-like Hfq protein increases OxyS RNA interaction with target mRNAs. *Mol Cell*, 9 (1): 11-22.

Zhang, A.; Wassarman, K. M.; Rosenow, C.; Tjaden, B. C.; Storz, G. & Gottesman, S. (2003) Global analysis of small RNA and mRNA targets of Hfq. *Mol Microbiol*, 50 (4): 1111-1124.

Zhang, J.; Fitz-James, P. C. & Aronson, A. I. (1993) Cloning and characterization of a cluster of genes encoding polypeptides present in the insoluble fraction of the spore coat of Bacillus subtilis. *J Bacteriol*, 175 (12): 3757-3766.

Zhang, J. Q.; Griffiths, K. K.; Cowan, A.; Setlow, P. & Yu, J. (2013) Expression level of Bacillus subtilis germinant receptors determines the average rate but not the heterogeneity of spore germination. *J Bacteriol*, 195 (8): 1735-1740.

Zhang, P.; Thomas, S.; Li, Y. Q. & Setlow, P. (2012) Effects of cortex peptidoglycan structure and cortex hydrolysis on the kinetics of Ca(2+)-dipicolinic acid release during Bacillus subtilis spore germination. *J Bacteriol*, 194 (3): 646-652.

Zheng, L.; Abhyankar, W.; Ouwerling, N.; Dekker, H. L.; van Veen, H.; van der Wel, N. N.; Roseboom, W.; de Koning, L. J.; Brul, S. & de Koster, C. G. (2016) Bacillus subtilis Spore Inner Membrane Proteome. *J Proteome Res*, 15 (2): 585-594.

Zheng, L. B.; Donovan, W. P.; Fitz-James, P. C. & Losick, R. (1988) Gene encoding a morphogenic protein required in the assembly of the outer coat of the Bacillus subtilis endospore. *Genes Dev*, 2 (8): 1047-1054.

Zheng, L. B. & Losick, R. (1990) Cascade regulation of spore coat gene expression in Bacillus subtilis. *J Mol Biol*, 212 (4): 645-660.

Zhuang, J.; Gutknecht, R.; Flukiger, K.; Hasler, L.; Erni, B. & Engel, A. (1999) Purification and electron microscopic characterization of the membrane subunit (IICB(Glc)) of the Escherichia coli glucose transporter. *Arch Biochem Biophys*, 372 (1): 89-96.

Zuker, M. & Stiegler, P. (1981) Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res*, 9 (1): 133-148.

Zweers, J. C.; Nicolas, P.; Wiegert, T.; van Dijl, J. M. & Denham, E. L. (2012) Definition of the sigma(W) regulon of Bacillus subtilis in the absence of stress. *PLoS One*, 7 (11): e48471.