Bistable Differentiation in an Isogenic Cell Population

Simon Syvertsson

Thesis submitted for the degree of
Doctor of Philosophy

Institute for Cell and Molecular Biosciences
Newcastle University

September 2013

Abstract

Single-cell organisms such as bacteria have traditionally been regarded as discrete units, which in turn has been reflected by the bulk-level methods used to study them. A growing culture of the bacterium *Bacillus subtilis* will exhibit a range of heterogeneous genetic developmental programmes such as motility, competence, and finally sporulation. As a popular choice for production of compounds in bioreactors, the bistable behaviours of *B. subtilis* may be undesirable traits, as they divert resources from their intended activity of synthesising a product.

This thesis investigates a novel observation that expression of a ribosomal subunit gene (*rpsD*) is elevated in the non-motile state of *B. subtilis*, using unstable GFP reporter constructs. The implications of using a proteolytically unstable protein as a reporter are also investigated with regard to the effect of protein degradation rates on the reporter construct, as well as presenting evidence for modulation of ClpXP activity in a $\Delta pnpA$ background.

Investigation of the motile/non-motile heterogeneous phenotype of *B. subtilis* posed a challenge for automated analysis pipelines. This thesis addresses this problem by developing and testing microscopy analysis pipelines designed to circumvent the traditional requirement for physically separated objects in a phase contrast channel, and instead using nucleoid or membrane stains to identify cells in a microscopy image.

Other factors impacting the activity of a proteolytically unstable P_{rpsD} reporter construct were investigated, including the rate of degradation of the reporter, and integration locus of the reporter construct. To assess the impact of locus positioning, a genetic tool was also created to survey changes in noise and overall expression levels from two homogeneously expressed promoters across different positions on the chromosome.

Declaration

I certify that this thesis contains my own work, except where acknowledged, and that no part of this material has previously been submitted for a qualification at this or any other university.

Acknowledgements

I would like to thank my supervisor, Dr. Leendert Hamoen for all the help and support, especially for his ability to hold the kind of vivid scientific discussion that could easily be mistaken for an argument outside the world of science. Our lab has had an incredibly high turnover rate during my stay here, but I'd like to thank past and present members of the Hamoen lab, including (in no particular order) Henrik, Pamela, Ilkay, Chris, Rita, Tom, Declan, Katka and Alex. If not for lab help, then for the occasional excuse to go to the pub or have lunch at Red Mezze. Dr. Phillip Aldridge and Prof. Colin Brooks really put me in my place during our assessment meetings, and while it wasn't always a very pleasant experience, I'm very grateful for it now.

Familjen Syvertssons tålamod med att ha en son och bror vars utlandsstudier växte från tre till sju år förtjänar också en eloge, och även om jag inte behövt mycket stöd hemifrån utöver en god jul och lite sommarfiske varje år så vet jag att ni hade funnits där om mer behövts. However, by far the most patient person close to me has been Martin, and I think if we can survive a Ph.D. together, then anything after that is just easy sailing.

Contributions to this work done by others are indicated in footnotes as they appear in the thesis, which include Prof. Remco Kort and Jojet Staal (carried out FACS experiments with motility reporter strain, chapter 4), Norbert Vischer (Author of ObjectJ, chapter 3), and Christopher Sauer (carried out ONPG assays for the β-Gal measurements in chapter 6).

This Ph.D. was funded by the Biotechnology and Biological Sciences Research Council (BBSRC).

Contents

List of Figures

List of Tables

Acronyms

AAA+ ATPases Associated with Various Cellular Activities
ATP Adenosine triphosphate
AU Arbitrary Units. Where referring to fluorescence intensity on a 16-bit scale

BCAA Branched-chain amino acids: L-isoleucine, L-leucine, or L-valine
BLAST Basic Local Alignment Search Tool
bp base pairs, can be prefixed kilo- (k) or mega- (m)

cam chloramphenicol
cat chloramphenicol acetyltransferase, confers resistance for cam
CCD charge-coupled device (digital camera)
cDNA Complimentary DNA
cfp cyan fluorescent protein
CH Casein hydrolysate
chDNA chromosomal DNA
CSV Comma-separated value file. Non-proprietary analogue to an Excel spreadsheet

Da Dalton; atomic mass unit. Usually prefixed kilo- (kDa) for describing protein mass
DAPI 4',6-diamidino-2-phenylindole
DI Deionised [water]
DNA Deoxyribonucleic Acid
dNTPs Deoxyribonucleotide mix of bases A, T, G, C

eDNA Extracellular DNA
EDTA Ethylenediaminetetraacetic acid
EPS extracellular polymeric substances
erm erythromycin

FACS Fluorescence-activated cell sorting
FP Fluorescent protein

g Gram. Can be combined with metric SI prefixes
GFP Green fluorescent protein

GTP Guanosine triphosphate

h hour
HTH Helix-turn-helix domain, a DNA binding protein domain

IMP inosine monophosphate
iNTP initiating nucleoside triphosphate, the first (+1) dNTP of a transcript
IPTG Isopropyl β-D-1-thiogalactopyranoside

kan kanamycin

l Litre. Can be combined with metric SI prefixes
LB Luria Bertani (media)

M Molar. Can be combined with metric SI prefixes
m metre. Can be combined with metric SI prefixes
MCS Multiple cloning site
min minute
mRNA Messenger RNA

N in nucleotide sequences: Any of the nucleotides A, T, G or C
NA Nutrient agar
nt Nucleotide

OD_{600} Optical density at 600 nm
ONPG *ortho*-Nitrophenyl-β-galactoside
ORF Open reading frame

PBS Phosphate buffered saline
PC Phase contrast
PCR polymerase chain reaction
PNK Polynucleotide kinase
PNPase polynucleotide phosphorylase
ppGpp guanosine tetraphosphate
pppGpp guanosine pentaphosphate

Q-PCR Quantitative PCR

RFP Red fluorescent protein
RNA Ribonucleic acid
RNAP RNA Polymerase
ROI Region of interest, the boundaries of a highlighted object in a computerised image
rRNA Ribosomal RNA

s second. Can be prefixed milli- (m)
SAS Small alarmone synthetase(s)
SDS Sodium dodecyl sulfate
SNP Single Nucleotide Polymorphism
SOE Splicing by Overlap Extension (PCR method)
spc spectinomycin

TA Toxin-antitoxin system
TCF threshold correction factor
tmRNA Transfer-Messenger RNA
TRIS tris(hydroxymethyl)aminomethane
tRNA Transfer RNA

wt Wild-type

X-gal 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside

yfp yellow fluorescent protein

Chapter 1

Introduction

A considerable advantage conferred upon multicellular life compared to single-celled organisms, is the ability to let cells differentiate. This allows them to assume useful niches that contribute to the survival of the organism as a whole. Bacteria have long been considered discrete unicellular units. As a consequence, many established methods used to study them often focus on their behaviour as a collective. As methods to investigate cellular behaviours on the single cell have become available, microbiologists have begun to appreciate a considerable population heterogeneity in cell fates and gene expression.

Electronic circuits are often used as analogies to describe the interplay of stimulus, regulators and effectors in a genetic network. While it is often an apt one, biological systems fail to live up to one aspect of the comparison in one important way. Computers are designed to reliably reproduce logical tasks, and are thus not subject to noise or variation in the outcome of a program executed repeatedly under the same conditions. A genetic circuit is subject to variations in biochemical conditions that differ from cell to cell, and may not be possible to control (noise). Such variation is not always to the detriment of the cell, but rather oppositely enables cellular differentiation and encourages the evolution of robust networks.

A good reason for bacteria to be generating multiple cell types in parallel is to take advantage of bet-hedging. Originally a term from the field of investments, bet-hedging can be described as combining an original investment with an additional investment in something that could counteract losses if an undesirable event takes place. An oversimplified example of bet-hedging taken from its original context could be an investor in solar power—he might consider a stake in an umbrella-making company to be a hedge against bad weather and the subsequent drop in demand for solar panels. Similarly, the cell fates of species such as *Bacillus subtilis* are bet-hedging strategies because bacteria cannot tell the future. That is, a portion of a culture might sporulate in anticipation of nutrient depletion while the remaining cells try their luck in the hope of finding new nutrient sources (Veening *et al.*, 2008b). In the case of either event, there

will be individuals in a good position for surviving and giving rise to future generations.

1.1 Bistability and gene regulation

Gene regulation in bacteria is predominantly done at the level of transcription. Genes are organised into transcriptional units known as operons, which are transcribed from promoters that recruit RNA polymerase (RNAP). Basal recruitment is provided by −35 and −10 elements on the DNA (the numbers here signify the position of the elements in relation to the first transcribed base), which generally increase in strength the closer to the consensus they are. An upstream promoter element (UP) sequence may further enhance RNAP recruitment by binding to the C-terminal domain of its α subunit (α-CTD). Transcription factors modulate basal gene expression by binding to operators, whose position in relation to the RNAP binding site determines if the transcription factor is an activator or repressor (figure 1.1). Though many variations on the setup exists, a simplified explanation of how transcription factors control a promoter is that repressors physically obstruct RNAP binding, and activators enhance recruitment of RNAP to the promoter (Browning and Busby, 2004). Activation may occur through a number of mechanisms, but the most basic mechanism relies on the transcriptional activator binding to its cognate motif, and recruiting RNAP by association to α-CTD. As the activity of a regulator may change based on the location of its binding motif in relation to the −35 and −10 elements, there are transcription factors that are able to exert both direct negative and positive regulation on target genes. Spo0A is an example of such a transcription factor in *B. subtilis* (Fujita *et al.*, 2005). The modulation of transcription factor activity varies, but possible mechanisms include allosteric inhibition by a metabolite (e.g., CodY by GTP), binding by an antagonist (e.g., SinR by SinI), or degradation (e.g., ComK by ClpXP, via MecA) (Ratnayake-Lecamwasam *et al.*, 2001,

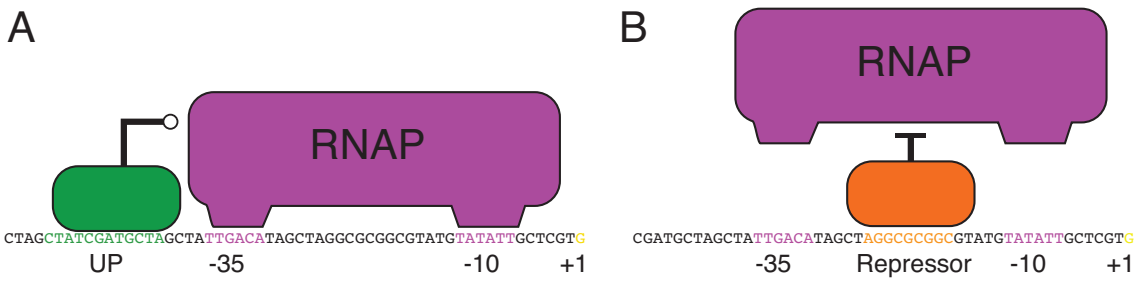

Figure 1.1: Simplified model of promoter interaction with RNA polymerase. Control of transcription initiation can be achieved through activation and repression. The RNA polymerase (RNAP) holoenzyme binds to −10 and −35 elements (pink base pairs). (A) Its interaction can be enhanced by binding of an activator (green) to an UP element, or (B) repressed by a repressor (orange) binding to a repressor site, which in some cases sterically block RNAP binding.

Bai *et al.*, 1993, Turgay *et al.*, 1998).

A second layer of regulation is available through the variety of σ (sigma) factors, a means of regulation that was first discovered in *B. subtilis* and later found to exist in many species of bacteria (Haldenwang, 1995). The first alternative σ factor to be discovered and characterised was σ^{A}, a sporulation σ factor (Moran *et al.*, 1982). RNA polymerase consists of three main subunits (α, β, and β') and a σ factor that interfaces with the -35 and -10 elements of the promoter. The σ factor is exchangeable, and several variants exist and have been ascribed to various situations. Examples range from general activity of the housekeeping σ^{A} factor expressed during normal growth, to σ^{D} for motility development, or more niche situations such as the coordination of sporulation development by σ factors σ^{E} to σ^{K} (Sonenshein *et al.*, 1993, p. 653-4). σ factors can be organised hierarchically, as is perhaps best demonstrated in sporulation, where σ factors exist for the different stages of sporulation. A particularly illustrative example is the stage of sporulation where the mother cell and pre-spore compartments arise, each with their separate requirements for gene expression. These requirements are catered for by separate σ factors—σ^{E}, then σ^{K} in the mother cell, and σ^{F}, then σ^{G} in the pre-spore (Hilbert and Piggot, 2004).

Due to the way the σ factor interacts with the RNAP holoenzyme, the response to a new σ factor being activated can be rapid. Each time transcription proceeds past 10 nucleotides, the σ factor detaches from RNA polymerase and a new σ factor will have to be recruited for the next transcription initiation (Österberg *et al.*, 2011). As a consequence, σ factors not only activate promoters under their control, but they also negatively affect genes under the control of competing σ factors, due to the need to compete against each other for inclusion as part of the RNA polymerase holoenzyme (Malik *et al.*, 1987, Österberg *et al.*, 2011).

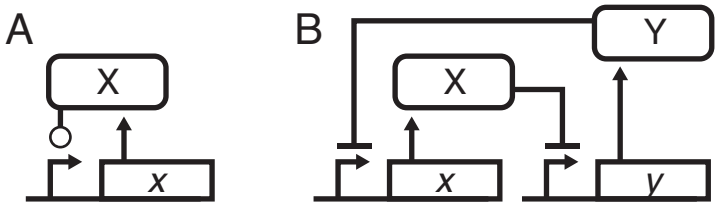

Figure 1.2: Feedback loops are required for generating bistability. A mechanism for maintaining a state once attained is essential for the generation of bistable populations. Two simplified cases using the hypothetical regulator X: (A) In a positive feedback loop, X is able to drive its own expression, and in real-life examples is often kept at low levels by various means. (B) In a double-negative feedback loop, expression of x is repressed by Y, which in turn is repressed by X. Reaching a threshold concentration of X will lead to alleviation of repression by Y, inducing a permanently high state of X. Arrows and shapes are explained in the caption of figure 1.4.

1.1.1 Bistability in the λ phage life cycle

Perhaps the earliest observed bistable phenotype in bacteria is that of the lytic and lysogenic life cycles of the λ phage, recognised in the 1950s to enter two distinct and mutually exclusive states (Lederberg, 1950). Upon infection of *Escherichia coli*, the phage may replicate and lyse the cell to release new virions in accordance with a simplified view of how viruses operate. An alternative fate is available to a subset of cells where the noise-driven accumulation of the repressor protein cI has reached threshold levels, which causes inhibition of lysis and promotes its own production, furthering maintenance of the lysogenic state. Such a mechanism for maintaining an attained genetic state is a requirement for bistable systems.

The regulatory mechanisms required for locking in bistability fall into one of two categories. The first is a positive feedback loop with a regulator driving its own transcription (figure 1.2A), as with the λ phage. The alternative is a double-negative feedback loop (figure 1.2B), as is the case in the system governing *B. subtilis* biofilm regulation, governed by SlrR and SinR down-regulating each other (Losick and Desplan, 2008). In the latter case, the production of a regulator is controlled by a repressor, whose production is in turn repressed by the regulator, and once a threshold level is reached, the regulator can prevent repression and enter a stable state (figure 1.2B). Either of these systems rely on the actors existing in finely tuned [low] concentrations that will be sensitive enough to react to input from noise-prone stochastic processes. This means that if one state is disproportionately favoured, input may not be integrated, and the switch does not yield bistability. Examples of both feedback mechanisms exist in *B. subtilis*, and are described in more detail in later sections of this chapter.

1.1.2 Noise in gene expression as a contributor to heterogeneity

In addition to the feedback loop needed to lock into a state, a second requirement for bistability is noise. More specifically, it is required in the early stages of generating a bistable state, where stochasticity can be amplified to have a greater effect. As a general rule, a system where actors (a copy of a regulatory protein, for example) are plentiful exhibits low noise. For example, the DNA replisome of bacteria proceeds at the order of several hundred of nucleotides incorporated per second, and under normal conditions the required dNTP is readily at hand for the polymerase, resulting in little variation in the execution of this process (Mirkin and Mirkin, 2005). As molecules become scarcer, noise increases—a phenomenon called the *finite-number effect* (Kærn *et al.*, 2005). Examples of systems with high noise are those with a concentration of a regulator that is stretched too thin to satisfy demand at every place on the chromosome where it could bind (figure 1.3D). To put it simple, when a regulator is too scarce and is in demand in many places

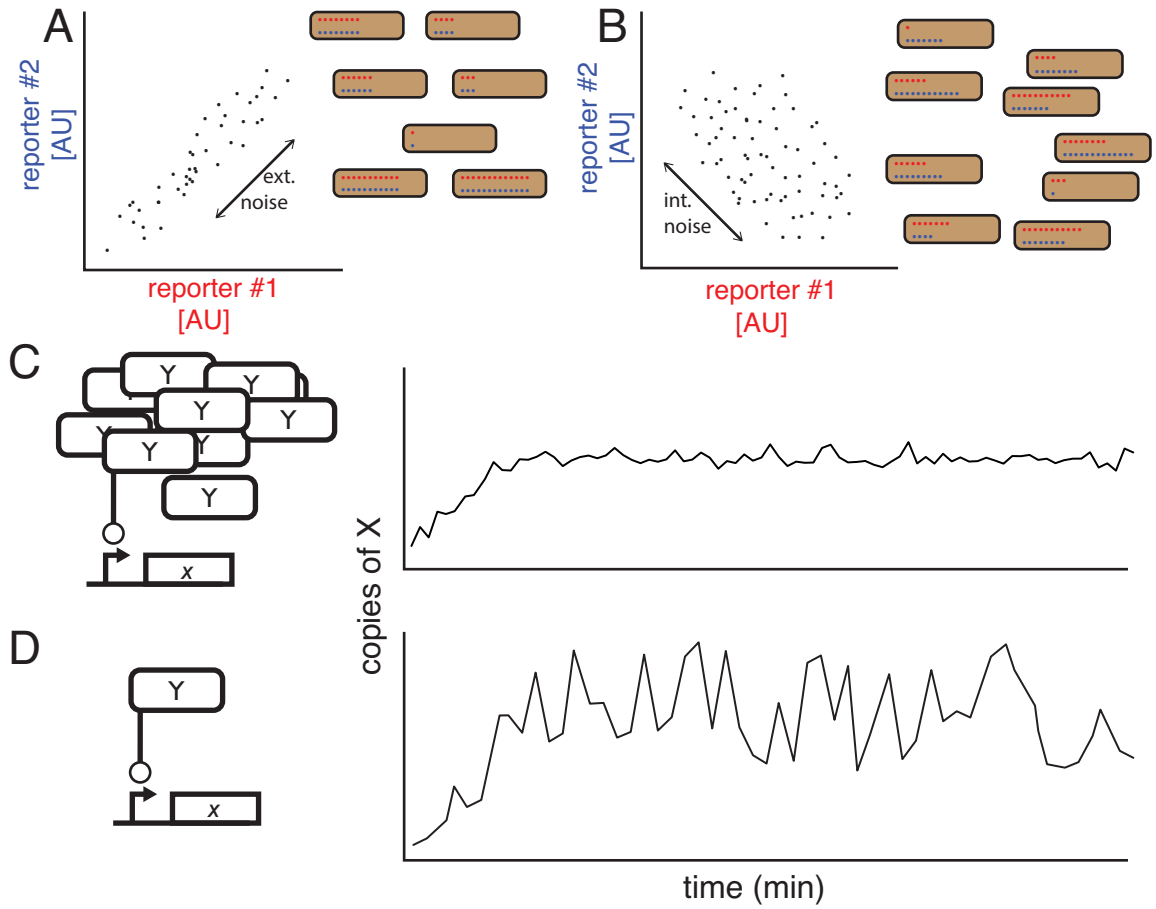

Figure 1.3: Noise in gene expression and its extrinsic and extrinsic components. (A–B) Left side of panels show the relative signal of multiple cells from two hypothetical reporter constructs, and the right panel show the level at which they are expressed in each cell (brown boxes). (A) A system with high extrinsic noise varies much from cell to cell, but the two reporter molecules are present at equal concentrations in each cell. (B) In a system with high intrinsic noise, the two reporters vary independently of each other. (C–D) left side of panels show the gene x being positively regulated by the protein Y, right side shows the levels of the protein X over time. (C) In a system where the regulator Y is plentiful, activation of P_x is frequent and there is little noise. (D) In a system where Y is scarce, the production of X is subject to noise due to the finite-number effect. Figures adapted from Elowitz *et al.* (2002) (A–B) and Kærn *et al.* (2005) (C–D).

simultaneously, the expected outcome is random, leading to a split in outcomes if this situation were to be played out in many cells at once. Compare the same situation, but with an increased concentration of the regulator (figure 1.3C); there are now sufficient copies of the regulator for it to be at hand everywhere and every time it is needed, and activation is a certainty in every parallel case [cell] where this situation is evaluated.

In simulating regulatory events such as these, stochastic methods are preferable over deterministic in order to evaluate the impact of randomness. For example, a typical simulation of protein levels in a cell may consist of a chance of the relevant gene being transcribed, the resulting mRNA being translated, as well as the chance of it being degraded, and finally, the chance of the protein being degraded (Wilkinson, 2009). Each parameter is evaluated at each time unit for each copy of the species present. This essentially follows the same rules as some board games—"If the player rolls a 6-sided dice for 2 or lower, this piece is removed from play". In the same way, the parameters set out at the beginning of a simulation could dictate that the chance of each copy of protein X to be degraded to be $\frac{1}{3}$. In fact, the elements of randomness shared with games of chance were not lost on the creator of the method, who named it the *Monte Carlo method* after the city with the famous eponymous casino (Metropolis, 1987).

Noise in cellular processes is divided into two components—intrinsic and extrinsic noise. Intrinsic noise (η_{int}) concerns comparable noise-prone events within a cell, such as the competition for a scarce transcription factor by commonly occurring binding sites on a chromosome. This category of noise is most often visualised experimentally by allowing identical promoters to drive distinguishable fluorescent proteins (FPs). An inducible P_{spac}-*cfp* / P_{spac}-*yfp* system used by Elowitz *et al.* (2002) successfully demonstrated the usefulness of this approach. They showed how the constructs under highly induced conditions (with low intrinsic noise) exhibited correlated CFP and YFP from cell to cell (figure 1.3A), and vice-versa showing disjoint fluorescence intensities in systems with high intrinsic noise (figure 1.3B).

Extrinsic noise (η_{ext}) affects the whole cell uniformly, meaning variance is present on an intercellular level. Here, examples include whether an individual cell has undertaken a global regulatory event, or simply entered the same growth state as its peers when grown in a batch culture. Reduction of extrinsic noise is an important part in population-level experimental methods. For example, cultures can be repeatedly diluted (synchronised) during exponential phase to encourage a larger proportion of the culture to be in the same same physiological state. An example of an induced synchronisation method available in *B. subtilis* is a temperature-sensitive mutant of the replication initiation protein DnaB, which allows a culture to start DNA replication at the same time after shifting to a permissive temperature (Wang *et al.*, 2007b). Extrinsic noise can also arise as a knock-on effect from changes in metabolic pathway utilisation,

which has been observed in *E. coli*, where a faster-growing subpopulation used the Entner-Doudoroff pathway instead of glycolysis (Labhsetwar *et al.*, 2013).

Together, the two components contribute to the total noise (η_{tot}), and as illustrated in figure 1.3, the variance of observations in either diagonal direction corresponds to one type of noise. Since they are perpendicular, the Pythagorean theorem can be used to calculate total noise, as shown below (Elowitz *et al.*, 2002):

$$\eta_{tot}^2 = \eta_{int}^2 + \eta_{ext}^2$$

1.2 Heterogeneity and bistability in *B. subtilis*

The Gram-positive soil bacterium *B. subtilis* is considered a paradigm for bistable and heterogeneous differentiation in bacteria, with numerous distinctive cell fates identified to date (López *et al.*, 2009b). The regulation of bistable and heterogeneous cell fates is achieved through a complex system of interlocking regulatory relationships which often ensure the deactivation of competing cell fates through switches and lockout mechanisms. Many of the described regulators in this section are depicted in figure 1.4.

1.2.1 Regulatory decisions in sporulation

Sporulation is a property of several species of mostly Gram-positive bacteria, where the cells respond to nutrient stress by developing into a dormant spore capable of withstanding harsh environmental punishment, including extreme heat, UV irradiation and aridity. Well-preserved spores from other *Bacillus* species have been found to remain viable for upwards of 20 million years and still be viable, which probably means that the lifetime of a spore is near-indefinite under favourable conditions (Cano and Borucki, 1995).

Sporulation in *B. subtilis* is controlled by the master regulator Spo0A, which is activated by phosphorylation. As described earlier, bistable systems require at least one of two autoregulatory mechanisms (figure 1.2), and Spo0A possesses both the basic types of loops described. Simple positive feedback is achieved by Spo0A controlling its own expression, and a double-negative repression loop consisting of the *spo0A* activator σ^{H} being kept at low levels by AbrB, whose expression in turn is repressed by Spo0A (Hahn *et al.*, 1995). Spo0A directly controls 121 genes (approximately one third repressed, and the remainder activated), but as some target genes encode transcription factors, its total sphere of influence extends to some additional ∼400 genes by indirect means (Piggot and Hilbert, 2004).

The core components responsible for the [phosphorylation] activation of Spo0A are organised in what is referred to as a phosphorelay, consisting of Spo0F, Spo0B and

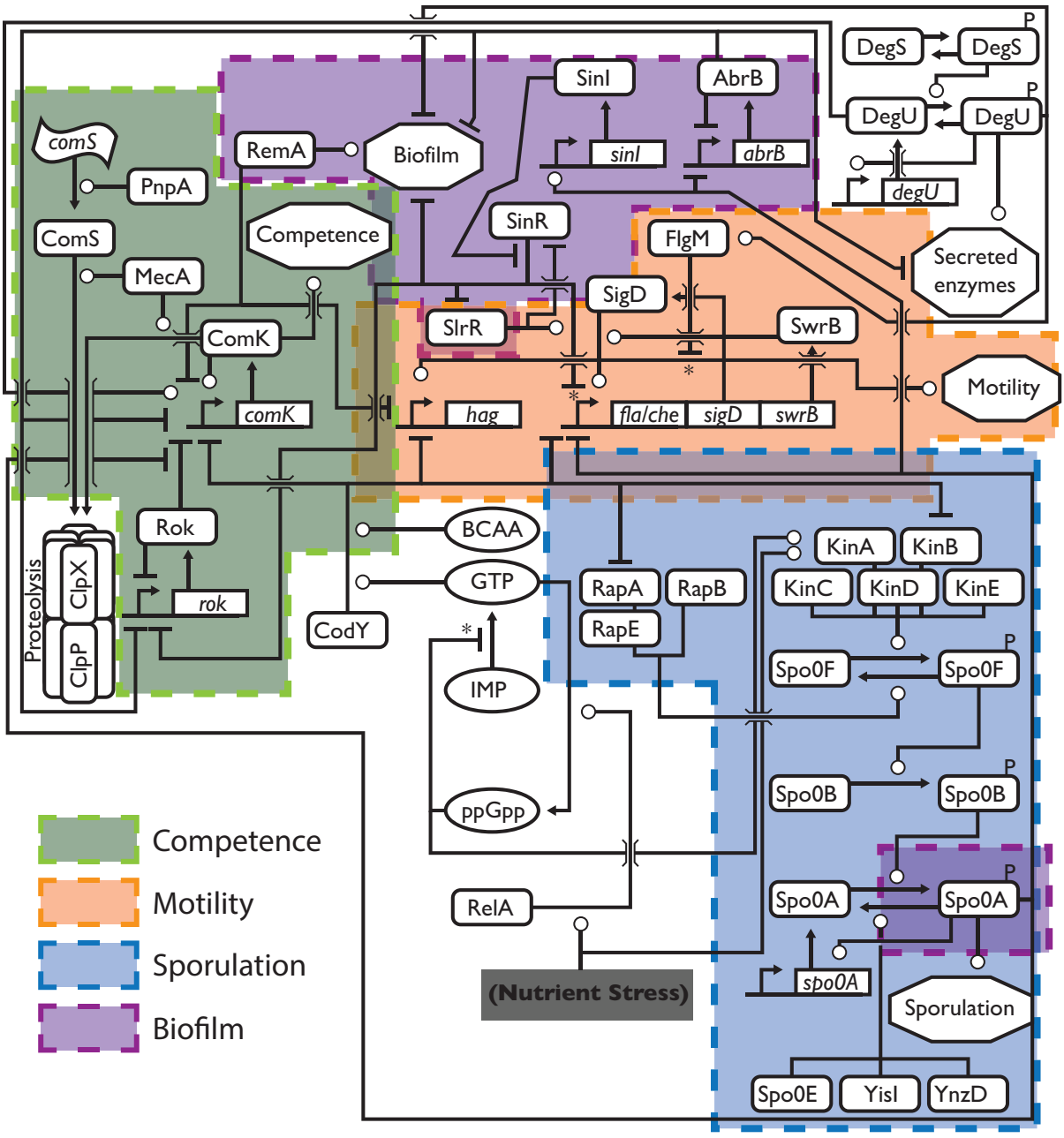

Figure 1.4: Overview of relationships between regulators considered in this thesis, put in the context of other known regulators of bistability in *B. subtilis*. Proteins are symbolised as rounded rectangles, genes are represented as small bent arrows with a rectangle (symbolising a promoter and ORF, respectively), metabolites in ovals, mRNA in a wavy rectangle, and cell fates as octagons. Arrowed lines connecting species represent a transition between states (e.g., a metabolite being processed, a gene being transcribed and translated into a protein.), T-bar lines represent repression, and circle-ended lines represent activation. Some indirect relationships are indicated with an asterisk where intermediate actors have been omitted. Much overlap exists between pathways, and coloured areas serve to give a *general indication* of where regulators are involved. A version of this figure showing references can be found in the appendix (figure B.1)

Spo0A. It begins with the upstream effector Spo0F, which is subject to phosphorylation by five histidine kinases (KinA–E) who respond to different cues. Contributions from KinA and KinB are the most well understood, and they appear to sense the impairment of cell respiration due to low nutrient levels, which increases the level of KinA to reinforce Spo0A phosphorylation (Eswaramoorthy *et al.*, 2010, Kolodkin-Gal *et al.*, 2013). The remaining kinases (KinCDE) are not central to sporulation under laboratory conditions, but may be more relevant to regulating Spo0A in the context of other stages of growth (Jiang *et al.*, 2000, McLoon *et al.*, 2011). Spo0F transfers its phosphate to Spo0B, which in turn transfers it to Spo0A, leading to sporulation given that a threshold level of Spo0A∼P can be reached and maintained (Burbulys *et al.*, 1991). Negative control by phosphoregulation of this pathway occurs at both the level of Spo0F and Spo0A. The CodY-repressed group of phosphatases RapAE act on Spo0F∼P (Molle *et al.*, 2003b), while Spo0E, YisI and YnzD are phosphatases of Spo0A∼P (Ohlsen *et al.*, 1994, Perego, 2001). It should perhaps also be noted that an effect of involving CodY here is that the stringent response becomes a factor in sporulation, due to the drop in GTP affecting the regulatory network via CodY (Geiger and Wolz, 2014). Furthermore, the stringent response is also involved in sporulation decision-making on the level of Spo0F activation, where it promotes production of KinAB (Tojo *et al.*, 2013).

Levels of Spo0A activity in a cell do not simply fall into states of ON or OFF; a more graded response has been shown to exist with genes in the Spo0A operon categorised as responding to low or high levels of Spo0A∼P. This phenomenon is caused by variation in the promoter regions, with $\text{Spo0A}^{\text{LOW}}$ promoters having a higher affinity for Spo0A∼P, thus being able to achieve activation at lower Spo0A∼P concentrations. Conversely, $\text{Spo0A}^{\text{HIGH}}$ promoters are sensitive to only higher levels of Spo0A∼P because they bind with a lower affinity. Logically, it is the $\text{Spo0A}^{\text{HIGH}}$ genes which are responsible for commitment to sporulation, while the $\text{Spo0A}^{\text{LOW}}$ regulon includes genes such as *skf* (cannibalism of sibling cells to recover nutrients) and *sdp* (sporulation delaying protein) (Fujita *et al.*, 2005). Alongside the positively regulated sporulation genes, entry into "competing" developmental pathways are inhibited by repression of *med* (positive *comK* regulator) at low $\text{Spo0A}^{\text{LOW}}$ levels, and at $\text{Spo0A}^{\text{HIGH}}$ levels by direct repression of the *fla/che* operon (motility) (Fujita *et al.*, 2005). As previously explained, $\text{Spo0A}^{\text{LOW}}$ regulatory elements have a high affinity for Spo0A∼P, so it would stand to reason that such promoters would also be active at $\text{Spo0A}^{\text{HIGH}}$ concentrations. However, in several cases (including *sdp*), an additional $\text{Spo0A}^{\text{HIGH}}$ regulatory element represses expression once a sufficient concentration has been reached, thus effectively limiting expression of the genes to a Spo0A∼P concentration with lower *and* upper bounds (Fujita *et al.*, 2005).

Competence is repressed by Spo0A indirectly, through pathways first identified as

belonging to the activation biofilm formation (also regulated by Spo0A). The repression of ComK is promoted by Spo0A by unraveling a chain of repressors ending with the ComK repressor Rok, which is inhibited by SinR, which in turn is repressed by SinI, which finally is activated by Spo0A (Fujita *et al.*, 2005, Bai *et al.*, 1993, Schultz *et al.*, 2009).

1.2.2 Competence development

A partial reason for the success of *B. subtilis* as a model organism is its natural competence for DNA uptake and integration. Natural competence develops as a response to nutrient limitation, and involves dedicated mechanisms for importing extracellular DNA through the cell wall, which can then be integrated into the chromosome. While researchers use this feature to integrate foreign DNA, it benefits the organism in a natural setting by facilitating uptake of more subtle changes, such as useful alleles that had arisen through mutations in sibling cells, which could then propagate through successive generations if a suitable selection pressure exists. Natural competence is prevalent among both Gram-positive and Gram-negative bacteria, and well-studied additional examples include *Streptococcus pneumoniae*, *Haemophilus influenzae*, and *Neisseria gonorrhoeae* (Claverys and Martin, 2003). Due to the differences in cell envelope structure in Gram-negative bacteria, some additional components exist that localise to the outer membrane. However, as a whole, the uptake and integration components are well-preserved across most bacterial species capable of natural competence, while the process of inducing competence often differs (Johnston *et al.*, 2014). There are examples of species of γ-Proteobacteria that make use of half the system, i.e., they import DNA for use as nutrients but do not integrate it (Finkel and Kolter, 2001). This behaviour is however not universal among γ-Proteobacteria, as *Vibrio cholerae* has recently been shown to have natural competence (Lo Scrudato and Blokesch, 2012).

Development of competence is controlled by the master competence regulator ComK, which also affects its own production through positive regulation of P_{comK} (van Sinderen and Venema, 1994, van Sinderen *et al.*, 1995). Autoregulation of *comK* makes up the main positive feedback loop of the bistable competence system, and noise is generated by constant repressive control of ComK levels by several systems (Hoa *et al.*, 2002, Serror and Sonenshein, 1996, Hamoen *et al.*, 2003a, Mirouze *et al.*, 2012). One such mechanism is provided by MecA, which is an adaptor of the ClpCP protease complex which targets several proteins, including ComK to its protease, where both the adaptor and target are degraded (Gur *et al.*, 2013). Alleviation of this repression is achieved by out-competing ComK for binding to MecA, which is done by the small protein ComS (Persuh *et al.*, 1999). The *comS* gene is situated within an ORF in the

much longer *srfAA* transcript responsible for surfactin production (D'Souza *et al.*, 1994). In order for it to be translated, the *srfAA* transcript requires additional post-transcriptional processing by polynucleotide phosphorylase (PNPase) (Luttinger *et al.*, 1996). The precise mechanism responsible for maturation of the *comS* transcript remains unknown, but competence can be induced by inducing *comS* production from an ectopic source (Liu *et al.*, 1996).

Further fine-tuning of *comK* expression is achieved by contributions from a large number of additional regulators (Hamoen *et al.*, 2003b). Rok (repressor of *comK*) represses expression of *comK* and its own gene, *rok* (Hoa *et al.*, 2002). *rok* is under repressive control from both AbrB and SinR, which are both under the control of Spo0A albeit in different contexts—AbrB is repressed at $\text{Spo0A}^{\text{LOW}}$ concentrations, SinR is activated at $\text{Spo0A}^{\text{HIGH}}$ (Fujita *et al.*, 2005, Schultz *et al.*, 2009, Hahn *et al.*, 1995). More direct involvement of Spo0A on *comK* activation is provided by antagonising binding of Rok to its regulatory element (Mirouze *et al.*, 2012). Additional repression of *comK* is also contributed by AbrB (transition state regulator), CodY (nutritional repressor, see section 1.3.3) (Hahn *et al.*, 1995). Similarly, some further input by regulators from competing differentiation pathways can also contribute by activation of *comK*; The biofilm formation repressor SinR provides indirect activation (Hoa *et al.*, 2002), as well as non-phosphorylated DegU (Hamoen *et al.*, 2000), which when phosphorylated is a regulator of degradative enzyme production. Competence development is mainly subject to being locked out by other programmes, except in the case of motility, where ComK promotes the production of FlgM, the anti-σ^{D} factor (Caramori *et al.*, 1996), providing a switch between the two states that is further reinforced by MecA also increasing σ^{D} levels (Liu and Zuber, 1998). Additional negative effects on *comK* regulation may arise from the frequency at which the ComK recognition motif (K-box) occurs across the chromosome. Such sites vastly outnumber the number of experimentally verified ComK regulon members, which may provide a sink that diverts ComK from its role in competence development (Hamoen *et al.*, 2002).

Following successful activation of *comK*, the state is maintained by constant production of ComK. The structural components of the uptake machinery are synthesised from *comC*, *comP*, and the *comGA* operon, which is free of direct involvement by the range of regulators affecting *comK*. The uptake and integration machinery imports single-stranded DNA as long as 20 kb, which is then integrated to the existing chromosome, provided an existing homology to a region on the chromosome. Regions of homology as short as a few hundred kbp have been shown to be successfully integrated, and if the DNA is of chromosomal origin (as opposed to a plasmid vector), close to 100% of macromolecules are integrated once imported (Dubnau, 1991).

1.2.3 Biofilm development

Biofilms are bacterial communities characterised by their complex architectural organisation, and the success of biofilm formation as a survival mechanism is particularly problematic from clinical and industrial perspectives. Biofilms account for over 65% of all hospital-acquired infections and cost billions of dollars annually, partly due to their ability to colonise medical implants and in-dwelling catheters, and resist removal (Smith and Hunter, 2008). Similarly, the cost of biofilms to industry is also significant, where they cause corrosion of equipment exposed to hydrated environments, as well as partial or complete blockage of flow in pipes.

Traditionally used compounds for specific targeting of bacteria (e.g., antibiotics), or more broadly acting biocides such as chlorine are ineffective against well-established biofilms (Smith and Hunter, 2008). This resilience is not due to any resistance mechanisms specific to the compounds used, but simply because of the inability of the agent to penetrate the biofilm through diffusion. Furthermore, as a consequence of insufficient diffusion through a thick biofilm, cells located at the centre of the structure are metabolically inactive due to low nutrient levels in their immediate vicinity. This makes them less vulnerable to antibiotics, which often exert their activity by targeting an active process at a vulnerable step.

B. subtilis biofilms are hydrophobic, a property which can be readily demonstrated by placing a droplet of water on top of a biofilm colony which will resist being soaked. This property is in part the result of the hydrophobic protein BslA (Kobayashi and Iwano, 2012, Hobley *et al.*, 2013), but also due to the topology of biofilms, which is conferred by extracellular polymeric substances (EPSs). The role of surface topology was demonstrated by Epstein *et al.* (2011), who demonstrated that a plastic resin mould of a biofilm surface could suspend water droplets with intermediate efficacy (compared to a flat plastic surface of the same material, and a wild-type biofilm surface). When the biofilm surface topology and hydrophobic proteins work in synergy, even droplets of solvents such as of 50% ethanol can be suspended on a wild-type biofilm or colony surface (Epstein *et al.*, 2011).

More active lifestyle choices localised within the biofilm structure were identified by Vlamakis *et al.* (2008) by following the localisation of fluorescent reporter strains in cross-section slices of *B. subtilis* colonies. This study further confirmed the long-held observation that the attachment surface-distal face of the biofilm is inhabited by cells on a developmental pathway headed for sporulation. This strategic placement of spores in a microenvironment is likely to aid their dispersal and has been compared to the fruiting body (sporocarp) of basidiomycetes and ascomycetes (higher fungi) (Branda *et al.*, 2001). Other localised cell niches identified by the study were σ^{D} activity (as identified by fluorescence signal produced from a P_{hag} reporter construct), which appeared to be

located predominantly towards the bottom of the structure (Vlamakis *et al.*, 2008). This inverse organisation, where cells only take on one role demonstrates the lock-out mechanisms present in many of the cell fate programs, seen in figure 1.4.

A central component of biofilm development is the production of an extracellular matrix, which acts as mortar if cells were compared to bricks in a wall, but also providing an overall coat for the biofilm. The extracellular matrix consists of a wide array of molecules, most notably (but not limited to) exopolysaccharides, proteins, and extracellular DNA (eDNA). The most widely studied strain used to study biofilm formation in *B. subtilis* is the wild-type isolate NCIB3610 (168CA lost its biofilm-producing capabilities during domestication) (Marvasi *et al.*, 2010, Cairns *et al.*, 2014). In NCIB3610, exopolysaccharide is in part produced by a pathway encoded by the 15-gene *epsA–O* operon, whose product is not yet entirely characterised. Chai *et al.* (2012) identified a role for the *epsA–O* operon in using up toxic UDP-galactose, and reported exopolysaccharide to be composed of glucose, *N*-acetyl-galactose, and galactose (in order of prevalence). However, a conflicting study by another group reports mannose, then glucose to be the most common components of the exopolysaccharide (Jones *et al.*, 2014). Consistent with the perceived purpose of the EPS—to just provide a tangled and tough structure without specific chemical properties—it may be that the exact composition of *epsA–O* exopolysaccharide is unimportant, and therefore governed by the availability of a wide range of substrates (Cairns *et al.*, 2014). In addition to the *epsA–O* of NCIB3610, many *B. subtilis* strains have their own EPSs (Marvasi *et al.*, 2010). Taken together, the EPSs are very diverse in size and properties, ranging from ∼0.5–128 kDa in mass, and similarly diverse in the charge and (non-) polarity of the polymers (Marvasi *et al.*, 2010). The protein component of the extracellular matrix is largely represented by the *tapA-sipW-tasA* operon, where SipW is a signal peptidase responsible for cleavage of TapA and TasA (Driks, 2011, Terra *et al.*, 2012). Attached to the outer surface of the peptidoglycan, TapA provides a nucleation and anchoring point for the polymerisation of extracellular TasA (Romero *et al.*, 2011).

The eDNA component of biofilms contributes to surface adhesion, which is an important step in establishing the biofilm (Das *et al.*, 2010). Molecular characterisation of the eDNA itself has across many species shown that the sequence is genomic—that is, a specialised sequence is not replicated and exported only for this purpose (Zafra *et al.*, 2012). The whole-genomic content of eDNA suggests that it is the product of cell lysis, and while this is certainly the case for some species, *B. subtilis* can export eDNA, as is evidenced by its presence in the supernatant of exponentially growing culture (Zafra *et al.*, 2012). eDNA from *B. subtilis* can also be used for transformation, which suggests that it plays a dual role, serving both competence and biofilm formation (Zafra *et al.*, 2012). It should perhaps be noted that the existence of secreted "cannibalisation" toxins

in conjunction with sporulation hints that eDNA from lysed cells also exists, but whether it is significant and expressed at the appropriate time is questionable. López *et al.* (2009c) showed that the presence of surfactin (here doubling as a signalling molecule in addition to contributing colony-spreading surfactant activity) promotes the production of both matrix genes and *skf* (spore killing factor).

A model system for studying biofilm formation has been the use of pellicles—a floating biofilm that forms on top of undisturbed liquid media. The architectural complexity of pellicles and colonies growing on solid media constitute phenotypes which are affected by genes responsible for biofilm formation, as evidenced by a reduction in their complexity upon deleting key genes such as the *eps* operon (Kobayashi and Iwano, 2012). A great deal of effort has gone into developing methods of clearing biofilms, and targeting the natural mechanisms for biofilm dispersal has been met with some success. The constituents of spent media following pellicle formation were analysed by Kolodkin-Gal *et al.* (2010) with the rationale that in order to avoid buildup of toxic compounds, a molecular trigger exists for dispersing the biofilm. A mix of D-amino acids with the ability to disrupt biofilms was identified in the spent medium. The incorporation of D-leucine, D-methionine, D-tyrosine, and D-tryptophan into the cell wall, causing release of TasA was first thought to be the underlying mechanism. However, the same group later showed that reversing a naturally occurring mutation in the gene encoding D-Tyr-tRNA de-acylase restored resistance to the D-amino acid mix by preventing incorporation of the amino acid into proteins, but not into the cell wall (Leiman *et al.*, 2013). Another component of biofilms that has been targeted is the eDNA, which through the action of a secreted DNAse of *Bacillus licheniformis* leads to the dispersal of both Gram-positive and Gram-negative biofilms (Nijland *et al.*, 2010). As an example of the complexity underlying biofilm disassembly, recent developments in the field saw one prominent group suggesting norspermidine as a trigger for disassembly, while another later provided evidence for the complete opposite (Kolodkin-Gal *et al.*, 2012, Hobley *et al.*, 2014).

The development of biofilms in *B. subtilis* requires alleviation of the biofilm repressor SinR, which is controlled by SinI. The production of SinI is mainly governed by Spo0A, which drives transcription of *sinI* when at $\text{Spo0A}^{\text{HIGH}}$ concentrations (Fujita *et al.*, 2005). The de-repression of SinR targets is achieved by SinI binding to SinR with a 1:1 stoichiometry (López *et al.*, 2009b). Once free of SinR repression, activation of the *epsA–O* and *tapA-sipW-tasA* operons is provided by RemA, which also provides negative regulation of motility by inhibition of *hag* transcription (Winkelman *et al.*, 2013). Additionally, the latter of the two operons also requires the activator *slrR*, whose transcription is also controlled by SinR (Chu *et al.*, 2008). SlrR also serves a purpose in maintaining cells as chains (as opposed to motile and unattached cells), as it binds to

SinR and inhibits the *lyABC* operon responsible for remodelling the cell wall upon cell separation (Chai *et al.*, 2010). The generation of heterogeneity in cell fates with regard to matrix production is largely down to the action of transcription factors mainly attributed to sporulation, with *sinI* being a member of the $Spo0A^{LOW}$ regulon (sensitive to intermediate levels of active Spo0A). A further layer of regulation is added by AbrB, a repressor acting (independently) on the same targets as SinI, and is repressed by active Spo0A (Chu *et al.*, 2008).

1.2.4 Antimicrobial-tolerant persister cells

Most examples of bacterial bistability exhibit relatively large subpopulations entering an alternative developmental pathway. In sporulation, competence or motility development, tenths of the whole population are often considered (Haijema *et al.*, 2001, Kearns and Losick, 2005). A far more subtle cell fate of great importance in a clinical and industrial context is the development of a persister state, which has been most widely observed in *E. coli* and *Pseudomonas aeruginosa*. Following harsh antibiotic treatment of a culture, about one in 10,000 cells can be found to enter a metabolically dormant (not sporulating) state which allows them remain viable. It was established early that the response is of a temporary and non-heritable nature, as persisters are culturable. That is, if subjected to the same treatment again, the same proportion of cells can be killed (Bigger, 1944). While a laboratory setting thus far yields a rather low proportion of persister cells, a recent study with *Salmonella* Typhimurium shows that the number of persister can increase by orders of magnitude if internalised by macrophages (Helaine *et al.*, 2014).

The exact nature of the regulatory network behind the small size of a persister population is unknown, but largely assumed to be of a stochastic nature. Several systems related to cell stress have been tied to the persister state, including the stringent response (see section 1.3.1), SOS response (cell cycle arrest in response to DNA damage), and toxin-antitoxin (TA) systems (Lewis, 2010). TA systems are genes usually found in pairs and as the name suggests, they consist of one toxin and an accompanying antitoxin which neutralises it. The toxin targets a sensitive process of the cell, such as cleaving mRNA at common motifs or damaging the cell membrane. Antitoxins inhibit the action of their toxins by binding to the toxin proteins themselves, or inhibiting them at a translational stage by sequestering their mRNA. Together with the other stress systems, it seems plausible that persistence is initiated by a mechanism whereby cells are slowed down by stress responses, then kept at a dormant state by activated TA systems (Balaban *et al.*, 2013).

1.2.5 Regulation and heterogeneity in motility

Motility in bacteria is broadly defined as the ability of cells to actively move, and includes some lesser considered modes of locomotion. These include twitching motility by pili, gliding motility by surface attachment points in the membrane, and sliding motility purely by physical displacement within a growing colony (Kearns, 2010). Motility in this thesis concerns the most commonly considered means—flagellar motility, by which the cell moves in liquid by rotating a long whip-like extracellular appendage.

The structure of the bacterial flagellum can be divided into two main parts—the filament and the hook-basal body. In some species, the process of assembling the flagellum is a highly ordered event, like in *Salmonella* Typhimurium where multiple classes of genes are activated just in time by waves of activation orchestrated by different sigma factors (Berg, 2003, Kutsukake *et al.*, 1994, Chadsey and Hughes, 2001). In *B. subtilis*, a single σ factor (σ^{D}) exists for the purposes of motility development, but some hierarchy is achieved by the addition of the σ^{D} co-factor SwrB to increase transcription of later stage genes (Kearns and Losick, 2005).

The basal body of the structure is organised into proteins polymerising into rings that are assembled from inside-out. At the centre is a membrane-embedded MS-ring providing attachment for the rod component that links to the filament itself (Berg, 2003). The MS-ring is situated in the cell membrane, and an additional ring of proteins called the C-ring is attached to its cytoplasm-facing side is responsible for accepting input from the chemotaxis system. Membrane-embedded stators form a loose ring formation on the outside of the MS-ring and provides rotation by turning the MS-ring. This provides torque via the rod and ultimately drives the rotation of the flagellum (Berg, 2003). The energy for rotation is generated by the proton motive force, and whether protons or other ions are used depends on the species. In *B. subtilis* the movement of protons (H^+) and sodium ions (Na^+) across the membrane drives the stators (Zhou *et al.*, 1998, Hirota and Imae, 1983).

The filament itself is composed of long, upwards of 10 μm polymers of predominantly a single protein (Hag, in the case of *B. subtilis*, named for being the H antigen in serotyping tests). Flagellin is threaded in an unfolded state through the hollow tube of the growing filament, and assumes its position at the end of the growing structure (Guttenplan *et al.*, 2012). Another important components of the filament is the hook (composed of the protein FlgE), a stiff and bent portion of the filament which translates the torque from the rod into larger movements. Finally, FliD serves as a cap for the filament, which is positioned at the end of the filament from the start of its growth, and directs polymerisation at the tip of the filament (Ikeda *et al.*, 1996). The method of assembly where the basal body constitutes a secretion apparatus hints at the likely evolutionary origins of the flagellum. The closely related Type III secretion system

found in some Gram-negative bacteria (including some members of the *Escherichia, Salmonella* and *Pseudomonas* genera) assembles a system of rings to create a passage spanning the periplasm, and exports proteins into eukaryotic host cells through a protruding syringe needle structure (Gerlach and Hensel, 2007).

Growing cultures of *B. subtilis* exhibit a bimodality with respect to motility, where the motility sigma factor σ^{D} is either expressed or not. Motile cells of a wild-type culture are readily identifiable without a fluorescent reporter construct, as they are physically detached from other cells and can move about freely if not immobilised on a microscope slide (Kearns and Losick, 2005). Conversely, non-motile ("sessile") cells do not as frequently detach from each other following cell division, and form long chains (Rashid and Sekiguchi, 1996).

Two subtly different motile behaviours exist with regard to flagellar motility; swimming and swarming motility. The first is simply defined as individual cells swimming in a liquid culture. Swarming motility is a more advanced phenotype which distinguishes itself by production of cell *rafts*—groups of cells joined together. Swarming involves hyperflagellation, which suggests that the inordinate amount of flagella may become entangled and contribute to intercellular adhesion. In addition to changes in flagellum production, a surfactant is also required for movement, as swarming takes place in a non-liquid environment (Angelini *et al.*, 2009). As is the case with biofilm formation, swarming is usually not a property observed in domesticated strains, because the aggressive spread of a colony is an undesirable trait in a bacterial model organism. Though a separate phenotype, the developmental networks of swarming and swimming motility are closely intertwined (Kearns and Losick, 2005).

The bulk of *B. subtilis* motility and chemotaxis genes are located in the 26 kbp *fla/che* operon, which is initially expressed at low levels by the housekeeping σ factor, σ^{A}. The penultimate ORF of the transcript encodes for the motility σ factor σ^{D}, which is maintained scarce due to the noise-prone low levels of *fla/che* transcript completion, but also by the anti-σ^{D} factor FlgM, which inactivates it by forming a heterodimer (Caramori *et al.*, 1996). The role of noise in transcript completion was demonstrated by moving *sigD* successively closer to the σ^{A}-dependent $\mathrm{P}_{fla/che}$ promoter, which resulted in increasingly larger proportions of motile sub-populations (Cozy and Kearns, 2010). The positive feedback loop in this bistable system is the autoregulation by σ^{D} from additional σ^{D}-dependent $\mathrm{P}_{fla/che}$ promoters; one located close to the original $\mathrm{P}_{fla/che}$ promoter, and another that has been mapped to the 8^{th} ORF (Cozy and Kearns, 2010). The shorter distance from promoter to *sigD* increases the bias towards transcript completion, and a positive feedback loop can eventually lock to maintain high σ^{D} levels.

Exactly where the line is drawn between swimming and swarming motility on the regulatory level is still unclear. σ^{D} on its own is not sufficient to induce motility (Kearns

and Losick, 2005), and two accessory transcription factors (SwrA, SwrB) exist to further enhance transcription of swimming and swarming motility genes (Kearns *et al.*, 2004). SwrB is located at the last position in the *fla/che* operon, and greatly enhances transcriptional activity from a promoter located at the beginning of the operon (Kearns and Losick, 2005). SwrA is non-functional in domesticated strains due to a frame shift mutation, but an insertion mutation at a repetitive region of A-T base pairs restores functionality, meaning that the gene is regulated through phase variation (Kearns *et al.*, 2004). Evidence shows that a surfactin production gene *sfp* which is mutated in domesticated strains will restore swarming if repaired, but not if surfactin is provided to the cell artificially (Kearns *et al.*, 2004). This suggests that the role of surfactin is not only that of reducing surface tension for non-liquid spread of a colony, but also for the purpose of signalling (López *et al.*, 2009a, Angelini *et al.*, 2009). A handful of σ^{D} regulon members are scattered outside the *fla/che* operon, including *hag*, and a few structural genes and genes encoding autolysins. Autolysins (*lytABC*) break down and remodel the peptidoglycan between chained cells, allowing them to physically separate (Smith *et al.*, 2000).

In contrast to more drastic changes in cell fate, such as sporulation or competence, no particular sacrifice in growth rate appears to be associated with being motile from an early stage of a culture (although swarming cells are observed to be longer) (Cozy and Kearns, 2010).

Lock-out mechanisms for limiting the expression of competing developmental pathways are common in the bistable responses of *B. subtilis*. Also, determining the appropriate time to end maintenance of such a phenotype is also important. Mechanical obstruction of the flagellum can be sensed as an environmental cue, and is integrated into the regulatory pathway by the kinase DegS, which activates DegU through phosphorylation, ultimately inhibiting further motility development (Cairns *et al.*, 2013, Belas, 2013). Another transitional mechanism is also provided by EpsE, which is a bifunctional enzyme that in addition to synthesising the biofilm EPS also functions as a clutch, which disengages the flagellum from its rotor (Blair *et al.*, 2008). DegU∼P promotes the production of more DegU and drives the production of biofilm matrix components (Marlow *et al.*, 2013). DegU∼P also disables further development of motility and competence, in that in its non-phosphorylated state it is a promoter of competence (Hsueh *et al.*, 2011, Hamoen *et al.*, 2000). As the DegU system does not reinforce itself by producing more phosphorylated DegU, the balance of DegU and DegU∼P makes the system more akin to a graded rheostat switch (analogous to a light dimmer) rather than an ON/OFF switch with only two possible states (Murray *et al.*, 2009). Like many regulators discussed thus far, the DegU-DegS two-component system does not limit itself to regulating biofilm formation, but it is also involved in generating

a bistable subpopulation of degradative enzyme-secreting cells (Veening *et al.*, 2008a, Mäder *et al.*, 2002).

1.3 Nutrient stress in *B. subtilis*

1.3.1 RelA and the stringent response

The stringent response is a large-scale regulatory response triggered by amino acid starvation and mediated by the small molecules pppGpp and ppGpp ((p)ppGpp). At times of nutrient limitation, the presence of unacetylated [uncharged] tRNAs will cause the ribosome to stall, which provides the initial cue for the (p)ppGpp synthetase RelA. The intracellular concentration of RelA has been determined to be 200-fold lower than that of ribosomes, which makes the speed at which critical levels of (p)ppGpp can be raised remarkable (Wendrich *et al.*, 2002). This quick response is due to the mechanism of action, where RelA recognises a ribosome/mRNA configuration, and if an unacetylated tRNA is present in the ribosome A site, (p)ppGpp is synthesised (Wendrich *et al.*, 2002). It should be noted that unlike the tmRNA system for rescuing stalled ribosomes, RelA does not rescue the ribosome, as the (p)ppGpp synthesis step simply releases RelA without disturbing the tRNA (Wendrich *et al.*, 2002). Because no requirements for its production are removed on synthesis and the reaction can by repeated many times, this enables the concentration of (p)ppGpp to quickly reach critical levels, as is required by the urgency of the situation that the stringent response signals.

In many studied systems including *E. coli*, an enzyme with a double synthetase/hydrolase role (SpoT) exists alongside another enzyme (RelA) which only synthesises (p)ppGpp (Magnusson *et al.*, 2005). No such counter-balancing enzyme has been identified in *B. subtilis*, where RelA is instead bifunctional, both synthesising and hydrolysing (p)ppGpp. The reaction equation shown below summarises the action of RelA:

$$\mathrm{ATP} + \mathrm{GDP} \underset{\text{RelA}}{\overset{\text{RelA}}{\rightleftharpoons}} \mathrm{AMP} + \mathrm{ppGpp}$$

Alongside RelA, two minor (p)ppGpp synthetases exist; YjbM and YwaC, termed small alarmone synthetases (SASs). They were identified as suppressor mutations arising in response to a Δ*relA* background (Srivatsan *et al.*, 2008). As a Δ*relA* background yields a total absence of (p)ppGpp upon artificial induction of the stringent response by addition of serine hydroxamate, their roles are likely unrelated to the stringent response (Wendrich and Marahiel, 1997). Nanamiya *et al.* (2008) found that the SAS genes play a role in environmental stress responses like alkaline shock. Their residual contribution of low (p)ppGpp levels in a Δ*relA* strain likely contributes to slower growth under normal

culture conditions. This can be seen in strain with deletions to all three (p)ppGpp genes, which partly regains the wild-type growth rate (Tagami *et al.*, 2012). This is not to be confused with the growth rate upon induction of the stringent response, where a Δ*relA* strain exhibits continued growth instead of halting, which is described as a *relaxed* phenotype, hence the name *relA*.

The stringent response affects nutrient spending on several levels, inhibiting DNA replication, transcription and production of ribosomes, as well as up-regulation of some *de novo* amino acid synthesis pathways (Wang *et al.*, 2007b). (p)ppGpp inhibits inosine monophosphate (IMP) dehydrogenase, the enzyme which catalyses the first step in the pathway for GTP production (Lopez *et al.*, 1981). Furthermore, the subsequent accumulation of IMP also leads to an increase in ATP, as IMP is a precursor of both GTP and ATP. The intracellular nucleotide pool is turned over very fast, and ceased synthesis of GTP leads to depletion within minutes (Werner, 1971). As a consequence of rapid GTP depletion, it is also important that genome replication ceases immediately, being an energy-intensive process. Additionally, replication needs to halt quickly because a GTP-deficient nucleotide pool may lead to an increased rate of mismatched base pairs being introduced by DNA polymerase (Kunz and Kohalmi, 1991). *B. subtilis* has two systems for arresting replication at times of nutrient stress. The first system acts through the replication termination protein RTP, which in addition to its role of terminating replication at the *ter* region, also gains the ability to stop replication during the stringent response at stop sites located close to *oriC* ($\sim$200-300 kbp) at either side (Autret *et al.*, 1999). Through a separate mechanism, (p)ppGpp directly targets the replication machinery through primase, which is a vital part of the replication machinery (Wang *et al.*, 2007a).

Gene regulation usually involves the binding of a transcription factor to operator sites of a promoter in order activate or repress transcription of a gene. Many members of the stringent response regulon are regulated by more global mechanisms that directly target transcription initiation on a nucleotide level. Initiation of transcription starts with an open complex of the DNA (promoter) and RNA Polymerase (RNAP), a configuration which is unstable until a ternary complex of DNA, RNAP and nascent mRNA has been assembled after the phosphodiester bond has formed between the +1 and +2 positions of the transcript (Walker *et al.*, 2004). The stringent response exerts control over all genes with guanosine as their initiating nucleoside triphosphate (iNTP). This affects transcription in two ways: firstly by direct competition for binding to RNAP by the related (p)ppGpp molecule, and secondly by also having made GTP scarcer through the inhibition of IMP dehydrogenase by (p)ppGpp (Krásný *et al.*, 2008, Jores and Wagner, 2003). Particularly relevant targets using this mechanism include rRNA and ribosomal subunit genes (Krásný *et al.*, 2008).

1.3.2 Regulation and operon organisation of ribosomal subunit production

The organisation of ribosomal protein subunits in *B. subtilis* is largely focused around a single 15 kbp S10-*spc*-α operon, named after the previously known homologous clusters from *E. coli*, where they are separate (Li *et al.*, 1997). Also different from *E. coli*, is the absence of *rpsD* from the α operon, whose product, S4 (protein 4 of the small ribosome subunit) autogenously regulates P_α (Sonenshein *et al.*, 1993, p. 676). While a ribosomal target for regulation by S4 also exists in *B. subtilis*, it is limited to its own monocistronic *rpsD* operon. The product inhibits translation by binding its own mRNA leader region due to a structural similarity to 16S rRNA, to which S4 binds in the assembled ribosome (Grundy and Henkin, 1991; 1992).

Aside from feedback control by subunits and rRNA, more global regulatory mechanisms affect the expression of ribosomal protein subunit and rRNA genes. It is logical that regulation of ribosome biogenesis is coupled to growth rate regulation, but the internal nucleotide pool and ppGpp concentrations also control expression. Being subject to rapid change as a response to overall nutrient levels, both the concentration of the iNTP of a transcript (often G in the case of ribosomal genes) and ppGpp contribute to controlling synthesis on a more global level, independent of transcription factors other than σ^A (the housekeeping σ factor) (Murray *et al.*, 2003).

1.3.3 The pleiotropic regulator CodY

CodY is a global regulator present in low G/C genome Gram-positive bacteria. When active, it represses genes appropriate for nutrient limitation, such as motility, competence and *de novo* amino acid synthesis (Molle *et al.*, 2003b). In some pathogenic strains, CodY also exerts control over virulence genes (Sonenshein, 2005). GTP and the branched-chain amino acids (BCAAs; L-isoleucine, L-leucine, or L-valine) are cofactors of CodY, and serve as an overall gauge of the nutrient levels in the cell, alleviating repression by CodY accordingly (Handke *et al.*, 2008). GTP is a particularly useful indicator of nutrient levels, as the intracellular nucleotide pool is turned over quickly as well as being an indicator of carbon, nitrogen, and phosphorus sources. Additionally, the sensitivity to GTP levels makes CodY an indirect effector of the stringent response (see section 1.3.1) leads to a drastic drop in GTP.

CodY forms a homodimer, and consists of an N-terminal GAF domain (present in cyclic nucleotide phosphodiesterases and adenylate cyclases) for cofactor binding, and a Helix-turn-Helix domain for DNA binding (Levdikov *et al.*, 2006). Cofactor binding to CodY is not competitive, meaning that CodY can accommodate both a BCAA and GTP, which increases CodY activity further than if only one of the two were present

(Handke *et al.*, 2008).

Δ*codY* strains exhibit increased sporulation rates (Handke *et al.*, 2008), and an inability to regulate competence or motility in response to nutrient levels (Serror and Sonenshein, 1996, Bergara *et al.*, 2003). In addition to its role in as a nutrient stress response regulator, CodY is one of several regulators responsible for controlling some major intersections of the metabolic network (Sonenshein, 2007). Together with CcpA, CodY has a role in the regulation of carbon metabolism, allowing for pyruvate to be redirected into lactate metabolism instead of entering the citric acid cycle under high glucose conditions (Sonenshein, 2007). A role in nitrogen metabolism together with TnrA is also known. Together they can inhibit the membrane transporter for arginine, and inhibit production by acting upstream of the pathways that control the availability of glutamine, which is the preferred nitrogen source in *B. subtilis* (Sonenshein, 2007).

1.4 Protein degradation in *B. subtilis*

Proteins are subject to quality control by mechanisms that ultimately result in their degradation by proteases if they are misfolded or denatured. The proteolytic action of AAA+ proteases can generally be classified into two constituent components with distinctive roles. A hexameric ring of ATPase recognises, unfolds, and translocates the protein into the internal space of the hexamer. A stack of two hexameric rings continues the tunnel started by the ATPase ring, and is made up of a peptidase with its catalytic site facing inward. In *B. subtilis*, nine proteins make up the seven known configurations of AAA+ proteases; the peptidase ClpP can pair with one of three APTases—ClpC, ClpE, or ClpX, and the peptidase ClpQ can pair with ClpY. In addition to the Clp proteins, three other closely related proteases exist where discrete functions are not assigned to separate proteins; LonA, LonB and FtsH (Molière and Turgay, 2013).

The quality control aspect of proteolysis is carried out throughout the lifetime of a protein. For example, a heat shock aimed at denaturing proteins has been shown to elicit increased proteolysis, as indicated by the appearance of fluorescent foci in strains where Clp proteases have been fused to fluorescent proteins (Kirstein *et al.*, 2008). Conversely, proteolysis may be required at the stage of translation in cases where termination has failed to occur. Such a case may for example arise where the appropriate acetylated tRNA has not been available to allow elongation to proceed, or the transcript has failed to produce a stop codon due to a frame shift mutation, premature transcript termination, or exoribonucleolytic degradation of the 3′ end that has resulted in the removal of the stop codon. The stalled ribosome is rescued by tmRNA, a small RNA species that mimics the ribosome binding of tRNA and carries a small ORF (*ssrA*). After binding to the ribosomal A site (facilitated by SmpB), tmRNA finishes elongation of the nascent

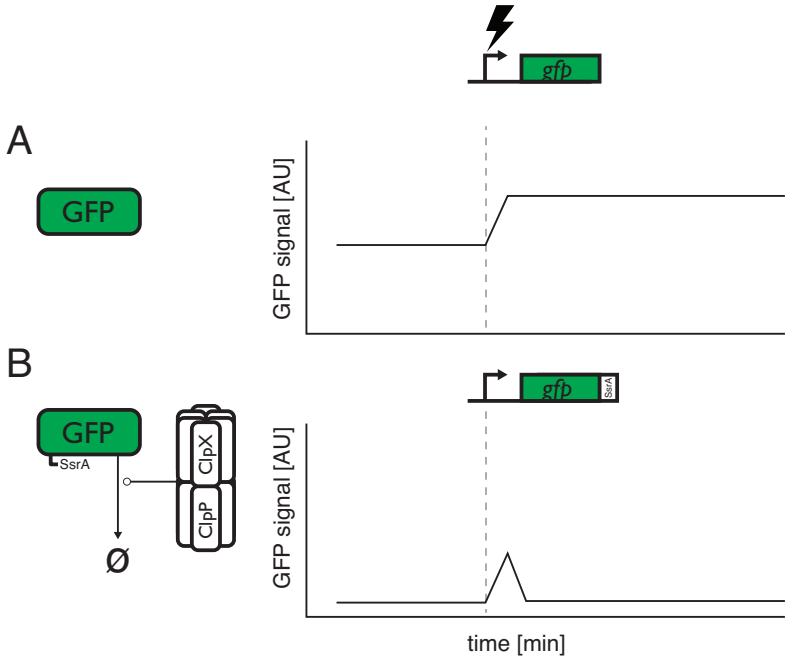

Figure 1.5: Use of unstable GFP to show rate of change and current transcription. Comparison between stable and unstable GFP (through SsrA-mediated proteolysis by ClpXP) (A) A stable GFP is not subject to significant degradation, and a burst in promoter activity has to be detected by measuring change in GFP signal (B) An unstable GFP construct only produces detectable signal for a short time after the promoter activity burst, as the GFP is quickly degraded. Periods of fluorescence in panel B relate to changes in GFP fluorescence in panel A, and thus eliminates the requirement for multiple observations to establish promoter activity.

peptide by shifting translation to *ssrA*, a process known as trans-translation (Withey and Friedman, 2003). tmRNA is already conjugated to the first alanine residue required by *ssrA*, and allows translation to proceed until termination, incorporating all 15 residues (in *B. subtilis*, AGKTNSFNQNVALAA) are incorporated (Ito *et al.*, 2002).

1.4.1 Use of SsrA to create proteolytically unstable GFP

Since its discovery in *Aequorea victoria*, the green fluorescent protein (GFP) has been improved to fit the needs of researchers conducting studies in a wide range of model organisms (Chudakov *et al.*, 2010). While it already folds and maturates without the requirement of any additional factors not present in other model organisms, the maturation rate of wild-type GFP is slow (Tsien, 1998). Some estimates suggest that signal would be detectable ∼2.5 hours after promoter activation (Leveau and Lindow, 2001). Modern GFP variants such as the superfolder GFP have been engineered to include several mutations with an effect on folding efficiency, maturation, and narrower excitation and emission spectra (Pédelacq *et al.*, 2005, Cormack *et al.*, 1996).

Once folded and maturated, GFP is a sturdy molecule with few exposed sites that

could serve as targets for proteases (Chiang *et al.*, 2001). This means that its half-life inside cells is unusually long, and any drop in GFP signal on the single-cell level would be primarily the result of dilution through growth. Tagging of GFP with the SsrA degron significantly reduces its half-life from hours to the order of minutes, depending on the variant of SsrA used. This addresses the problem with GFP signal persisting past the activity of the promoter (figure 1.5). The use of a wild-type SsrA tag is often not desired due to the efficacy of degradation eclipsing the maturation speed of GFP, but the last residues of SsrA can be modified to remedy this, effectively modulating the degradation efficiency (Andersen *et al.*, 1998).

The SsrA/ClpXP system itself is a relatively self-contained mechanism and does not require much supporting machinery. This makes it a candidate for an independent degradation machinery for destabilising substrate in organisms that lack a native SsrA system. The last four residues of SsrA are universally mapped to ClpX binding across species that possess the system, and where an adaptor protein is present, it binds to the first four residues. Contrary to *E. coli*, where SspB acts as an adaptor between SsrA and ClpX, *B. subtilis* SsrA binds directly to ClpX (Kirstein *et al.*, 2009). However, artificially introduced *E. coli* SspB has been shown to function with the *B. subtilis* ClpXP (Griffith and Grossman, 2008), and ClpXP from *E. coli* has conversely been transplanted into *Saccharomyces cerevisiae* without including SspB, where SsrA-tagged substrate was successfully degraded (Grilly *et al.*, 2007). Furthermore, a recent study by Sen *et al.* (2013) demonstrated ClpX unfolding activity of SsrA-tagged substrate *in vitro*, without the need for any other protein components other than ClpXP and the substrate.

1.5 Preliminary data: Transcriptome analysis of motile *B. subtilis*[1]

1.5.1 Introduction

Gene expression profiling through the use of DNA microarrays has been a frequently used tool in the study of bacterial transcriptomes. Microarrays allow for the measurement of the relative differences in transcript abundance between two populations. Typically, a reference population is compared to one that has been subject to a condition of interest, e.g., nutrient stress or the deletion of a regulatory mechanism. Transcription profiling of motile *B. subtilis* cells has thus far been limited to investigating the effect of deleting *sigD*, the gene encoding the motility sigma factor σ^{D} (Serizawa *et al.*, 2004). However, removing a key regulator such as σ^{D} can be considered an artificial situation, and expression profiling may not be revealing upstream regulatory events that feed into regulatory decisions relating to motility development. *B. subtilis* grows heterogeneously with respect to motility throughout the course of a liquid culture. The two developmental pathways are readily distinguishable from each other, and consist of non-motile chains of cells and freely swimming cells without cell-to-cell attachment by the cell wall at cell poles. Previous studies have used a P_{hag}-*gfp* construct to report σ^{D} activity by epifluorescence microscopy (Kearns and Losick, 2005). When collecting preliminary data for this project, a collaborating group used the construct to distinguish and separate motile cells from non-motile cells using fluorescence-activated cell sorting (FACS). This allowed the comparative transcriptome analysis of motile and non-motile subpopulations derived from the same culture. It also eliminated the requirement for artificial manipulation of motility by targeting σ^{D} levels through the deletion of a regulator.

1.5.2 Transcriptome analysis indicates an inhibition of ribosome synthesis during motility

This section describes the collection of transcriptome data relating to motile *B. subtilis*, and was conducted in order to identify potential co-regulated heterogeneously expressed genes which may have gone undetected without the use of single-cell methods to study gene expression. An exponentially growing culture of the P_{hag}-*gfp* strain DS901 (Kearns and Losick, 2005) was separated using FACS into high and low GFP signal subsamples, representing σ^{D}-ON and OFF states, respectively. A total of 3.00×10^7 and 1.00×10^7 particles were sorted for the GFP^+ and GFP^- subpopulations, though as the cell chains often prevalent in the GFP^- subpopulation will be counted as a single particle, this

[1]Raw data in this section was collected and statistically scored by Jojet Staal and Remco Kort—collaborators at TNO Quality of Life, The Netherlands. Analysis and presentation of results was carried out by the author as part of the thesis.

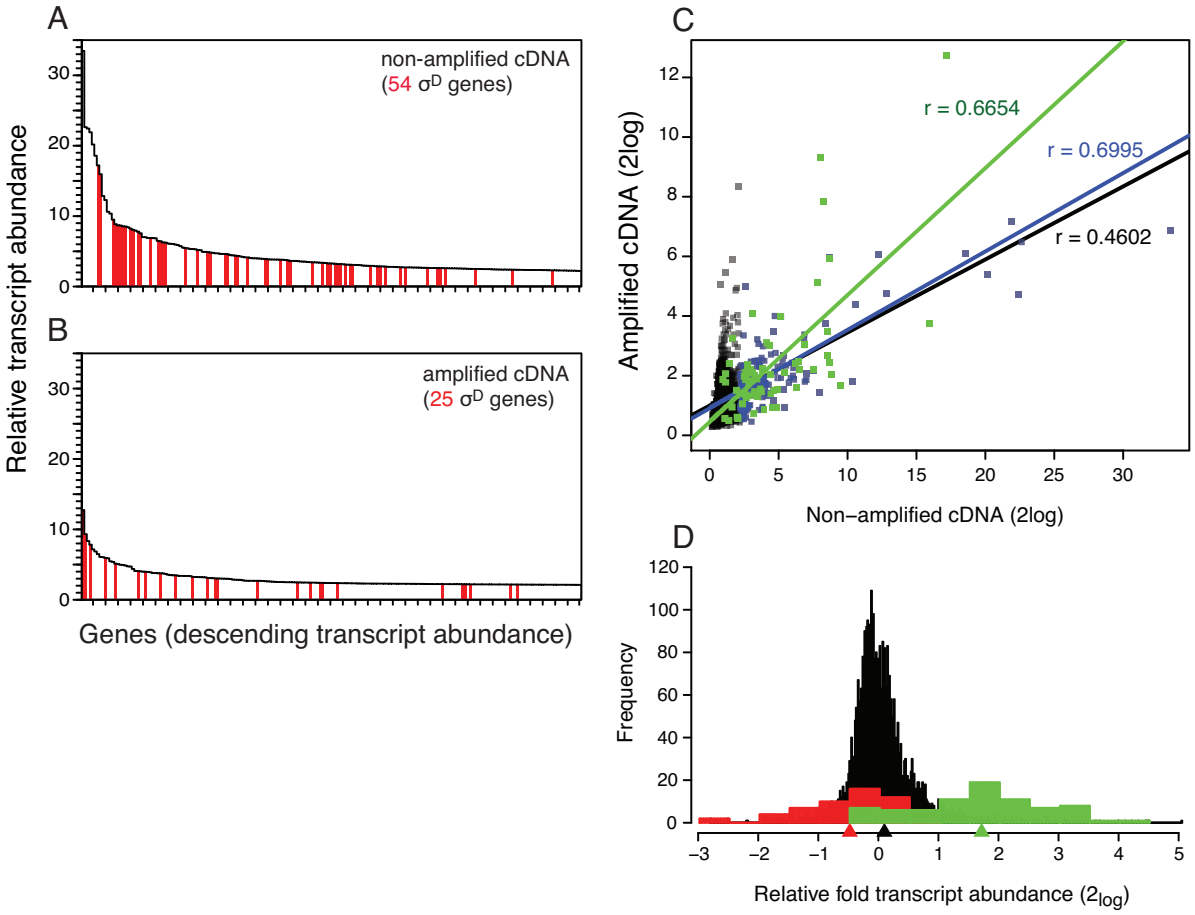

Figure 1.6: Effect of non-specific cDNA amplification on transcription profiling. Bar charts in A–B show the relative regulation of the top 200 up-regulated genes in a σ^{D}-ON population, red bars indicate member of the σ^{D} regulon. (A) In a sample where cDNA was not non-specifically amplified, 54 genes of the σ^{D} regulon were found among the top 200 genes. (B) Where non-specific amplification was carried out to increase cDNA yield, 25 σ^{D}-dependent genes were among the top 200 up-regulated genes. (C) Overall correlation between amplified and non-amplified cDNA samples on the whole genome level. Black points, line and r statistic indicate all genes; blue, top 200 up-regulated genes in the non-amplified sample; green, σ^{D}-dependent genes. (D) Histogram showing the relationship between relative transcript levels for ribosomal subunit genes (red), σ^{D}-dependent genes (green), and all 4,104 gene transcripts probed for (black). Arrowheads below the plot show the mean values of the groups potted in their corresponding colours.

Table 1.1: Table of the 252 most up-regulated ORFs in a motile sub-population of *B. subtilis*. σ^D-ON and OFF populations were separated according to their P_{hag}-*gfp* activity by FACS. Values correspond to the sample which was not amplified using Klenow fragment (see section 1.5.2 for details). Genes displayed in boldface are members of the σ^D regulon, genes up-regulated during the stringent response are underlined.

Gene	Identity/similarity/function	Regulation (fold change)	Variation (standard deviation)
hisI	phosphoribosyl-AMP cyclohydrolase/phosphoribosyl-ATP pyrophosphatase protein	33.46	7.38
hisB	imidazoleglycerol-phosphate dehydratase	22.63	3.40
hisA	1-(5-phosphoribosyl)-5-[(5-phosphoribosylamino) methylideneamino] imidazole-4-carboxamide isomerase	22.44	1.51
hisF	imidazole glycerol phosphate synthase subunit HisF	21.91	5.11
hisH	imidazole glycerol phosphate synthase subunit HisH	20.17	0.21
hisD	histidinol dehydrogenase	18.58	1.18
ylqB	**hypothetical protein**	17.21	0.37
hag	**flagellin protein**	15.95	0.84
hisZ	ATP phosphoribosyltransferase regulatory subunit	12.85	1.41
hisG	ATP phosphoribosyltransferase	12.28	2.42
yuiF	hypothetical protein	10.61	0.63
ilvC	ketol-acid reductoisomerase	10.35	2.87
flgL	**flagellar hook-associated protein**	9.50	0.62
yxkC	**hypothetical protein**	8.86	2.64
flgK	**flagellar hook-associated protein**	8.70	2.54
mcpA	**methyl-accepting chemotaxis protein**	8.67	2.31
fliT	**flagellar protein**	8.55	1.65
motB	**flagellar motor protein**	8.52	1.20
yjbJ	hypothetical protein	8.42	0.99
yfmT	**hypothetical protein**	8.23	0.99
yfmS	**hypothetical protein**	8.06	0.11
ilvH	acetolactate synthase small subunit	7.94	2.10
mcpB	**methyl-accepting chemotaxis protein**	7.79	1.31
fliD	**flagellar hook-associated protein**	7.56	0.29
xkdH	hypothetical protein	7.02	0.16
xkdJ	hypothetical protein	7.01	0.17
leuA	2-isopropylmalate synthase	6.90	3.42
motA	**flagellar motor protein**	6.89	0.48
leuB	3-isopropylmalate dehydrogenase	6.87	0.40
leuC	isopropylmalate isomerase large subunit	6.85	1.18
fliS	**flagellar protein FliS**	6.49	2.83
lytF	**gamma-D-glutamate-meso-diaminopimelate muropeptidase (major autolysin) (CWBP49')**	6.43	0.24
yjfB	**hypothetical protein**	6.27	0.79
yscB	**hypothetical protein**	6.24	0.41
hom	homoserine dehydrogenase	6.11	0.57
yvyD	hypothetical protein	6.07	0.40
xkdI	hypothetical protein	6.04	0.46
ywcJ	hypothetical protein	5.96	0.18
xkdK	hypothetical protein	5.88	1.58
ilvB	acetolactate synthase large subunit	5.74	1.35
xhlB	holin	5.54	2.50
yvyC	**flagellar protein FlaG**	5.45	0.50
dps	DNA-protecting protein	5.41	1.77

Continued on next page

Table 1.1 – *Continued from previous page*

Gene	Identity/similarity/function	Regulation (fold change)	Variation (standard deviation)
yuiA	hypothetical protein	5.34	2.68
yxaL	hypothetical protein	5.32	0.81
xhlA	involved in cell lysis upon induction of defective prophage PBSX	5.29	1.49
cheV	**modulation of CheA activity in response to attractants (chemotaxis)**	5.17	3.51
aroA	3-deoxy-7-phosphoheptulonate synthase	5.16	0.24
thrC	threonine synthase	4.93	2.06
albB	antilisterial bacteriocin (subtilosin) production	4.92	0.44
flgE	**flagellar hook protein**	4.88	2.19
ylxF	**hypothetical protein**	4.83	0.97
cydB	cytochrome bd ubiquinol oxidase (subunit II)	4.78	1.07
thrB	homoserine kinase	4.72	0.12
xkdG	hypothetical protein	4.72	0.13
ylxG	flagellar basal body rod modification protein	4.64	0.96
yxjG	hypothetical protein	4.63	1.70
flgN	**hypothetical protein**	4.63	1.36
flgM	**anti-sigma factor repressor of σ^{D}-dependent transcription**	4.57	0.06
yxiE	hypothetical protein	4.45	0.08
ybdO	**hypothetical protein**	4.42	1.21
fliM	**flagellar motor switch protein**	4.39	0.06
lytC	***N*-acetylmuramoyl-L-alanine amidase (major autolysin) (CWBP49)**	4.34	0.44
alsD	alpha-acetolactate decarboxylase	4.28	0.45
xkdP	hypothetical protein	4.20	0.26
ilvA	threonine dehydratase	4.14	0.04
hemAT	**haem-based aerotactic transducer**	4.14	1.65
ywaA	branched-chain amino acid aminotransferase	4.04	0.11
xkdM	hypothetical protein	4.01	1.03
aroH	chorismate mutase (isozymes 1 and 2)	4.01	0.92
cheY	two-component response regulator	3.98	1.26
xkdQ	hypothetical protein	3.96	2.24
xkdF	hypothetical protein	3.94	0.86
fliG	**flagellar motor protein**	3.92	1.83
cheD	**required for methylation of methyl-accepting chemotaxis proteins (MCPs) by CheR**	3.86	0.52
cydC	ABC membrane transporter (ATP-binding protein)	3.84	0.04
aroB	3-dehydroquinate synthase	3.84	0.41
yuaI	hypothetical protein	3.83	0.54
yvyF	hypothetical protein	3.82	0.09
lytA	**membrane bound lipoprotein**	3.81	1.80
csrA	carbon storage regulator	3.81	0.26
xlyA	*N*-acetylmuramoyl-L-alanine amidase	3.76	0.42
fliZ	**flagellar protein**	3.72	0.19
fliI	**flagellum-specific ATP synthase**	3.65	0.35
alsS	alpha-acetolactate synthase	3.64	1.21
lctP	L-lactate permease	3.64	0.41
ilvD	dihydroxy-acid dehydratase	3.61	0.18

Continued on next page

Table 1.1 – *Continued from previous page*

Gene	Identity/similarity/function	Regulation (fold change)	Variation (standard deviation)
lytB	modifier protein of major autolysin LytC (CWBP76)	3.55	1.52
fliF	**flagellar M-ring protein**	3.55	0.82
ydbM	hypothetical protein	3.53	1.19
albA	antilisterial bacteriocin (subtilosin) production	3.48	0.33
leuD	3-isopropylmalate dehydratase (small subunit)	3.45	0.37
fliP	**flagellar biosynthesis protein**	3.43	2.05
yvzD	hypothetical protein	3.43	0.45
cydD	ABC membrane transporter (ATP-binding protein)	3.42	1.27
xkdE	hypothetical protein	3.38	1.34
flhB	**flagellar biosynthesis protein**	3.37	0.56
fliK	**flagellar hook-length control**	3.34	0.55
flhP	**flagellar hook-basal body protein**	3.28	0.01
sigD	**sigma factor σ^{D}**	3.28	1.52
albC	antilisterial bacteriocin (subtilosin) production	3.26	0.98
fliY	**flagellar motor switch protein**	3.23	1.11
fliH	**flagellar assembly protein**	3.19	0.16
ydjH	hypothetical protein	3.18	0.95
bcd	leucine dehydrogenase	3.16	0.82
yviE	hypothetical protein	3.12	0.77
xkdS	hypothetical protein	3.11	1.43
mcpC	**methyl-accepting chemotaxis protein**	3.10	0.68
flhO	**flagellar basal-body rod protein**	3.07	0.42
trpA	tryptophan synthase subunit alpha	3.07	0.14
serC	phosphoserine aminotransferase	3.05	0.27
yxxG	hypothetical protein	3.03	0.88
ldh	L-lactate dehydrogenase	2.98	0.30
xkdO	hypothetical protein	2.96	1.51
cydA	cytochrome bd ubiquinol oxidase (subunit I)	2.95	0.12
lytD	N-acetylglucosaminidase (major autolysin) (CWBP90)	2.93	1.02
yuiB	hypothetical protein	2.93	0.21
ycnK	hypothetical protein	2.88	1.46
yviF	hypothetical protein	2.86	0.37
ydjI	hypothetical protein	2.86	0.26
yrvJ	hypothetical protein	2.85	0.01
cheA	two-component sensor histidine kinase	2.83	0.06
ycnJ	hypothetical protein	2.80	0.40
oppA	oligopeptide ABC transporter (binding protein)	2.79	0.65
lytR	membrane-bound transcriptional regulator LytR	2.79	0.79
ycnI	hypothetical protein	2.78	0.17
xkdR	hypothetical protein	2.75	0.53
epr	extracellular serine protease	2.74	0.69
yabO	hypothetical protein	2.72	0.26
cheC	**inhibition of CheR-mediated methylation of methyl-accepting chemotaxis proteins (MCPs) through control of the binding of CheD to the MCPs**	2.72	0.07
xtmA	PBSX defective prophage terminase (small subunit)	2.72	1.03
oppD	oligopeptide ABC transporter (ATP-binding protein)	2.71	0.66
bdbC	thiol-disulfide oxidoreductase	2.70	0.80
cysP	sulfate permease	2.70	0.52

Continued on next page

Table 1.1 – *Continued from previous page*

Gene	Identity/similarity/function	Regulation (fold change)	Variation (standard deviation)
cysK	cysteine synthetase A	2.69	0.01
xkdU	hypothetical protein	2.68	0.96
ybaJ	hypothetical protein	2.66	0.13
fabHB	3-oxoacyl-(acyl carrier protein) synthase	2.66	1.75
ydjG	hypothetical protein	2.65	0.76
yebD	hypothetical protein	2.63	0.43
yydF	hypothetical protein	2.63	0.61
sucC	succinyl-CoA synthetase subunit beta	2.62	1.09
ylxL	hypothetical protein	2.61	0.69
cheB	chemotaxis-specific methylesterase	2.61	0.46
aroF	chorismate synthase	2.61	0.52
yjcQ	hypothetical protein	2.59	0.10
ypdC	hypothetical protein	2.59	0.34
yxjI	hypothetical protein	2.58	0.55
tyrA	prephenate dehydrogenase	2.58	1.00
albD	antilisterial bacteriocin (subtilosin) production	2.58	0.24
ysbB	antiholin-like protein LrgB	2.57	0.56
xkdT	hypothetical protein	2.56	0.59
fhuD	ferrichrome ABC transporter (ferrichrome-binding protein)	2.55	0.55
flhF	**flagellar biosynthesis protein**	2.54	0.77
hemH	ferrochelatase	2.52	0.15
xtmB	PBSX defective prophage terminase (large subunit)	2.51	0.41
hpr	transcriptional regulator (MarR family)	2.50	0.83
cheW	**modulation of CheA activity in response to attractants (chemotaxis)**	2.50	0.09
xkdN	hypothetical protein	2.49	1.00
yplQ	hypothetical protein	2.49	0.85
ykzG	hypothetical protein	2.48	1.12
ylxH	hypothetical protein	2.48	1.06
phrF	regulator of the activity of phosphatase RapF	2.48	0.17
appB	oligopeptide ABC transporter (permease)	2.46	1.15
ydgF	hypothetical protein	2.44	1.51
yjcF	hypothetical protein	2.42	0.74
serA	phosphoglycerate dehydrogenase	2.42	0.07
xepA	lytic exoenzyme associated with defective prophage PBSX	2.41	0.26
dat	D-alanine aminotransferase	2.41	0.90
yvbX	hypothetical protein	2.40	0.00
lysA	diaminopimelate decarboxylase	2.40	0.78
buk	butyrate kinase	2.40	0.80
pspA	phage shock protein A homolog	2.39	0.50
rapF	response regulator aspartate phosphatase	2.38	0.07
ywdK	hypothetical protein	2.38	1.09
ylmG	hypothetical protein	2.36	1.22
xlyB	N-acetylmuramoyl-L-alanine amidase	2.36	0.23
sat	sulfate adenylyltransferase	2.36	0.25
ycsD	hypothetical protein	2.35	0.18
yxzC	hypothetical protein	2.35	0.77
rnhB	ribonuclease HII	2.35	0.44
xkdV	hypothetical protein	2.34	0.21

Continued on next page

Table 1.1 – *Continued from previous page*

Gene	Identity/similarity/function	Regulation (fold change)	Variation (standard deviation)
yjcN	hypothetical protein	2.34	0.17
yteJ	hypothetical protein	2.34	0.88
ptb	phosphate acetyltransferase	2.33	0.71
nucA	nuclease	2.33	1.27
ylnD	hypothetical protein	2.31	0.31
yydA	hypothetical protein	2.31	0.07
fliL	**flagellar basal body-associated protein**	2.31	0.21
ywfH	hypothetical protein	2.30	0.23
yppF	hypothetical protein	2.28	0.37
hisC	histidinol-phosphate aminotransferase	2.28	0.88
yvjD	hypothetical protein	2.28	0.72
flhA	**flagellar biosynthesis protein**	2.28	0.07
yvfI	hypothetical protein	2.26	0.19
resB	required for cytochrome c synthesis	2.25	1.16
ycbJ	hypothetical protein	2.25	0.06
oppF	oligopeptide ABC transporter (ATP-binding protein)	2.25	0.56
lexA	LexA repressor	2.22	0.63
yhdT	hypothetical protein	2.21	0.18
oppC	oligopeptide ABC transporter (permease)	2.21	0.07
resC	required for cytochrome c synthesis	2.20	0.17
ylqH	hypothetical protein	2.20	0.44
opuCB	glycine betaine/carnitine/choline ABC transporter (membrane protein)	2.19	0.12
xkdW	hypothetical protein	2.19	0.25
yyaK	hypothetical protein	2.19	0.77
yaaK	hypothetical protein	2.18	0.37
yeeI	hypothetical protein	2.17	0.28
aspS	aspartyl-tRNA synthetase	2.17	0.10
ytxG	hypothetical protein	2.16	0.10
ezrA	septation ring formation regulator EzrA	2.16	0.34
yerI	hypothetical protein	2.15	0.10
murAB	UDP-N-acetylglucosamine 1-carboxyvinyltransferase	2.13	0.32
yclQ	hypothetical protein	2.13	0.03
yceC	hypothetical protein	2.12	0.36
yjcM	hypothetical protein	2.12	0.35
acpD	acyl carrier protein phosphodiesterase	2.12	1.66
asd	aspartate-semialdehyde dehydrogenase	2.12	0.03
gudB	glutamate dehydrogenase	2.11	0.99
yczJ	hypothetical protein	2.11	1.06
yrvI	D-tyrosyl-tRNA deacylase	2.11	0.09
sucD	succinyl-CoA synthetase alpha subunit	2.09	0.08
ysbA	murein hydrolase regulator LrgA	2.09	0.17
yutF	hypothetical protein	2.08	0.16
yvcL	hypothetical protein	2.08	0.82
appC	oligopeptide ABC transporter (permease)	2.08	0.15
clpC	class III stress response-related ATPase	2.08	0.73
yuaF	hypothetical protein	2.08	0.36
aroE	3-phosphoshikimate 1-carboxyvinyltransferase	2.08	1.14
ydbL	hypothetical protein	2.08	0.48

Continued on next page

Table 1.1 – *Continued from previous page*

Gene	Identity/similarity/function	Regulation (fold change)	Variation (standard deviation)
dra	deoxyribose-phosphate aldolase	2.06	0.22
yjcG	hypothetical protein	2.06	0.15
yodF	hypothetical protein	2.05	0.21
yaaN	hypothetical protein	2.04	0.03
yqaP	hypothetical protein	2.04	0.12
flgC	**flagellar basal body rod protein**	2.04	0.43
appA	oligopeptide ABC transporter (oligopeptide-binding protein)	2.04	0.10
ydbN	hypothetical protein	2.03	0.10
yvlB	hypothetical protein	2.03	0.32
ydiS	hypothetical protein	2.02	0.52
yjcI	cystathionine beta-lyase	2.02	0.21
yloV	hypothetical protein	2.02	0.42
oppB	oligopeptide ABC transporter (permease)	2.02	0.41
ywzB	hypothetical protein	2.02	1.15
spo0J	site-specific DNA-binding protein	2.01	0.33
ydjP	hypothetical protein	2.01	1.00
hisS	histidyl-tRNA synthetase	2.01	0.54
hemA	glutamyl-tRNA reductase	2.01	0.34
pdp	pyrimidine-nucleoside phosphorylase	2.01	0.33
ywfI	hypothetical protein	2.01	0.06
yetG	hypothetical protein	2.01	0.36
yvrP	hypothetical protein	2.00	0.02

Table 1.2: Table of the 95 most down-regulated ORFs in a motile sub-population of *B. subtilis*. σ^{D}-ON and OFF populations were separated according to their P_{hag}-*gfp* activity by FACS. Values correspond to the sample which was not amplified using Klenow fragment (see section 1.5.2 for details). Ribosomal protein subunit genes are displayed in boldface, members of the stringent response regulon are underlined

Gene	Identity/similarity/function	Regulation (fold change)	Variation (standard deviation)
pyrAA	carbamoyl-phosphate synthase small subunit	0.14	0.02
pyrB	aspartate carbamoyltransferase catalytic subunit	0.14	0.02
yukA	hypothetical protein	0.15	0.06
mrgA	metalloregulation DNA-binding stress protein	0.15	0.06
rpsR	**ribosomal protein S18**	0.18	0.02
rpsD	**30S ribosomal protein S4**	0.18	0.04
pyrC	dihydroorotase	0.22	0.04
yueB	hypothetical protein	0.22	0.06
purB	adenylosuccinate lyase	0.26	0.01
yukD	hypothetical protein	0.27	0.02
gltB	glutamate synthase (small subunit)	0.27	0.04
ycdA	hypothetical protein	0.27	0.07

Continued on next page

Table 1.2 – *Continued from previous page*

Gene	Identity/similarity/function	Regulation (fold change)	Variation (standard deviation)
pyrK	dihydroorotate dehydrogenase electron transfer subunit	0.27	0.01
pyrD	dihydroorotate dehydrogenase	0.28	0.05
ytiP	hypothetical protein	0.28	0.03
purC	phosphoribosylaminoimidazole-succinocarboxamidesynthase	0.28	0.08
yueC	hypothetical protein	0.28	0.02
purS	phosphoribosylformylglycinamidine synthase	0.29	0.01
yrrN	hypothetical protein	0.29	0.02
pyrP	uracil permease	0.29	0.02
speB	agmatinase	0.30	0.04
yxaI	hypothetical protein	0.30	0.11
katA	vegetative catalase 1	0.31	0.01
pyrE	orotate phosphoribosyltransferase	0.32	0.02
pbuX	xanthine permease	0.32	0.10
purQ	phosphoribosylformylglycinamidine synthase subunit I	0.33	0.01
rplC	**50S ribosomal protein L3**	0.33	0.00
yrrO	hypothetical protein	0.33	0.13
rpmGA	**highly similar to ribosomal protein L33**	0.33	0.11
udk	uridine kinase	0.33	0.10
glnR	transcriptional regulator	0.34	0.08
tenA	transcriptional regulator	0.34	0.06
rpsO	**30S ribosomal protein S15**	0.35	0.02
yflE	hypothetical protein	0.35	0.16
rpsJ	**30S ribosomal protein S10**	0.35	0.00
yukB	hypothetical protein	0.36	0.05
yrrM	hypothetical protein	0.36	0.14
purE	phosphoribosylaminoimidazole carboxylase catalytic subunit	0.36	0.15
yukE	hypothetical protein	0.37	0.07
purL	phosphoribosylformylglycinamidine synthase subunit II	0.37	0.01
bslA	hypothetical protein	0.37	0.21
pit	low-affinity inorganic phosphate transporter	0.37	0.12
ytmQ	tRNA (guanine-$N(7)$-)-methyltransferase	0.37	0.06
purK	phosphoribosylaminoimidazole carboxylase II	0.37	0.08
purF	amidophosphoribosyltransferase	0.38	0.05
rpsT	**ribosomal protein S20 (BS20)**	0.39	0.04
pyrR	pyrimidine regulatory protein PyrR	0.40	0.03
nadC	nicotinate-nucleotide pyrophosphorylase	0.40	0.02
ylxM	hypothetical protein	0.40	0.01
rplA	**50S ribosomal protein L1**	0.40	0.02
yefA	hypothetical protein	0.41	0.08
thiS	hydroxyethylthiazole phosphate biosynthesis (thiamine biosynthesis)	0.41	0.10
ywnC	hypothetical protein	0.41	0.02
rplD	**50S ribosomal protein L4**	0.42	0.02
yxjA	hypothetical protein	0.42	0.20
guaC	guanosine 5′-monophosphate oxidoreductase	0.42	0.09
rplU	**ribosomal protein L21 (BL20)**	0.42	0.06
goxB	glycine oxidase	0.43	0.04

Continued on next page

Table 1.2 – *Continued from previous page*

Gene	Identity/similarity/function	Regulation (fold change)	Variation (standard deviation)
ylqC	hypothetical protein	0.43	0.02
yjhA	hypothetical protein	0.43	0.01
ykaA	hypothetical protein	0.43	0.23
nadB	L-aspartate oxidase	0.44	0.16
thiC	thiamine biosynthesis protein ThiC	0.44	0.19
purA	adenylosuccinate synthetase	0.44	0.05
yukC	hypothetical protein	0.44	0.01
pyrG	CTP synthetase	0.44	0.07
yebC	hypothetical protein	0.44	0.05
xpt	xanthine phosphoribosyltransferase	0.44	0.02
pbuG	hypoxanthine/guanine permease	0.44	0.03
ykbA	hypothetical protein	0.44	0.08
pyrF	orotidine 5′-phosphate decarboxylase	0.45	0.03
speD	S-adenosylmethionine decarboxylase proenzyme	0.45	0.10
speA	arginine decarboxylase	0.45	0.02
safA	morphogenetic protein associated with SpoVID	0.45	0.15
groEL	chaperonin GroEL	0.45	0.10
accB	acetyl-CoA carboxylase	0.46	0.16
yqdB	hypothetical protein	0.46	0.11
ywqL	hypothetical protein	0.46	0.12
murD	UDP-*N*-acetylmuramoyl-L-alanyl-D-glutamatesynthetase	0.46	0.20
yczI	hypothetical protein	0.46	0.03
rpsF	**ribosomal protein S6 (BS9)**	0.47	0.04
gcaD	UDP-*N*-acetylglucosamine pyrophosphorylase	0.47	0.07
carB	carbamoyl-phosphate transferase-arginine (subunit B)	0.47	0.44
ffh	signal recognition particle-like (SRP) component	0.49	0.03
purN	phosphoribosylglycinamide formyltransferase	0.49	0.18
pstS	phosphate ABC transporter (binding protein)	0.49	0.23
rplW	**50S ribosomal protein L23**	0.49	0.02
comZ	late competence gene	0.49	0.20
ytpS	hypothetical protein	0.49	0.01
rocA	pyrroline-5 carboxylate dehydrogenase	0.49	0.05
yjbV	hypothetical protein	0.50	0.03
purM	phosphoribosylaminoimidazole synthetase	0.50	0.01
yqeT	ribosomal protein L11 methyltransferase	0.50	0.08
ylaH	hypothetical protein	0.50	0.05
rpsS	**ribosomal protein S19 (BS19)**	0.50	0.00

figure does not carry much significance. cDNA from the two samples was hybridised to an expression profiling microarray with probes for 4,104 known and predicted *B. subtilis* genes, each appearing 4 times as technical replicates (Kunst *et al.*, 1997). However, the biomass separated by FACS yielded less than the amount of mRNA (1 μg) recommended for use in a microarray experiment. To address this concern, cDNA was non-specifically amplified using Klenow fragment with random oligonucleotides as primers. To assess the impact of non-specific amplification on the quality of data, the

presence of transcripts from the σ^D regulon (as identified by Serizawa *et al.* (2004)) was used as an indicator of quality. Figure 1.6A–B compares non-amplified and amplified samples for the presence of σ^D-dependent ORF transcripts in the 200 most abundantly detected transcripts in a σ^D-ON population, relative to that of cells in a σ^D-OFF state. A non-amplified cDNA yield appears to be more accurate on the grounds that in the 200 most abundant transcripts, there are more than double the number of σ^D-dependent genes (54) compared to that of an amplified sample (25) (figure 1.6A-B). Therefore, it seemed that non-specific amplification was detrimental to the quality of the data. Furthermore, the cDNA of the non-amplified population does not appear to have been proportionally amplified. Figure 1.6C shows this by comparing the relative levels of the 200 most abundant ORFs in the non-amplified sample compared to their levels in an amplified sample, showing low levels of correlation.

The significance of gene regulation change was scored using T-profiler (Boorsma *et al.*, 2005), and as expected, many transcripts relating to motility are found among those up-regulated in a σ^D-ON population (table 1.1 shows the 252 significantly up-regulated transcripts in a motile subpopulation). Interestingly, a large number of ribosomal subunit genes are found in the significantly down-regulated portion (consisting of 95 genes) of the σ^D-ON transcriptome (table 1.2 and figure 1.6D). Out of 53 ribosomal protein subunit genes present on the *B. subtilis* genome, 13 appeared in the portion of the dataset deemed to be significantly down-regulated, and none in the up-regulated group. This reciprocal relationship between motility and down-regulation of biomass generation seems logical, because motility is a response to nutrient stress. However, as the apparent down-regulation in ribosomal promoter activity of motile cells occurs during the exponential growth phase—when nutrients are plentiful—this relationship was further studied.

1.6 Aims

The main aim of this thesis was to test the hypothesis that the motile subpopulation of cells in a growing culture of *B. subtilis* exhibited a higher rate expression from ribosomal subunit promoters. This hypothesis was based on preliminary data collected by collaborators, which suggested that following separation of motile from non-motile cells in a growing culture, a difference in transcript abundance from ribosomal subunit genes was present. In order to investigate this hypothesis, populations were investigated on the single-cell level using epifluorescence microscopy. This required the use of proteolytically unstable GFP as a reporter to overcome the problem of excess production of a stable reporter protein under control of a strong promoter. A possible regulatory cause for the hypothesised difference between motile and non-motile cells was also investigated, with

the alternative hypothesis that such changes in ribosome production would be brought on by more global cellular events such as nutrient availability.

The methods used for analysing the type of microscopy data generated by this project are preferably analysed by means of automated image processing. However, standard methods focus on segmentation of cells using the assumption that cells physically segregate on division—a criteria that *B. subtilis* does not satisfy under certain growth conditions. As such, another aim of this thesis is to establish automated analysis methods which will work well with the cell chaining phenotype of *B. subtilis*.

Chapter 2

Materials and Methods

2.1 Strains and Media

Any strains mentioned in this thesis are listed in table 2.1. For *B. subtilis*, the laboratory strain 168CA was used unless stated specifically otherwise. Where necessary, solid or liquid media was supplemented with antibiotics (50 μg/ml spectinomycin (spc), 2 μg/ml kanamycin (km), 10 μg/ml tetracycline (tet), 5 μg/ml chloramphenicol (cam), 2 μg/ml erythromycin (erm)). 1% Nutrient agar (Oxoid) was used for growth on solid media.

Genetic transformation of *B. subtilis* was done by inducing natural competence using a method based on the work of Anagnostopoulos and Spizizen (1961). Competence media consisted of Spizizen minimal media (SMM) (0.2% $(NH_4)_2SO_4$, 1.4% K_2HPO_4, 0.6% KH_2PO_4, 0.1% $Na_3C_6H_5O_7$, 0.02% $MgSO_4$), supplemented with 0.5% glucose, casamino acids, ferric ammonium citrate, and tryptophan. Starvation media consisted of SMM and glucose at the aforementioned concentrations. Overnight cultures were grown in competence media and diluted 1:20 into fresh media, after 3 hours of shaking growth at 37 °C, pre-warmed starvation media was added to the culture at a 1:1 ratio. After 1 hour of starvation, 400 μl of culture was added to a a 2 ml Eppendorf tube containing donor strain chDNA or integrating plasmid, and allowed to shake for 1 hour before spreading on nutrient agar plates containing the appropriate antibiotic.

B. subtilis cultures for microscopy were grown in casein hydrolysate (CH) media consisting of 10.64 g/l casein hydrolysate, 3.83 g/l L-glutamic acid (sodium salt), 1.33 g/l L-alanine, 1.48 g/l L-asparagine, 1.45 g/l KH_2PO_4, 0.57 g/l NH_4Cl, 114 mg/l Na_2SO_4, 102 mg/l NH_4NO_3, 1.04 mg/l $FeCl_3{\cdot}6\,H_2O$, adjusted to pH 7.0 and autoclaved. Prior to use, CH was supplemented with 0.1 mM $CaCl_2$, 0.4 mM $MgSO_4$, 0.15 mM $MnSO_4$, 20 μg/ml tryptophan (Harwood and Cutting, 1990, p. 396–7).

2.2 Molecular cloning

Unless stated otherwise, the *E. coli* strain DH5α was used for amplification of plasmids, grown at 37 °C shaking in Luria-Bertani (LB) media supplemented with 50 μg/ml ampicillin. Purification of plasmid DNA from *E. coli* was performed using a QIAprep Spin Miniprep Kit (QIAGEN). Genetic transformation of chemically competent ($CaCl_2$) *E. coli* with plasmids was carried out using the method of Dagert and Ehrlich (1979). In brief, transformation steps consisted of incubating cells on ice with plasmid DNA, heat shocking briefly at 42 °C, then chilling on ice again before phenotypic expression at 37 °C in LB and plating on amp nutrient agar plates. Where the use of DNA methylation-sensitive restriction enzymes was required, the plasmid was purified from *E. coli* strain ER2925.

Polymerase chain reactions (PCRs) were carried out using Phusion High-Fidelity DNA Polymerase (New England BioLabs). Where PCR reactions are described and only melting temperatures and extension times are stated, other timings and temperatures and concentrations are those suggested by the supplier. Oligonucleotide primers were supplied by Eurogentec (Liège, Belgium) and were designed to have a t_m of or close to 60 °C according to the simplified equation $t_m = 4S + 2W$, where S and W denote the number G/C and A/T base pairs, respectively (Rychlik and Rhoads, 1989). 168CA chDNA was used as a template when amplifying DNA from the *B. subtilis* chromosome unless otherwise stated, using the genome available through the SubtiList genome browser as a reference (Moszer *et al.*, 1995).

Purification of PCR product or digested DNA was carried out using a QIAquick PCR Purification Kit, or a QIAquick Gel Extraction Kit (QIAGEN) when separation of DNA fragments using gel electrophoresis was required prior to purification. DNA gels were made with 1% agarose in TAE buffer, stained with 0.5 μg/ml ethidium bromide (Sigma), and if necessary visualised under ultraviolet light using a GelDoc (BioRad). Restriction enzymes were provided by Roche, Promega or NEB. Ligations were performed using T4 DNA ligase (Roche) in 25 μl reactions containing provided buffer, 2 μl ligase, and 10 μl mix of 1:1 vector:insert unless quantification shows one of the two to be of a disproportionally low concentration. Reactions are incubated at room temperature for 15 or 30 min (for sticky- and blunt-ended ligations, respectively) and heat-inactivated for 20 min at 65 °C before transforming into DH5α *E. coli*.

Constructs were verified by restriction digests targeting both the vector and predicted insert, then visualised using gel electrophoresis. Colonies showing successful cloning were further verified on a nucleotide level by Sanger sequencing (DNA Sequencing and services, Dundee, UK), covering restriction sites used for ligation, as well as regions amplified by PCR.

Where *amyE* was used as the ectopic target locus for double crossover integration, successful integration was tested by growing *B. subtilis* strains on NA plates supplemented with 0.5% starch (Sigma), and detected by staining with iodine (Fluka). When targeting the *aprE* locus, integration was verified by PCR, using primer pair KS92/KS93 (60 °C, 2 min 30 sec).

B. subtilis chromosomal DNA (chDNA) was prepared using a modified protocol based on the method described by Harwood and Cutting (1990, p. 145). The strain was cultured overnight at 30 °C in selective LB supplemented with 0.5% glucose, after which 2 ml of culture was washed twice in TES buffer. The sample was resuspended to a final volume of 750 μl in TES, incubated for 5 min at 37 °C with 0.33 mg/ml lysozyme, then a further 30-60 min with 0.66 mg/ml pronase and 1.2% sarkosyl (Sodium lauroyl sarcosinate, Sigma). Phenol and chloroform were added to the sample, mixed vigorously and separated out by centrifugation. A further chloroform step was carried out on the upper phase, then mixed, centrifuged, and separated again. DNA from the upper phase was precipitated in 100% ethanol and resuspended in 200 μl TES buffer.

2.2.1 Site-directed mutagenesis by linearising PCR primers

Site-directed mutagenesis and insertions of sequences too short to ligate as a fragment was performed by linearising the plasmid using divergent primer pairs carrying the desired modification in 5′ overhangs. PCR was carried out at a $t_m = 55$ °C and an extension time of 30 sec/kb, after which excess template was degraded using *Dpn*I and cleaned up using a PCR purification kit. The 5′ ends of the PCR product were kinated using T4 polynucleotide kinase (PNK) (New England BioLabs) in a 20 μl reaction, then ligated by sizing up the reaction to 40 μl and adding T4 DNA Ligase (Roche) and proceeding with ligation, heat inactivation and transformation as previously described (section 2.2).

2.2.2 Construction of a gfp/lacZ-carrying transposon delivery vector

A unique cloning site in pMarB (Le Breton *et al.*, 2006) was created by linearising the plasmid with the divergent PCR primer pair oSS330/oSS331 (55 °C, 4 min), both carrying non-complementary 5′ overhangs with the *Mun*I recognition sequence. The PCR product was purified and its ends were treated with *Mun*I, purified again and ligated using T4 DNA ligase (Roche). This plasmid was designated pSS121.

The components of the core reporter constructs were assembled using SOEing (Splicing by Overlap Extension) PCR (Horton, 1995) on DNA fragments with ∼20-nucleotide 3′ G/C-clamped overlaps of a calculated t_m of 60 °C (see page 38). *sfGFP* was amplified from pSG1729-sfGFP using PCR primer pair oSS346/oSS258

(59 °C, 11 sec), P_{veg} from 168CA chDNA with oSS259/oSS260 (60 °C, 11 sec), *lacZ* from pMUTIN4 with oSS261/oSS262 (54 °C, 50 sec). The sequence for P_{spac} was short enough to allow for introduction by overhangs in the reverse and forward primers (oSS260/oSS261) of two adjacent fragments that were to be joined by means of SOE. The three fragments were joined stepwise using oSS346 as the terminal *sfGFP* primer, and oSS347 as the terminal *lacZ* primer, which also introduce *Bam*HI sites in at both ends of the final construct.

Components of the reporter construct were stepwise subcloned into the pUC19 multiple cloning site (MCS), consisting of (in order) *Eco*RI, *Sac*I, *Bam*HI, *Pst*I, *Sph*I, *Hind*III. The first convergent terminator, t_{ezrA}/t_{braB} was PCR amplified from 168CA chDNA using primer pair oSS344/oSS345 (55 °C, 5 sec) with overhangs carrying the *Sac*I recognition sequence, inserting it into the pUC19 *Sac*I site produced pSS116. The *Bam*HI site of pSS116 was used to introduce the P_{veg}-*sfGFP*/P_{spac}-*lacZ* construct, producing pSS117. A divergent terminator on the opposite side of the construct was introduced using the *Sph*I site of pSS117, inserting t_{ywoG}/t_{ywoF} PCR amplified from 168CA chDNA using primer pair oSS348/oSS349 (55 °C, 5 sec) carrying the *Sph*I recognition sequence, creating pSS118.

*Eco*RI and *Hind*III were used to release the transcriptionally isolated assembled construct from pSS118, creating a 4.7 kbp fragment, and the pMarB-derived vector pSS121 was opened using *Mun*I. Both fragments were blunted using Klenow fragment (Promega) and ligated to create pSS125.

2.2.3 Construction of P_{rpsD} reporter constructs

The P_{rpsD}-*gfp-ssrA* reporter plasmid pSS102 for integration at the *amyE* locus was created using the CloneJet system (Thermo Scientific). 550 bp PCR products of the 3′ and 5′ ends of *amyE* were amplified using primer pairs oSS243/oSS244 and oSS251/oSS252, respectively (57 °C, 15 sec). A spectinomycin resistance cassette was amplified from pSG1154 using oSS245/oSS246 (55 °C, 30 sec), a 333 bp region of P_{rpsD} extending until the start codon of *rpsD*, from 168CA chDNA using oSS247/oSS248. *sfGFP* was amplified from pSG1729-sfGFP using oSS249/oSS250 (57 °C, 15 sec), with oSS250 providing part of the *ssrA* in an non-complementary overhang sequence. Using SOEing PCR, the fragments were joined in the following order: *amyE* 5′, *spcR*, P_{rpsD}, *sfGFP-ssrA*, *amyE* 3′, and joined using the terminal primer pair oSS243/oSS252 (58 °C, 1 min 30 sec). The resulting 3.3 kbp fragment was cloned into PJET1.2Blunt using the protocol and T4 DNA ligase provided by the supplier, then transformed into DH5α *E. coli*.

A P_{rpsD}-*gfp* reporter construct (lacking *ssrA*) was created from pSS102 by linearising the plasmid with primer pair oSS190/oSS429 by PCR (57 °C, 3 min 30 sec).

The linearised plasmid was treated with *Dpn*I, PNK and ligase as described in section 2.2.1 to create pSS164.

2.2.4 Construction of P_{hag} reporter constructs

A P_{hag}-*mCherry* reporter construct was made, based on the P_{hag}-*gfp* construct by Kearns and Losick (2005). A 377 bp P_{hag} fragment was amplified from 168CA chDNA using primers oSS411/oSS370 (55 °C, 15 sec), and a 749 bp *mCherry* fragment from DS1675 chDNA using primers oSS412/oSS413 (55 °C, 15 sec). The promoter and reporter were joined by SOEing PCR using 1 μl of each purified fragment as template, and oSS370/oSS413 as primers (55 °C, 25 sec). The resulting 1103 bp P_{hag}-*mCherry* fragment was cut using *Eco*RI and ligated into pAPNC213Cat linearised using the same enzyme and treated with calf intestine alkaline phosphatate (Roche). The resulting plasmid (pSS153) was screened for variants where the inserted P_{hag} did not promote transcription in the same direction as the P_{spac} promoter present upstream of the *Eco*RI, as this appeared to be artificially increase mCherry signal even in the absence of IPTG.

A tetracycline-resistant version—pSS157, was created by digestion of pSS153 with *Bgl*II and *Stu*I to excise the chloramphenicol resistance cassette *cat*. The remaining vector was ligated to the cut PCR product amplified from a strain using the *tetL* cassette to delete a gene (*ezrA::tetL*), by using primer pair oSS416/oSS417 (55 °C, 30 sec).

2.2.5 Construction of intermediately stable ssrA variants

Plasmids carrying a *gfp-ssrA* construct were mutagenised using PCR and linearising primer pairs oSS424/oSS425 or oSS424/oSS426 (55 °C, 30 sec/kbp) to create mutations "DAV" and "AAV", respectively (names signify the three most C-terminal residues, wild-type being "LAA"). The PCR product was treated with *Dpn*I and cleaned up to remove template product. The linearised plasmid was kinated using T4 PNK, then ligated using T4 DNA Ligase (Roche), sequentially and in the same reaction before being transformed into DH5α *E. coli*.

The above method applies to the construction of pSS160 and pSS161 from pSS102, as well as pSS167 from pSS114 .

2.2.6 Site-directed mutagenesis with QuikChange

Some plasmids, (pSS66 to pSS68) were constructed using the QuikChange site-directed mutagenesis method, which is largely equivalent to the method described in 2.2.5. Mutagenic primer pairs were designed using the "QuikChange Primer Design tool" provided online (Agilent Technologies, Inc., 2013). The pairs oSS169/oSS170, oSS171/oSS172, oSS173/oSS174, and oSS175/oSS176 were used to mutagenise

pSS65-pSS68 (in the same order), to carry SsrA tags with the three terminal residues as follows: "DAA", "AAA","DAV", and "AAV", respectively. The PCR reaction was mixed as normally specified for PFU Turbo Polymerase (Agilent), with template plasmid (pSS59) concentration at ∼0.2 ng/μl. Cycler conditions: initial denaturation at 93 °C for 1 min, 18 cycles of 95 °C for 50 sec, 52-58 °C for 50 sec, 68 °C for 9 min, then final extension at 72 °C for 10 min.

2.2.7 Construction of inducible gfp constructs

The *gfp-ssrA* vector plasmid pSS52 was created by cutting pSG1164 with *Spe*I, blunting the ends with Klenow fragment, then cutting again with *Bsr*GI (which cuts in the middle of *gfp*). An insert was created by amplifying *gfp* from pSG1164 using primer pair oSS63/oSS106 (55 °C, 20 sec), carrying an *ssrA* overhang. The insert was cut with *Bsr*GI and ligated blunt/sticky ends into the vector.

An IPTG-inducible $P_{hyper\text{-}spank}$-*gfp-ssrA* construct was created with vector plasmid pDR111 using *Hind*III and *Sph*I. A *gfp-ssrA* insert was amplified from pSS52 using primer pair oSS159/oSS158 (60 °C, 15 sec), and a *gfp* insert was amplified from pSG1164 (oSS159/160, 60 °C, 15 sec). *gfp-ssrA* and *gfp* were ligated into pDR111, creating pSS59 and pSS60, respectively.

2.2.8 Construction of P_{rpsD} lacking a negative feedback mechanism

A P_{rpsD}-*gfp-ssrA* construct (termed Δf) lacking a secondary structure responsible for allosteric feedback in its transcript between positions +21 to +156 was created by linearising pSS102 by PCR, using primers oSS332/oSS333 (55 °C, 3 min) that exclude this region. The resulting linearised plasmid was cut with *Bam*HI and ligated to itself without an insert, resulting in pSS114.

2.2.9 Relocation of reporter constructs from amyE to aprE

To create pSS64, an *aprE*-integrating version of the P_{rpsD}-*gfp-ssrA* plasmid pSS1, the reporter construct was amplified using primers oSS131/PrpsD-F (60 °C, 40 sec), and cut with *Sph*I, ligated into pAPNC213Cat opened with *Sph*I and *Sma*I.

To reverse the whole integrating region of the plasmids pSS1 and pSS64, the primer pairs oSS189/oSS190 (60 °C, 2 min 30 sec) and oSS193/oSS194 (62 °C, 2 min) were used to amplify the *amyE* and *aprE* arms and non-integrating regions of the plasmids (plasmid backbones). The integrating region (reporter and antibiotic resistance cassettes) of pSS1 and pSS64 were amplified with primers oSS191/oSS192 (53 °C, 1 min) and oSS195/oSS196 (55 °C, 1 min 45 sec). Using homologous overhangs introduced by the primer pairs used to amplify the integrating regions, backbone and integrating

region fragments were joined using an In-Fusion HD PCR cloning kit (Clontech) according to supplier's instructions.

2.3 Microscopy

2.3.1 Continuous culture microscopy

Overnight cultures for microscopy were grown at 37 °C in CH supplemented with 0.5% glucose and an appropriate antibiotic. Cultures of 1:20 diluted overnights in fresh CH and 0.025% glucose were grown for ∼2 h then further diluted to OD_{600} = 0.05 in pre-warmed media. Samples were washed by pelleting and resuspending in PBS, and 0.3 μl was spotted on 1% agarose slides prepared using 125 μl GeneFrames (AbGene, Surrey, UK). If staining was required, 0.2 μg/ml Nile Red (9-diethylamino-5-benzo-[α] -phenoxazinone) and/or 2 μg/ml DAPI (4',6-diamidino-2-phenylindole) was supplemented in the agarose when making the slide. Unless otherwise specified, microscopy was carried out using a Zeiss M200 microscope with a CoolSNAP HQ^2 CCD camera (Photometrics), controlled by MetaMorph software (Molecular Devices). Excitation/emission wavelengths and exposure times of fluorophores as follows: GFP (470/525) for 500 ms, mCherry or Nile red (560/630) for 500 ms and 250 ms (respectively), DAPI (350/460) for 500 ms.

2.3.2 Time lapse microscopy

Slides for time lapse microscopy were created using 1% 2-hydroxyethyl agarose (low gelling temperature agarose/type VII, Sigma) in CH media, moulded in a 125 μl GeneFrame (AbGene, Surrey, UK). Strips of agarose were cut out of the moulded slab to allow for aeration and physical separation of samples spotted on the same slide. A growing culture with a low OD_{600} (2-3 h growth at 37 °C after 1:100 dilution of an overnight culture) was diluted a further ∼200 times in CH before applied to a microscopy slide. 2 μl of sample was spotted on an agarose strip, and excess liquid was allowed to evaporate before applying a cover slip. This procedure is demonstrated in further detail in video material accompanying the protocol by de Jong *et al.* (2011a).

Microscopy was performed at 30 °C using a DeltaVision suite (Applied Precision) consisting of an Olympus IX71 microscope, a CoolSNAP HQ CCD camera (Photometrics), and SoftWoRx imaging software. The following filter settings were used (excitation/emission): FITC for detecting GFP (490/526), 1,000 ms, RD-TR-Cy3 for detecting mCherry (555/617), 1,000 ms, both were filtered using an ND2 (50% transmission) neutral density filter.

2.4 Microscopy analysis

Several methods of analysing microscopy images are used in this thesis:

1. Manual tracing of cells, where a 1 pixel thick line is used to bisect a cell along its z-axis, and a mean fluorescence intensity value is measured. This method was used to analyse time-lapse microscopy movies in chapter 5.

2. Automated analysis by CellProfiler, using nucleoid staining to identify cells (see below in section 2.4.1).

3. Semi-automated detection of cells within chains by ChainTracer (ObjectJ/ImageJ), using a membrane stain (described in chapter 3)

4. Automated analysis by NucTracer (ObjectJ/ImageJ)—a method based on method 2, but rewritten for wider distribution to the research community (described in chapter 3)

All methods are carried out on 16-bit `.tiff` files, and measurements are blanked against the average background value of the image.

2.4.1 Software-based analysis pipeline

For experiments investigating cell chaining and using membrane stains, ObjectJ/ImageJ was used to process microscopy data (Vischer *et al.*, 1994). Data presented in chapter 3 was analysed using this method.

Where both GFP and mCherry were used in the same strain (data in chapter 4), nucleoids were stained with DAPI and identified using thresholding methods in the cell image analysis software CellProfiler (Carpenter *et al.*, 2006). After identifying nucleoids, CellProfiler uses their outlines as ROIs to capture fluorescence intensities of specified channels, which are subsequently outputted into a `.csv` file along with metadata (strain ID, image ID, time point, or any other data that can be identified from the file name). A final step overlays the outlines over the DAPI image and saves the file for quality control. The advantages of CellProfiler over ObjectJ are in its ability to run processing jobs as "batch"—that is to say, perform a group of tasks with only initial inputs from the user and no need for interruptions in order to await user input. The CellProfiler pipeline is described in figure 2.1 and can be found in the appendix (A.1).

As a consequence of human error in capturing microscopy images, the fine focus of DAPI channel microscopy images has a significant impact on ROI quality, and is remedied here by adjusting the *threshold correction factor* (TCF) in the CellProfiler pipeline. No quality checking takes place *during* the image processing step in CellProfiler, and this problem was solved by executing three parallel runs of the same

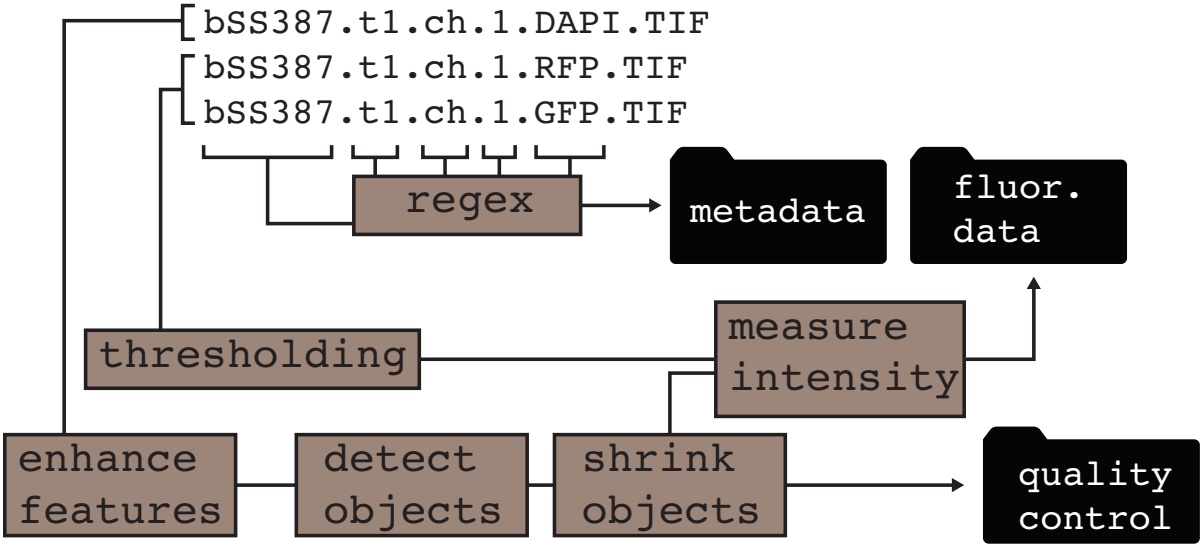

Figure 2.1: Outline of CellProfiler pipeline file. Diagrammatic representation of the CellProfiler pipeline used (referred to as pipeline.cp in figure 2.2) Brown boxes represent processing of data. Metadata such as strain name, time point and media are identified by a regular expression (regex) that extracts data in file names (in this example, strain name, time point, media, distinguishing number, and channel) according to a predictable pattern. The channels with visible reporter signal (GFP and RFP) are thresholded to subtract background fluorescence. The DAPI image is enhanced to bring out intensity peaks, which are detected in the next step. Objects are shrunk to ensure that mis-alignment between the GFP, RFP and DAPI channels do not become a source of error. Finally the outlines are used to capture intensity data from the thresholded images. A quality control image representing the outlines used as ROIs is also created.

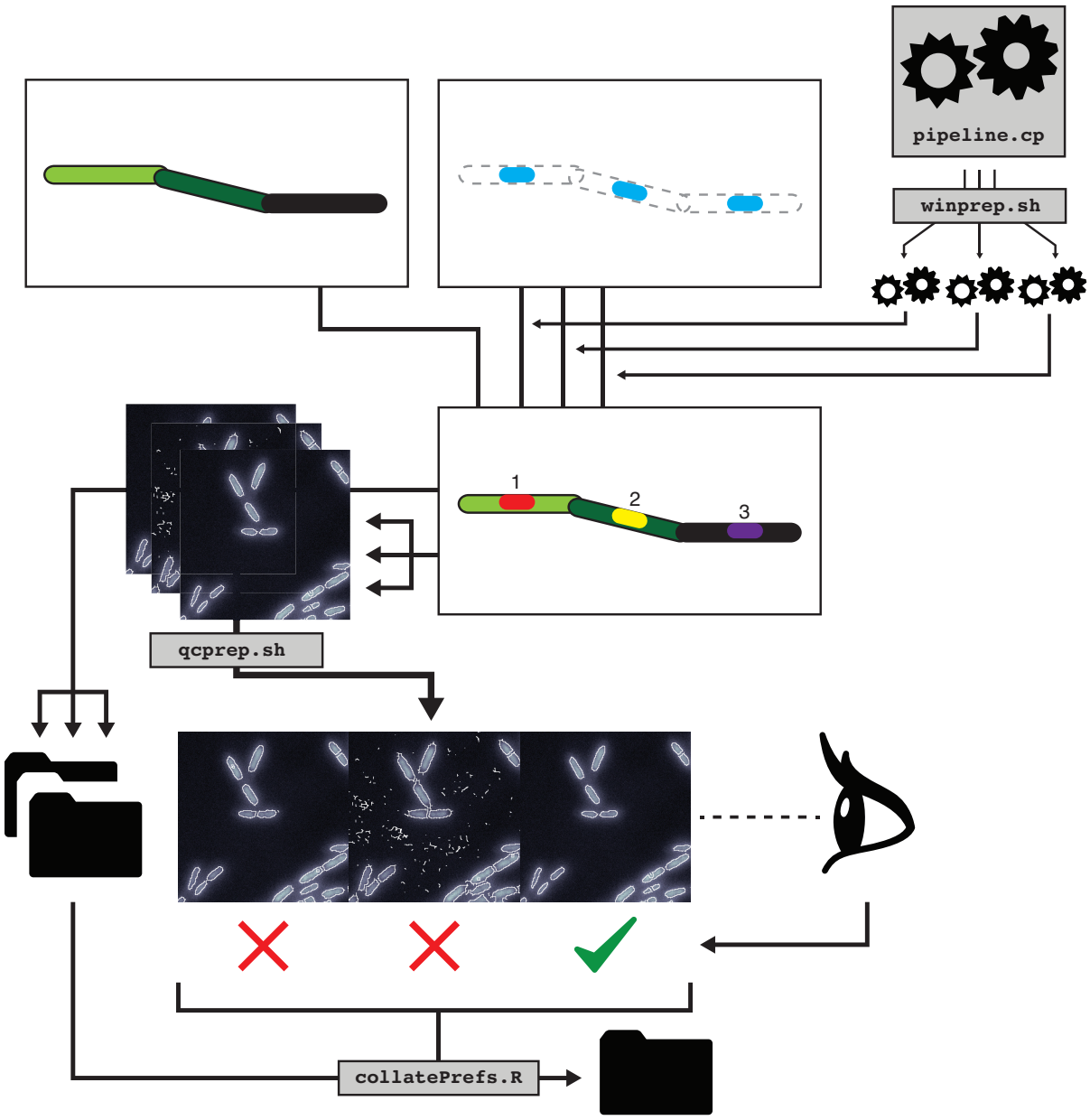

Figure 2.2: Outline of procedure used for detection of cell ROIs by nucleoid stains. The setup script winprep.sh duplicates the CellProfiler pipeline (see figure 2.1 for details) as copies with three different threshold correction factors. The pipeline takes input DAPI and reporter fluorescence images and quantifies fluorescence which is output as three files (one per TCF), as well as three quality control images. The setup script qcprep.sh makes montages of the three quality control images, which are then manually inspected, and preferences are recorded. The R script collatePrefs.R collects the results of preferred runs per image set into a final file containing fluorescence data.

analysis with only the TCF values varying between them. A UNIX shell script was used to set up this processing pipeline, which also enables the user to set up the experiment on another computer running Windows, as personal workstations can be unsuitable for image analysis, and a purpose-built high-end system is often available at many laboratories. Following execution of the three pipelines, a second shell script merges output images from CellProfiler into montages that enable the user to easily inspect the results and specify which run was preferable on a per-image basis. A third R script is used to take the user preferences and create a file with results taken from the three output `preferredTCF.csv` files, image by image and merging the results into a file representing the cell fluorescence intensities based on user quality control of DAPI regions. This work flow is illustrated in figure 2.2, and the auxiliary setup scripts can be found in Data analysis scripts (section A.1.1).

2.5 Western blot for detection of ClpX and ClpP

Samples for western blotting were pelleted and frozen at $OD_{600} \approx 0.5$ from cultures of *B. subtilis* growing in CH at 37 °C. Samples were resuspended in a buffer containing 2 mM EDTA (Ethylenediaminetetraacetic acid) and 100 mM Tris (tris(hydroxymethyl) aminomethane) and protease inhibitor (Roche). Cells were broken up using lysozyme (10 μg/ml, 10 min) and sonicated at 40 Hz. Cell lysates were pelleted using a tabletop centrifuge for 2 min and the protein concentration of supernatants were equalised by OD_{600} values recorded at the time of harvesting. Samples were prepared for loading onto a NuPAGE gel running in MOPS NuPAGE buffer at 200 V/300 mA using NuPAGE LDS sample buffer and sample reducing agent (Life technologies). Proteins were transferred overnight at 30 mA onto a Hybond nitrocellulose membrane (Amersham) using a wet transfer technique in the following buffer: 12.5 mM Tris, 96 mM Glycine, 0.05% SDS (Sodium dodecyl sulfate), 20% methanol.

Prior to blotting, the membrane was stained with Ponceau S Red stain (Sigma) to ensure even loading, then rinsed in DI water to destain. Blocking and blotting was performed with 5% (w/v) powdered skimmed milk (Marvel, Chivers Ireland, Ltd., Dublin) in PBST (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 2 mM KH_2PO_4 (pH 7.4), 0.1% Tween 20 (Polyoxyethylene (20) sorbitan monolaurate)) for 1 hour. Primary α-ClpX and α-ClpP antibodies (rabbit) were diluted 1:7,000 and 1:20,000 (respectively) and blotted for 1 hour. Secondary α-rabbit goat antibody-HRP conjugate was used at a 1:10,000 dilution for both primary antibodies, blotted for 1 hour. Chemiluminescence was performed using Pierce ECL 2 Western Blotting Substrate (Thermo), and detected using an ImageQuant LAS 4000 mini system (GE Healthcare). Densitometry was performed using gel analysis functionality of ImageJ (National Institutes of Health, USA).

2.6 β-Galactosidase assay

β-Galactosidase activity assays were carried out according to protocols set out by Miller (1972). Culture samples were collected at the same time microscopy was done and snap frozen in liquid nitrogen. Cells were resuspended in working buffer, diluted 1:8 and their OD_{600} was recorded. Lysis was done by adding 20 mg/ml lysozyme and incubating at room temperature for 30 min. ONPG was added at a 1:8 volume ratio from 20 mM stock in Z buffer, then reactions were allowed to proceed until a colour change had been observed. Reactions were stopped by adding 1 M Na_2CO_3 in Z-buffer at a ratio of 1:2.25, then OD_{450} and A_{420} was recorded. Promoter activity in Miller units was calculated using the below equation, using A (absorbance at specified wavelength), OD (optical density at specified wavelength), t (reaction time), v (reaction volume):

$$\frac{1,000 \times (A_{420} - (1.75 \times OD_{550}))}{t \times v \times OD_{600}}$$

Z-buffer composition: 60 mM $Na_2HPO_4{\cdot}2\,H_2O$, 40mM $NaH_2PO_4{\cdot}H_2O$, 10 mM KCl, 1 mM $MgSO_4$, adjusted to pH 7.0. Working buffer was made from 50 mM β-Mercaptoethanol in Z-buffer.

2.7 *In vivo* GFP half-life assay

The half-life of GFP-SsrA variants was calculated from the half-life of the GFP signal as observed *in vivo* using epifluorescence microscopy. Strains were grown as described in section 2.3.1, but upon reaching $OD_{600} \approx 0.5$ split into two 200 μl aliquots in 2 ml Eppendorf tubes and placed on a shaking 37 °C heat block. At 0 min, samples were treated with 250 μg/ml chloramphenicol from an aliquot of 100 mg/ml, and as a control the same volume (0.5 μl) of the solvent for the chloramphenicol (100% ethanol) was added. A sample was spotted on a slide prior to addition of chloramphenicol and used to represent the original GFP intensity levels. Microscopy images were captured taken using an epifluorescence microscope at regular intervals until a sufficient drop in GFP signal had been observed.

The nuclei or cell membranes of samples were not stained due to time and volume constraints, therefore cells for each time point, condition, and strain were measured manually using ImageJ and processed in R (method 1 in section 2.4).

Analysis was done in R (R Development Core Team, 2008) using the script shown in the appendix (A.5), which takes data from a comma-separated values (`preferredTCF.csv`) file with the following data per cell: sequential cell ID within an image, the first always being a background measurement, mean GFP value, parent image ID, treatment (sequential), time point (sequential), strain (sequential), as well as

a list of treatments, time points and strains to correspond to the sequential IDs. The script plots the data, then takes user input on lines 68–69 to determine the range of time points to use when fitting a straight line to a portion of the data leading up to a curve levelling off. Using the straight line equations (2.1) of the chloramphenicol treated (*cam*) time series and the control (*c*) series, the slope is adjusted for background decrease in fluorescence by subtracting subtracting the slope of the control (equation 2.2), represented on lines 77-87.

$$y = mx + c \tag{2.1}$$

$$y_{adj} = (m_{cam} - m_c)x + c_{cam} \tag{2.2}$$

2.8 Transposon screen for genes affecting the $P_{rpsD}{}^{HIGH}$ phenotype

2.8.1 Mariner transposon mutagenesis using pMarB

The host *B. subtilis* strain was transformed with pMarB using a normal transformation protocol (see section 2.1) and incubated for 2 nights at 30 °C (permissive temperature for pMarB replication) on an erythromycin nutrient agar plate. Candidate colonies are grown in LB for ∼3–5 hours at 30 °C, spread on both kanamycin and erythromycin nutrient agar plates at dilutions of 10^{-3} to 10^{-5}, and incubated overnight at 50 °C (non-permissive temperature for replication). The culture yielding the highest kan:erm cfu/ml count (typically ∼30:1) was chosen for further experiments, and if more colonies were required, frozen stock saved from the previous step could be plated out.

2.8.2 Creation of transposant pool chDNA

Using the calculated kanamycin-resistant cfu/ml, a volume of frozen stock equalling a number of transposants (∼40,000) likely to yield insertions that cover the chromosome by linkage, was spread on enough large (150 mm diameter, Sarstedt) petri dishes required for 2,500 colonies/plate. Colonies were suspended into a small volume of PBS by mechanical scraping using a sterile Pasteur pipette bent into an "L" shape using a Bunsen burner. chDNA was extracted from the total colony suspension by diluting sufficient volume to yield a suitable amount for chDNA extraction (2 ml sample at $OD_{600} = 2.0$, see section 2.2 for details).

2.8.3 Microscopy-based screen

A receptor strain was transformed with transposant pool chDNA, selecting for kanamycin resistance and also screening for erm^S colonies. The resistance of any relevant

reporter construct not related to the transposon mutagenesis process was also selected for. Colonies were grown overnight in CH and diluted 1:1,000 in 800 μl fresh CH medium in 2 ml Eppendorf tubes attached to a shaking rack. After 6–7 hours, each culture was imaged using a microscope (see section 2.3 for general microscopy methodology).

2.8.4 Microplate-based screen

Colonies picked for a microplate-based screen for loss of a $P_{rpsD}{}^{\mathrm{HIGH}}$ phenotype were grown overnight in CH at 37 °C. Overnight cultures were diluted 1:100 into fresh media and allowed to grow for 5–6 hours, then diluted 1:40 times into a volume of 200 μl in 96-well microtitre plates (BD plastic) and covered with a clear lid. Wells on the edge of the microtitre plate were left with uninoculated media to prevent excess evaporation of liquid. Measurements were carried out on a BMG Labtech FLUOstar optima plate reader, using a 595 nm filter to record OD_{600}, and 485/520 nm (excitation/emission) filters to record GFP.

2.9 Arbitrary PCR to locate Tn*YLB-1* from pMarB

Determination of Tn*YLB-1* transposon insertion loci was done by arbitrary PCR, adapted from Knobloch *et al.* (2003). PCR reactions were carried out using GoTaq polymerase (Promega) in 50 μl reactions using primer pairs Arb1/MarB1 or Arb1/MarB2 (sequencing from the 5′ and 3′ ends of *kan*, respectively). Thermocycler program as follows: initial denaturation at 95 °C for 5 min, 6 cycles of 95 °C for 1 min, 30 °C for 1 min, 72 °C for 1 min, 30 cycles of 95 °C for 30 sec, 50 °C for 30 sec, 72 °C for 1 min, then final extension at 72 °C for 7 min.

A second PCR reaction using 2 μl of the previous finished reaction as template was carried out using primer pairs Arb2/MarB1 or Arb2/MarB2 (to match the corresponding primer pair in the previous reaction). Thermocycler program as follows: initial denaturation at 95 °C for 5 min, 6 cycles of 95 °C for 30 sec, 50 °C for 30 sec, 72 °C for 30 sec, 30 cycles of 95 °C for 30 sec, 58 °C for 30 sec, 72 °C for 30 sec, then final extension at 72 °C for 7 min.

The reaction was cleaned up using a PCR cleanup kit, then Sanger sequenced using the primers MarB1N or MarB2N. Sequencing output was compared to the genome sequence of *B. subtilis* 168CA available on *SubtiList* (Moszer *et al.*, 1995) using BLAST to determine the Tn*YLB-1* insertion locus.

Table 2.1: List of strains used. Where a strain originated from this study, its construction is described by recipient strain listed to the left, donor strain (or plasmid) on the right. Unless stated otherwise, all strains belong to the *B. subtilis* species.

Strain	Relevant genotype	Construction or Reference
168CA	*trpC2* wild-type *B. subtilis*	Laboratory strain
PY79	prototroph wild-type *B. subtilis*	Youngman *et al.* (1984)
AH7	P_{hag}-*sfGFP*	L. Hamoen, unpub.
LH95	Δ*clpX*	L. Hamoen, unpub.
BSB1	168CA *trpC2*$^+$	Nicolas *et al.* (2012)
BG546	BG1 *trpC2 thr5 pnpA*::*kan*	Wang and Bechhofer (1996)
CB100	*sigD*::*cat*	Mirel and Chamberlin (1989)
DS1675	*thrC*::($P_{fla/che}$-*mCherry spc*)	D. Kearns
DS901	PY79 *amyE*::(P_{hag}-*gfp cat*)	Kearns and Losick (2005)
GP736	*sinR*::*tet*	Bai *et al.* (1993)
JK47	*clpX*::*cat*	Kirstein *et al.* (2008)
JWV026	*spo0A*::*kan*	J.-W. Veening, unpub.
TW20	*relA*::*mls amyE*::(P_{spac}-*relA cat*)	Wendrich and Marahiel (1997)
BD2590	*clpP*::*mls*	Turgay *et al.* (1998)
SG82	*lacA*::*tet*	Daniel *et al.* (1997)
ezrA::*tetL*	*ezrA*::*tetL*	L. Hamoen, unpub.
PS258	*codY*::*erm*	Serror and Sonenshein (1996)
PS259	*codY*::*cat*	Serror and Sonenshein (1996)
DH5α	*E. coli* F− ϕ80*lacZ*ΔM15 Δ(*lacZYA-argF*) U169 *recA*1 *endA*1 *hsdR*17 (rK−, mK+) *phoA supE*44 λ− *thi*-1 *gyrA*96 *relA*1	Commercial (Invitrogen)
ER2925	*E. coli ara*-14 *leuB*6 *fhuA*31 *lacY*1 *tsx*78 *glnV*44 *galK*2 *galT*22 *mcrA dcm*-6 *hisG*4 *rfbD*1 R(*zgb*210::Tn*10*)TetS *endA*1 *rpsL*136 *dam*13::Tn*9 xylA*-5 *mtl*-1 *thi*-1 *mcrB*1 *hsdR*2	Commercial (NEB)
bSS2	PY79 *amyE*::(P_{rpsD}-*gfp-ssrA spc*)	Syvertsson, MRes project
bSS107	168CA *amyE*::(P_{rpsD}-*gfp-ssrA spc*)	168CA × bSS2
bSS135	168CA *amyE*::($P_{hyper\text{-}spank}$-*gfp-ssrA spc*)	168CA × pSS59
bSS136	168CA *amyE*::($P_{hyper\text{-}spank}$-*gfp spc*)	168CA × pSS60
bSS161	168CA *amyE*::($P_{hyper\text{-}spank}$-*gfp-ssrA*$_{DAV}$ *spc*)	168CA × pSS67
bSS162	168CA *amyE*::($P_{hyper\text{-}spank}$-*gfp-ssrA*$_{AAV}$ *spc*)	168CA × pSS68
bSS166	PY79 *aprE*::(P_{rpsD}-*gfp-ssrA spc*)	PY79 × pSS64
bSS174	PY79 *amyE*::(P_{rpsD}-*gfp-ssrA spc*) (reversed)	PY79 × pSS72
bSS184	PY79 *aprE*::(P_{rpsD}-*gfp-ssrA spc*) (reversed)	PY79 × pSS73
bSS250	168CA *amyE*::($P_{hyper\text{-}spank}$-*gfp spc*) *codY*::*cat*	bSS136 × PS259
bSS251	168CA *amyE*::($P_{hyper\text{-}spank}$-*gfp-ssrA*$_{DAV}$ *spc*) *codY*::*cat*	bSS161 × PS259
bSS277	168CA *amyE*::(P_{rpsD}-*sfGFP-ssrA spc*)	168CA × pSS102
bSS283	168CA *amyE*::(P_{rpsD}-*sfGFP-ssrA spc*) *codY::cat*	bSS277 × PS259
bSS322	168CA *amyE*::(P_{rpsD}Δ*f*-*sfGFP-ssrA spc*)	168CA × pSS114
bSS332	168CA *relA*::*mls*	168CA × TW20
bSS339	168CA *amyE*::(P_{hag}-*gfp cat*)	168CA × DS901
bSS340	168CA *amyE*::(P_{rpsD}-*sfGFP-ssrA spc*) *pnpA*::*kan*	bSS277 × BG546
bSS352	168CA *trpC2*$^+$ *lacA*::*tet*	BSB1 × SG82
bSS353	168CA *trpC2*$^+$ *lacA*::*tet aprE*::(*lacI cat*)	bSS352 × pAPNC213Cat
bSS385	168CA *aprE*::(P_{hag}-*mCherry cat*)	168CA × pSS153
bSS387	168CA *amyE*::(P_{rpsD}-*sfGFP-ssrA spc*) *aprE*::(P_{hag}-*mCherry* cat)	bSS277 × bSS385

Continued on next page

Table 2.1 – *Continued from previous page*

Strain	Relevant genotype	Construction or Reference
bSS394	168CA *amyE*::(P_{rpsD}-*sfGFP-ssrA spc*) *aprE*::(P_{hag}-*mCherry* cat) *pnpA*::*kan*	bSS387 × BG546
bSS395	168CA *amyE*::(P_{rpsD}-*sfGFP-ssrA spc*) *aprE*::(P_{hag}-*mCherry* cat) *codY*::*mls*	bSS387 × PS258
bSS396	168CA *amyE*::(P_{rpsD}-*sfGFP-ssrA spc*) *aprE*::(P_{hag}-*mCherry* cat) *relA*::*mls*	bSS387 × bSS332
bSS397	168CA *amyE*::(P_{rpsD}-*sfGFP-ssrA spc*) *aprE*::(P_{hag}-*mCherry* cat) *spo0A*::*kan*	bSS387 × JWV026
bSS399	168CA *amyE*::($P_{rpsD}\Delta f$-*sfGFP-ssrA spc*) *aprE*::(P_{hag}-*mCherry cat*)	bSS322 × bSS385
bSS402	168CA *amyE*::(P_{rpsD}-*sfGFP-ssrA* spc) *aprE*::(P_{hag}-*mCherry tet*)	bSS387 × pSS157
bSS407	168CA *amyE*::(P_{rpsD}-*sfGFP-ssrA* spc) *aprE*::(P_{hag}-*mCherry tet*) *sigD*::*cat*	bSS402 × CB100
bSS410	168CA *amyE*::(P_{rpsD}-*sfGFP-ssrA*$_{AAV}$ *spc*)	168CA × pSS161
bSS421	168CA *amyE*::(P_{rpsD}-*sfGFP spc*)	168CA × pSS164
bSS429	168CA *amyE*::(P_{rpsD}-*sfGFP spc*) *aprE*)::(P_{hag}-*mCherry cat*)	bSS421 × bSS385
bSS432	168CA *aprE*::(P_{hag}-*mCherry cat*) *amyE*::(P_{rpsD}-*sfGFP-ssrA*$_{AAV}$ *spc*)	bSS385 × bSS410
bSS440	168CA *amyE*::($P_{rpsD}\Delta f$-*sfGFP-ssrA*$_{\mathrm{DAV}}$ *spc*)	168CA × pSS167
bSS475	168CA *amyE*::(P_{rpsD}-*sfGFP-ssrA spc*) *aprE*::(P_{hag}-*mCherry* cat) *sinR*::*tet*	bSS387 × GP736
bSS477	PY79 *amyE*::($P_{rpsD}\Delta f$-*sfGFP-ssrA*$_{\mathrm{DAV}}$ *spc*)	PY79 × bSS440
bSS479	168CA *amyE*::($P_{hyper\text{-}spank}$-*gfp-ssrA*$_{DAV}$ *spc*) *pnpA*::*kan*	bSS161 × BG546
bSS483	168CA *amyE*::($P_{hyper\text{-}spank}$-*gfp spc*) *pnpA*::*kan*	bSS136 × BG546
bSS485	168CA *amyE*::($P_{hyper\text{-}spank}$-*gfp-ssrA spc*) *clpX*::*cat*	bSS135 × JK47
bSS492	168CA *amyE*::(P_{rpsD}-*sfGFP-ssrA*$_{DAV}$ *spc*) *aprE*)::(P_{hag}-*mCherry cat*)	bSS385 × bSS411

Table 2.2: List of plasmids used in this study. If its construction is described in this document, a page reference is provided.

Plasmid	Relevant genotype	Reference
pAPNC213Cat	*aprE*::(MCS *lacI cat*) *bla*	Morimoto *et al.* (2002)
pJet1.2/blunt	*bla rep* p_{lacUV5}-*eco47IR'* (blunt) *'eco47IR*	GenBank: EF694056
pMarB	*erm bla himar9* Tn*YLB-1*::(*kan*)	Le Breton *et al.* (2006)
pSG1164	P_{xyl}-*gfpmut1 bla cat*	Lewis and Marston (1999)
pDR111	*amyE*::($P_{hyper\text{-}spank}$ *lacI spc*) *bla*	Britton *et al.* (2002)
pMUTIN4	P_{spac} MCS *lacZ lacI bla erm*	Vagner *et al.* (1998)
pSG1729-sfGFP	*amyE*::(*sfGFP spc*) *bla*	F. Bürmann, unpub.
pSG1154	*amyE*::(P_{xyl}-MCS-*gfp spc*) *bla*	Feucht and Lewis (2001)
pSS1	*amyE*::(P_{rpsD}-*gfp-ssrA spc*) *bla*	Syvertsson, MRes project
pSS52	*amyE*::(*gfp-ssrA spc*) *bla*	page 42
pSS59	*amyE*::($P_{hyper\text{-}spank}$-*gfp-ssrA spc*) *bla*	page 42
pSS60	*amyE*::($P_{hyper\text{-}spank}$-*gfp spc*) *bla*	page 42
pSS64	*aprE*::(P_{rpsD}-*gfp-ssrA spc*) *bla*	page 42
pSS65	*amyE*::($P_{hyper\text{-}spank}$-*gfp-ssrA*$_{DAA}$ *spc*) *bla*	page 41
pSS66	*amyE*::($P_{hyper\text{-}spank}$-*gfp-ssrA*$_{AAA}$ *spc*) *bla*	page 41

Continued on next page

Table 2.2 – *Continued from previous page*

Plasmid	Relevant genotype	Reference
pSS67	*amyE*::($P_{hyper\text{-}spank}$-*gfp-ssrA*$_{DAV}$ *spc*) *bla*	page 41
pSS68	*amyE*::($P_{hyper\text{-}spank}$-*gfp-ssrA*$_{AAV}$ *spc*) *bla*	page 41
pSS72	*amyE*::(P_{rpsD}-*gfp-ssrA spc*) (reversed) *bla*	page 42
pSS73	*aprE*::(P_{rpsD}-*gfp-ssrA spc*) (reversed) *bla*	page 42
pSS102	*amyE*::(P_{rpsD}-*sfGFP-ssrA spc*) *bla*	page 40
pSS114	*amyE*::($P_{rpsD}\Delta f$-*sfGFP-ssrA spc*) *bla*	page 42
pSS116	t_{ezrA}/t_{braB} *bla*	page 40
pSS117	t_{ezrA}/t_{braB} P_{veg}-*sfGFP* P_{spac}-*lacZ bla*	page 40
pSS118	t_{ezrA}/t_{braB} P_{veg}-*sfGFP* P_{spac}-*lacZ* t_{ywoG}/t_{ywoF} *bla*	page 40
pSS121	*erm bla himar9* Tn*YLB-1*::(*kan Mun*I)	page 39
pSS125	*erm bla himar9* Tn*YLB-1*::(*kan* t_{ezrA}/t_{braB} P_{veg}-*sfGFP* P_{spac}-*lacZ* t_{ywoG}/t_{ywoF})	page 40
pSS153	*aprE*::(P_{hag}-*mCherry cat*) *bla*	page 41
pSS157	*aprE*::(P_{hag}-*mCherry tetL*) *bla*	page 41
pSS160	*amyE*::(P_{rpsD}-*sfGFP-ssrA*$_{DAV}$ *spc*) *bla*	page 41
pSS161	*amyE*::(P_{rpsD}-*sfGFP-ssrA*$_{AAV}$ *spc*) *bla*	page 41
pSS164	*amyE*::(P_{rpsD}-*sfGFP spc*) *bla*	page 41
pSS167	*amyE*::($P_{rpsD}\Delta f$-*sfGFP-ssrA*$_{\mathrm{DAV}}$ *spc*) *bla*	page 41

Table 2.3: List of oligonucleotides used in this study. Capitals denote a complementary portion of the sequence, underlined text indicates restriction site(s) if one is specified.

Name	Sequence (5′ to 3′)	Restriction site	Description
Arb1	ggccacgcgtcgactagtacNNNNNNNNNNGATAT		Arbitrary primer (step 1)
Arb2	GGCCACGCGTCGACTAGTAC		Nested primer for Arb1 (step 2)
MarB1	GTTAGACCGGGGACTTATC		Transposon primer (forward, step 1)
MarB1N	GCGCCTACGAGGAATTTGTATC		Nested transposon primer (forward, step 1)
MarB2	GCATTTAATACTAGCGACGCC		Transposon primer (reverse, step 2)
MarB2N	GGGAATCATTTGAAGGTTGG		Nested transposon primer (forward, step 2)
KS92	CTCTACGGAAATAGCGAGAG		*aprE* integration check (forward)
KS93	AGAAGCAGGTATGGAGGAAC		*aprE* integration check (reverse)
PrpsD-F	cgcggcatgcATGTTTTTATCACCTAAAAGTTTACC	*Sph*I	P_{rpsD}(forward)
oSS63	GAATTCCTGCAGATGAGTAAAGG		pSG1164 (forward)
oSS106	gcgcgcaagcttgcgcttaggcagctaatgctacgttttggttaaa-actgttagttttgcccgcggatctTTTGTATAGTTCATCCATGCC	*Hind*III	*gfp* (reverse), *ssrA* in overhang
oSS131	TTATTATTATTTTTGACACCAGACC		binds past *ssrA* in pSS1
oSS158	gcgccggcatgcGCGCTTAGGCAGCTAATGC	*Sph*I	*ssrA* (reverse)
oSS159	gcgcgcaagcttTAAGGAGGCCTGCAGATGAGTAAAGGAG	*Hind*III	*gfp* (forward)
oSS160	gcgccggcatgcAGTGGATCTGAAGTCTGGAC	*Sph*I	*gfp* (reverse)
oSS169	TTGCATGCGCGCTTAGGCAGCgtcTGCTACGTTTTGGTTAAAACT		QuikChange mutagenesis “DAA” on pSS59 (forward)
oSS170	AGTTTTAACCAAAACGTAGCAgacGCTGCCTAAGCGCGCATGCAA		QuikChange mutagenesis “DAA” on pSS59 (reverse)
oSS171	TTGCATGCGCGCTTAGGCAGCcgcTGCTACGTTTTGGTTAAAACT		QuikChange mutagenesis “AAA” on pSS59 (forward)
oSS172	AGTTTTAACCAAAACGTAGCAgcgGCTGCCTAAGCGCGCATGCAA		QuikChange mutagenesis “AAA” on pSS59 (reverse)
oSS173	CCACCGAATTAGCTTGCATGCGCGCTTA-aaCAGCgtcTGCTACGTTTTGGTTAAAACTGTT		QuikChange mutagenesis “DAV” on pSS59 (forward)
oSS174	AACAGTTTTAACCAAAACGTAGCAgacG-CTGttTAAGCGCGCATGCAAGCTAATTCGGTGG		QuikChange mutagenesis “DAV” on pSS59 (reverse)

Continued on next page

Table 2.3 – *Continued from previous page*

Name	Sequence (5′ to 3′)	Restriction site	Description
oSS175	CCACCGAATTAGCTTGCATGCGCGCTTA-aaCAGCggcTGCTACGTTTTGGTTAAAACTGTT		QuikChange mutagenesis "AAV" on pSS59 (forward)
oSS176	AACAGTTTTAACCAAAACGTAGCAgccG-CTGttTAAGCGCGCATGCAAGCTAATTCGGTGG		QuikChange mutagenesis "AAV" on pSS59 (reverse)
oSS183	CTGTCTGTTGTGAGCGGAG		*relA* deletion check (forward)
oSS184	GAGATTATTATGTGCTCCGG		*relA* deletion check (reverse)
oSS189	GATGGCTGGACAGCCTGAG		Plasmid backbone of pSS1 (forward)
oSS190	ATTCTGCGTGACATCCCATC		Plasmid backbone of pSS1 (reverse)
oSS191	gatgggatgtcacgcagaatCACATTGTGAAATCTATTGACC		Integrating region of pSS1, with overhang for In-fusion (forward)
oSS192	ttcctcaggctgtccagccatcTAAATAGTACATAATGGATTTCC		Integrating region of pSS1, with overhang for In-fusion (reverse)
oSS193	TTGAGTCAATTCCGCTGTCG		Plasmid backbone of pSS64 (forward)
oSS194	GGCGTTAGCCCAAGCGC		Plasmid backbone of pSS64 (reverse)
oSS195	gacagcggaattgactcAACGAATTCGAGCTCGCCC		Integrating region of pSS64, with overhang for In-fusion (forward)
oSS196	gatgcgcttgggctaacgccgCTTCACTAAATTAAAGTAATAAAG		Integrating region of pSS64, with overhang for In-fusion (reverse)
oSS243	GTCTAGCCTTGCCCTCAATG		*amyE* upstream (forward)
oSS244	gattaataattACTGACGATTACCTTGCGTG		*amyE* upstream (reverse), with SOE overhang
oSS245	gtaatcgtcagTAATTATTAATCTGTAGACAAATTGTG		*spcR* (forward)
oSS246	GATAAAAACATGGTCTTTGAAGCATGCAAATGTCAC		*spcR* (reverse)
oSS247	gcttcaaagACCATGTTTTTATCACCTAAAAG		P_{rpsD}, with overhang for SOE (forward)
oSS248	ctttgctcatctgcaggaattccATAATGTGACTCCTCCTTTG		P_{rpsD}, with overhang for SOE (reverse)
oSS249	gaattCCTGCAGATGAGCAAAGGAGAAGAACTTTTC		*sfGFP*, with SOE overhang (forward)
oSS250	ctaatgctacgttttggttaaaactgtta-gttttgcccgcTTTGTAGAGCTCATCCATGC		*sfGFP*, with *ssrA*/SOE overhang (reverse)
oSS251	gttttaaccaaaacgtagcattagctgccta-ataaGTGTATTTTGTGTATTCCAGTC		*amyE* downstream (forward), *ssrA* in overhang
oSS252	ATGTTTGCAAAACGATTCAAAAC		*amyE* downstream (reverse)

Continued on next page

Table 2.3 – *Continued from previous page*

Name	Sequence (5′ to 3′)	Restriction site	Description
oSS258	gtggatgcaATGGGTACCCTGCAGATGAG		sfGFP (reverse), with overhang for SOEing PCR
oSS259	cagggtacCCATTGCATCCACCTCACTAC		p_{veg} (forward), with overhang for SOE
oSS260	caattccacacattatgccacaccttgtagat-aaagtcaacaaCCGTAATACGCTGACAAGAGAG		p_{veg} (reverse), with P_{spac}/SOE overhang
oSS261	GTGGCATAATGTGTGGAATTGTGAGCCGCT-CACAATTATATAGGGAAAAGGTGGTGAAC		*lacZ* (forward), with P_{spac}/SOE overhang
oSS330	ggttcaattgGCCTACGAGGAATTTGTATC	*Mun*I	Introduces *Mun*I in pSS113 (forward)
oSS331	ggttcaattgGGACCCCTATCTAGCGAAC	*Mun*I	Introduces *Mun*I in pSS113 (reverse)
oSS332	gcgcggatccAAGGCTGCAATATTTCACAC	*Bam*HI	Linearises pSS102 to remove negative feedback (forward)
oSS333	gcgcggatccAATAAAACCAAAGGAGGAGTC	*Bam*HI	Linearises pSS102 to remove negative feedback (reverse)
oSS344	gcggagctctaaCTGCGCCAGGTGCAGTTG	*Sac*I	*ezrA*/*braB* terminator (forward)
oSS345	gcggagctctaaGTGCTTCCAGGAATCGTCG	*Sac*I	*ezrA*/*braB* terminator (reverse)
oSS346	gcaggaattcggatccttattaTT-TGTAGAGCTCATCCATGCCATG	*Bam*HI, *Eco*RI	*sfGFP* (forward)
oSS347	gcaggaattcggatccATTATTATTTTTGACACCAGACC	*Bam*HI, *Eco*RI	*lacZ* (reverse)
oSS348	ggatcgcatgctaaTTAACAGCGGGGCTGTTCG	*Sph*I	*ywoG*/*ywoF* terminator (forward)
oSS349	ggatcgcatgctaaCCTATGATTTTTCGTTAGATCC	*Sph*I	*ywoG*/*ywoF* terminator (reverse)
oSS370	cgcggaattcTTATCGCGGAAAATAAACGAAG	*Eco*RI	P_{hag} (forward)
oSS393	GACGGTTGCTAGGTAACTG		*pnpA* deletion check (forward)
oSS394	GGGCTTAACTTCGTACTCC		*pnpA* deletion check (reverse)
oSS411	gtagttcctccttatgtgctagcGAATATGTTGTTAAGGCACGTC		P_{hag} (reverse), with SOE overhang
oSS412	GCTAGCACATAAGGAGGAACTACTATGGTCAGCAAGGGAGAGG		*mCherry* (forward)
oSS413	gcgcgaattcATCCTTATTTGTATAATTCGTCC	*Eco*RI	*mCherry* (reverse)
oSS416	gcgcaggcctTTATTGCAATAAAATTAGCCCTG	*Stu*I	*tetL* (forward)
oSS417	gcgcagatctCAACATTTAACATTTCTCAACAAG	*Bgl*II	*tetL* (reverse)
oSS422	taaaactgttagttttgcccgcTTTGTATAATTCGTCCATTCCAC		Linearises pSS153 (reverse), *ssrA* in overhang

Continued on next page

Table 2.3 – *Continued from previous page*

Name	Sequence (5′ to 3′)	Restriction site	Description
oSS423	accaaaacgtagcattagctgccTAAGGATGAATTCGAGCTCGC		Linearises pSS153 (forward), *ssrA* in overhang
oSS424	TGCTACGTTTTGGTTAAAACTG		*ssrA* mutagenesis (reverse)
oSS425	gacGCTGttTAATAAGTGTATTTTGTGTATTCC		*gfp-ssrA* mutagenesis "DAV" (forward)
oSS426	gccGCTGttTAATAAGTGTATTTTGTGTATTCC		*gfp-ssrA* mutagenesis "AAV" (forward)
oSS429	tcattaTTTGTAGAGCTCATCCATGCC		*gfp* (reverse), 2 stop codons in overhang

Chapter 3

Semi-automated analysis of single cells defined by membrane or DNA staining, using ImageJ with the plugin ObjectJ

3.1 Introduction

Isogenic populations of bacteria show a remarkable variability in behaviour, especially when growth conditions become challenging. For example, cells can become motile, whereas others might become genetically competent or form spores (López *et al.*, 2009b), depending on the species. In biofilms, differentiation into various cell types has also been well-documented (Vlamakis *et al.*, 2008). This variation in cell types in isogenic populations is considered a bet-hedging strategy, as it prepares the species for abrupt environmental changes (Veening *et al.*, 2008c, de Jong *et al.*, 2011b). It is therefore important to study bacterial gene expression at a single cell level. Flow cytometry has been used for this (Shah *et al.*, 2006, Robert *et al.*, 2010), but due to the small size of bacteria, this technique requires relatively strong fluorescence signals. Flow cytometry is also impractical when dealing with species that form cell chains. Therefore, many single-cell gene regulation studies with bacteria use epifluorescence light microscopy. To obtain data from a sufficient sample of cells it is desirable to have automatic analysis software that can interpret the microscopy images. Several software packages have been developed to do this, including *CellProfiler*, *MicrobeTracker*, and plugins for ImageJ like *TLM-Quant* (Carpenter *et al.*, 2006, Sliusarenko *et al.*, 2011, Piersma *et al.*, 2013). These methods often use thresholding of the phase contrast image to outline cells. This works well with bacteria such as *E. coli*, *Salmonella* Typhimurium or *Caulobacter crescentus*, which divide by a visible constriction followed by cell separation. However, this is not the case with the model system *B. subtilis*. During exponential growth, *B. subtilis* cells do not immediately separate after septum synthesis has completed and often form long cell chains. Moreover, since *B. subtilis* divides by forming a crosswall (septum) that cannot be observed by phase contrast microscopy, it is often impossible to define the boundaries of a discrete cell. This hinders the use of phase contrast images to identify cells, both in manual and automated analysis. In this chapter, two methods are tested

that enable semi-automated single cell measurements in cell chains. The first method is called *ChainTracer* and uses membrane stain images to define cell boundaries within a chain. For cases where ChainTracer cannot be used because images are too crowded, a second method called *NucTracer* is presented, which uses fluorescently stained nucleoids as a proxy for single cells. Both methods run under the plugin *ObjectJ*, which in turn is connected to the popular Java-based image processing program *ImageJ* (Schneider *et al.*, 2012). ObjectJ supports non-destructive hierarchical marking, and integrates analysis across many multi-channel images, while maintaining active links between marked images and results for easy navigation between the two. The use of ChainTracer is contextualised by measuring motility development in a growing *B. subtilis* culture as an example.

B. subtilis exhibits a wide range of adaptations used to survive in soil, its natural habitat. Most of these differentiation processes are only activated in a subset of cells within the population. Examples of this bimodal regulation are motility, natural competence, and sporulation (Kearns and Losick, 2005, Maamar and Dubnau, 2005, Schultz *et al.*, 2009). Motility is switched on by induction of σ^{D}, the sigma factor responsible for transcription of the flagellum genes (Márquez-Magaña and Chamberlin, 1994). σ^{D} activates its own expression together with the transcription factor SwrB, which is encoded by the neighbouring and co-transcribed gene *swrB*. This positive feedback together with the positioning of *sigD* at the end of a very long mRNA, makes it more prone to RNA degradation noise, which results in a bimodal induction of motility during the exponential growth phase (Cozy and Kearns, 2010). Despite the positive feedback regulation, cells have the possibility to revert back to non-motile cells (Norman *et al.*, 2013). However, when stationary phase is reached, nutritional conditions ensure that most cells have become motile. The induction of motility is accompanied by the production of autolysins that act to release motile cells from cell chains. The effectiveness of ChainTracer was assessed by measuring the activity of the flagellin promoter P_{hag} in single cells (Kearns and Losick, 2005). In addition, it is demonstrated how ChainTracer can be used to measure cell length and diameter, as well as cell chain length. The results show that ChainTracer is a useful automatic image analysis tool when the outline of bacterial cells is difficult to visualise by phase contrast.

3.2 Results and Discussion

3.2.1 ChainTracer project file

Individual *B. subtilis* cells growing as chains and viewed under phase contrast illumination are difficult to outline by both automated and manual methods (figure 3.1A). However, staining cell membranes of these cells with a fluorescent dye will

unambiguously reveal their outlines. Currently, no software exists that can use a membrane stain to identify bacterial cells in microscopy images. this problem was approached by writing two task-specific project files for the ImageJ plugin ObjectJ (Vischer *et al.*, 1994, Pandey *et al.*, 2013). ObjectJ extends the way user-defined regions of interest (ROIs) are handled. The plugin collects raw data, ROIs, results, and task-specific scripts into a self-contained `.ojj` project file. The key advantage of ObjectJ is that users can easily navigate between collected data and the original images. Many anomalies can be quickly detected, as ObjectJ allows for sorting by any measured property, so the user can browse through cells with extreme values. Unwanted cell ROIs can then either be deleted or temporarily disqualified. The scripted tasks required for analysis are contained in two separate `.ojj` project files corresponding to the input images used; ChainTracer analyses membrane-stained cells, and NucTracer analyses nucleoid-stained cells. Images are stored in a hyperstack format, which is a `.tiff` file containing multiple images arranged by channel and frame dimensions.

ChainTracer is used for membrane-stained cells and carries out four semi-automated steps to obtain measurements of individual cells within a chain. In the first step, filamentous shapes (cell chains) are identified from a phase contrast image, and are marked as chain objects. In the second step, the filaments are straightened to aid analysis. In the third step, fluorescence intensity peaks along the chain axis in the membrane stain channel are used to automatically detect septa (figure 3.1B). The septa are indicated by triangle markers placed along the cell chain (figure 3.1C). In the fourth step, the chains are resolved into cells, which are marked as cell objects, which in turn have items associated with them (figure 3.1D–F). Between each step, the user can manually check and modify objects and items via the graphical user interface. For example, filament traces can be added, removed or edited; within a filament, septum positions can be added or deleted in cases where automatic detection has not worked (figure 3.1B–C). Once the individual cells have been identified, the amount of integrated fluorescence per cell in the green channel (GFP) is measured inside a box between two adjacent septa (figure 3.1C, cell #4). Alternatively, the mean intensity along the cell axis can be used. This may be preferable in cases where filaments form clusters, which would cause measurement boxes to sometimes overlap with a neighbouring cell (figure 3.1C, cell #3).

3.2.2 Single cell measurements with ChainTracer

To test ChainTracer with membrane-stained cells, it was used to measure the induction of motility in an exponentially growing *B. subtilis* culture. Motility was monitored by using the flagellin promoter P_{hag} fused to *gfp* as reporter (Kearns and Losick, 2005). Samples of growing *B. subtilis* were collected at 4 time points (figure 3.2A), cells were

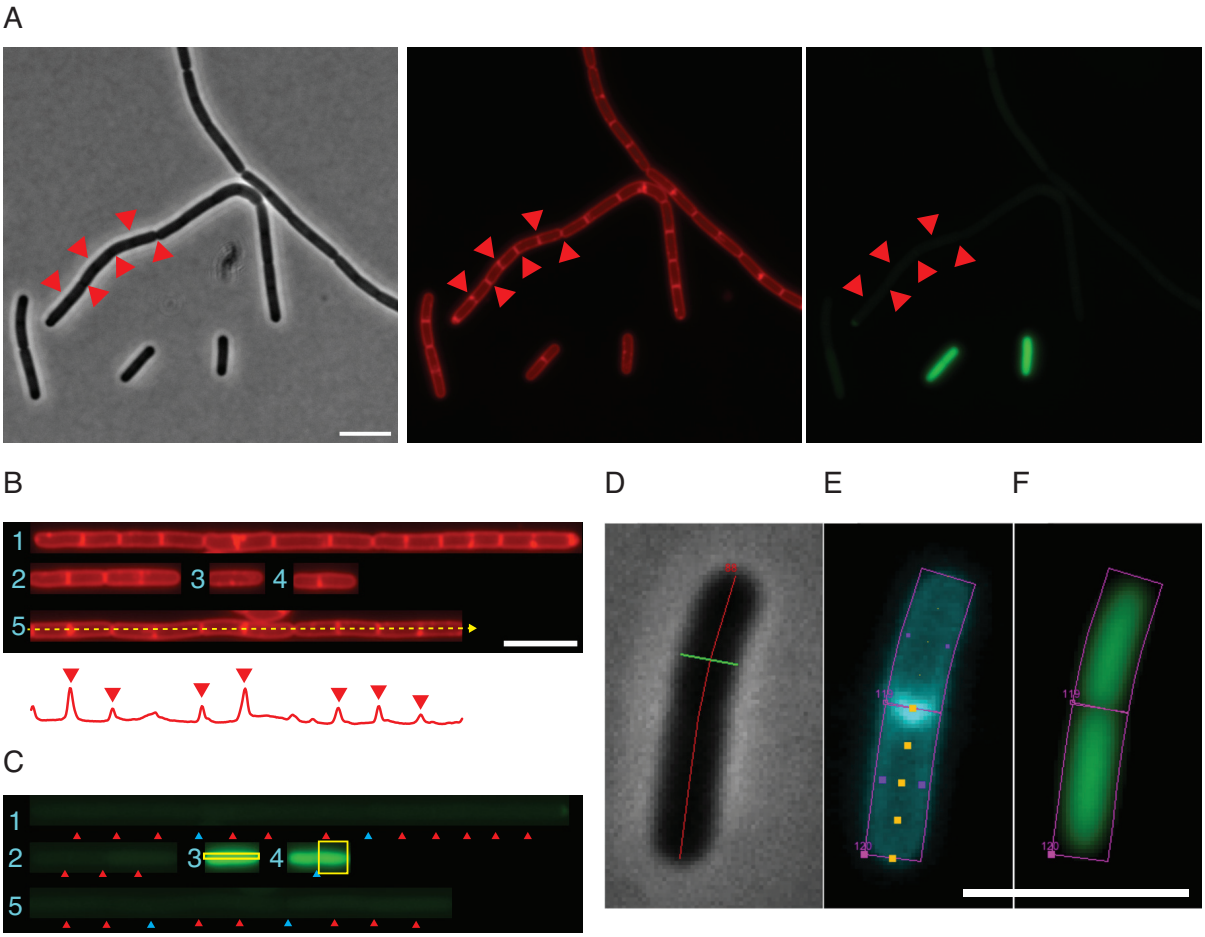

Figure 3.1: Microscopy images showing a heterogeneously motile *B. subtilis* culture during exponential growth in liquid medium. (A) Division septa (some indicated by red triangles at consistent positions across channels in A) can be difficult to identify from phase contrast images. from left to right: Phase contrast, fluorescent membrane stain, and GFP images. (B) ChainTracer screenshot showing straightened filaments. Filaments (objects) are numbered. A line scan (example shown as a dashed yellow arrow running through object #5, results shown in red graph underneath) is used to measure the fluorescence intensity along a filament. Intensity profile of object #5, with corresponding line scan output shown below the screenshot. Automatically detected peaks (septa) are indicated by red triangles over the intensity profile. (C) ChainTracer screenshot of the corresponding GFP channel. Automatically detected septa (items) are indicated by red triangles; manually added septa are indicated by blue triangles. The yellow boxes represent the two methods of fluorescence intensity data capture; a box encompassing an entire cell (cell #2 in filament #4) captures integrated GFP fluorescence, and a narrow box (cell #1 in filament #3) captures mean GFP measurement. (D-F) Summary of ObjectJ items making up an object in ChainTracer. (D) Cells shown in phase contrast, traced by a chain axis item (red), and bisected by a chain diameter (green) item. (E) The same cells as in D in a membrane stain channel resolved into two cells, each bounded by a cell box item (magenta), and an individual cell traced by a cell axis item (yellow dots) (F) The cell in the GFP channel, with cell box items (magenta). Scale bars are 5 μm.

stained with the fluorescent membrane dye Nile red and imaged using fluorescence microscopy. During analysis by ChainTracer, roughly 1 in 10 septa were not detected or were incorrectly specified (figure 3.2B). It should be mentioned that these values can be improved by optimising the conditions for membrane staining. Figure 3.2C shows fluorescence intensity data produced using ChainTracer as intended (when making manual adjustments between automated steps). The resulting histograms show clear bimodal distributions over time, with subpopulations exhibiting low and high GFP signal. This is in good agreement with previous reports (Kearns and Losick, 2005). To assess the accuracy of the ChainTracer data, a manual analysis of GFP intensities of individual cells was performed using ImageJ (figure 3.2D). The distribution of values obtained from a fully manual analysis shows a very good agreement with the data obtained with ChainTracer.

3.2.3 Chain length and cell length measurements

Motility development is accompanied by the induction of autolysins that help to release the cell from its non-motile siblings. ChainTracer records the hierarchy of chain objects and the number of cell objects they contain. This functionality was tested with the data set used in figure 3.2C. Since motile cells can still divide, cells were grouped in three classes; singlets and doublets, 3 to 4 cells joined together, and 5 or more cells joined together. The resulting stacked bar chart (figure 3.3A) shows how a culture shifts from predominantly longer cell chains, to single or double-cell chain configurations. Figure 3.3B shows that the induction of flagellin production is indeed strongly correlated to chain length.

ChainTracer automatically measures cell lengths, which made it possible to examine whether the length of motile cells changed over time. Based on the bimodal distribution in figure 3.2C, and cell chain data of figure 3.3B, a GFP fluorescence cut-off at 1,000 AU was used to classify cells as motile or non-motile (figure 3.3C). The overall cell length decreases at the end of the exponential growth phase, when the culture approaches the stationary growth phase, which is not surprising since slower-growing cells have more time to complete septation. Interestingly, in early log-phase, the motile cells are shorter compared to non-motile cells, suggesting that these cells grow slower. This difference levels off at later time points. While data for cell diameters was also collected for this data set, no significant change dependent on motility could be observed.

3.2.4 Single cell measurements with NucTracer

NucTracer is used for nucleoid stained cells, as detecting single cells by using a membrane stain comes with some limitations. Firstly, many fluorescent membrane dyes,

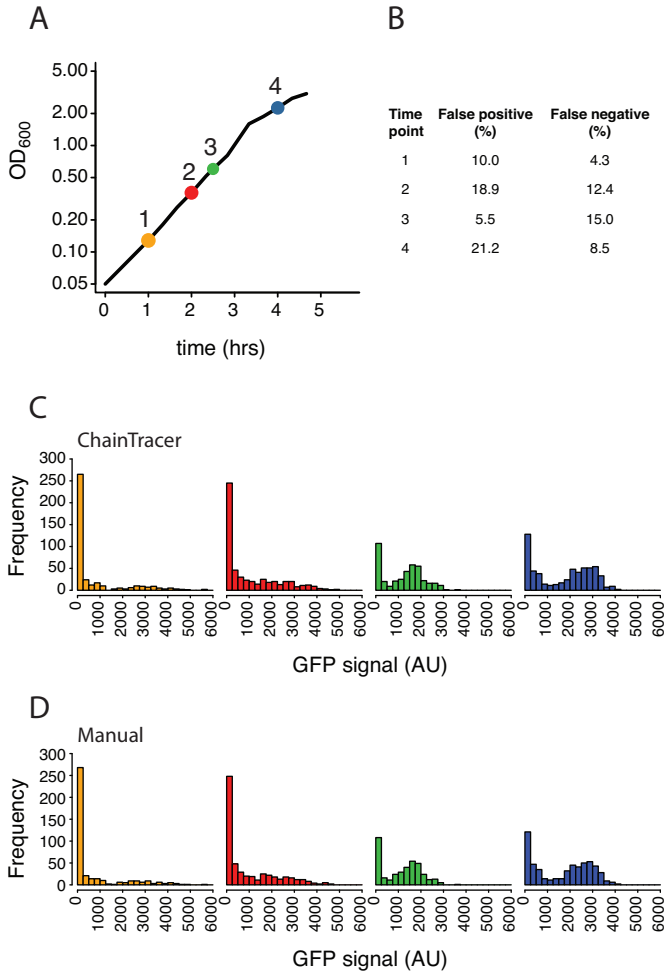

Figure 3.2: Bimodal expression of P$_{hag}$-*gfp* during growth, assessed by both ChainTracer and a fully manual analysis. (A) Growth curve of bSS339 carrying the P$_{hag}$-*gfp* reporter fusion grown in CH medium. Coloured circles indicate time points where samples were collected for microscopy. Colours are re-used in subsequent figures of this chapter. (B) Table showing success rates for septum detection by ChainTracer for each time point (C) Histograms of single-cell measurements obtained with ChainTracer showing GFP fluorescence intensities at time points 1 to 4 in panel A. $n = 401, 544, 405, 585$, respectively (D) Manually acquired GFP fluorescence intensities at time points 1 to 4 in panel A. $n = 397, 532, 386, 565$, respectively.

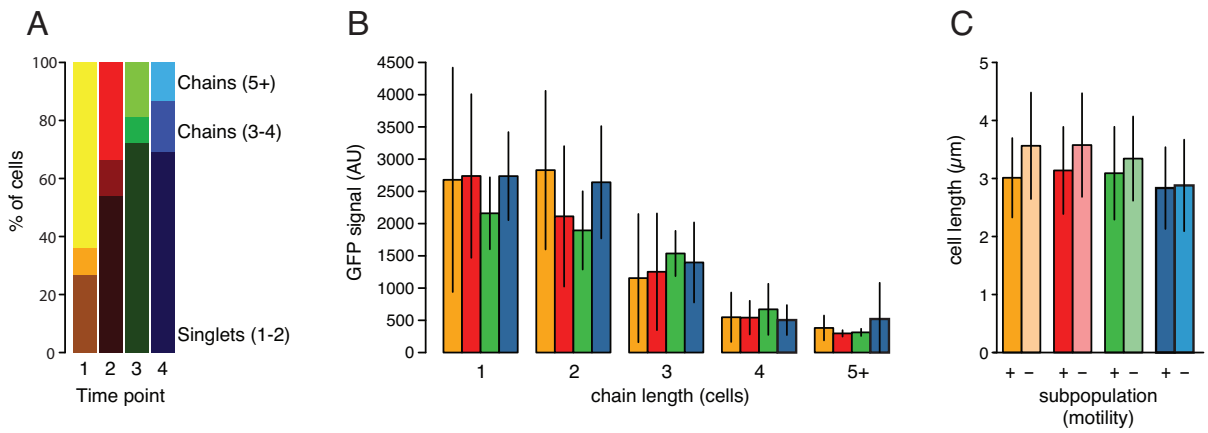

Figure 3.3: Cell chain and cell length measurements. (A) Stacked bar chart showing the distributions of chain lengths as expressed in number of cells in a chain. Dark colours indicate singlets and doublets, intermediate colours indicate 3-4 cells, and light colours indicate chains of 5 or more cells, respectively. (B) P_{hag}-*gfp* expression in relation to cell chain length. (C) Cell length measurements split by time points, and by the state of motility (on/off). The threshold value for determining motility was 1,000 AU GFP, see main text for details. For all panels, the colours used for bars and lines correspond to time points in figure 3.2A. Error bars indicate one standard deviation.

such as FM4-64 and Nile red have excitation and emission spectra that are incompatible with red fluorescence reporter proteins such as mCherry. Secondly, automated detection of cell chains requires that they should not make contact. However, statistical testing often requires large cell numbers, and crowded images are preferred in order to efficiently sample a culture. Therefore, another method was also developed to measure gene expression in single cells. This method uses the nucleoid stain as proxy for a single cell. The advantage is that; (i) the commonly used nucleoid stain DAPI is compatible with both GFP and red fluorescent proteins, (ii) nucleoids of different cells are always physically separated, which allows for high cell densities, and (iii) nucleoids are always located at the centre of the cell. Of course, a cell often contains two nucleoids and therefore this method does not identify single cells in absolute terms. However, when a sufficient sample of cells is counted, this method will present an accurate picture of the distribution in gene expression over the cells in a culture.

To collect data with the help of nucleoid staining, NucTracer was developed and tested by analysing the same microscopy images used for testing ChainTracer in figure 3.2. This was possible since the cells were stained both with Nile red and DAPI. NucTracer detects local maxima of DAPI intensity peaks, and a small circle with a diameter slightly less than that of a typical *B. subtilis* cell (0.66 μm) serves as the ROI used for GFP intensity measurement (figure 3.4A). Figure 3.4B shows the GFP intensity distribution over the nucleoid-determined ROIs at the 4 different time points, without any manual input during data collection. The distribution is clearly bimodal and is

comparable with the manually collected measurements shown in figure 3.2C. The number of ROIs is clearly higher than the number of cells (compare n in figure 3.2 and figure 3.4), but this discrepancy does not influence the overall distribution of fluorescence intensities (figure 3.4C).

3.3 Conclusion

This chapter describes the development of two methods that can be used to automatically measure gene expression in cells that form long chains and that are indistinguishable by phase contrast. The advantage of the membrane staining method to trace cells is that cell length, and cell chain length can be determined. This can potentially be used as a proxy for motility, without needing to use a P_{hag} reporter construct. The advantage of the nucleoid staining method is that it is compatible with both green as well as red fluorescent reporter proteins, and that crowded images can be quickly analysed. Finally, to assess the requirement for manual input when using ChainTracer, the frequency distribution of GFP fluorescence intensities was used as an indicator of data quality. Figure 3.4C shows that performing only the automated steps of ChainTracer does not produce data of sufficient quality, as its distribution deviates greatly from that of the two other methods.

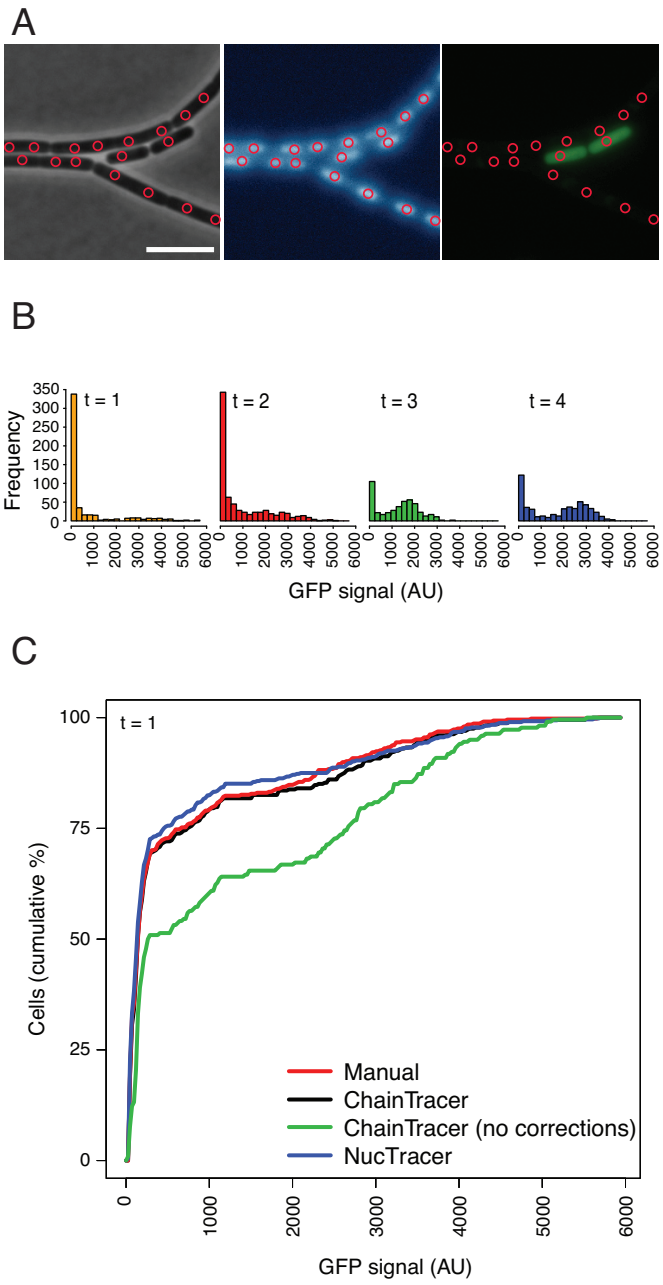

Figure 3.4: Single cell measurements using DAPI staining and NucTracer. The same dataset used in figure 3.2 was used to identify cells based on their DAPI intensity peaks. (A) Phase contrast, DAPI and GFP channels of the same microscopy image. Pink circles indicate locations of intensity peaks in the DAPI channel, which are used to collect GFP fluorescence intensity values. Scale bar is 5 μm (B) GFP fluorescence intensities collected using NucTracer. $n = 496, 725, 454, 524$. Colours correspond to the time points used in figure 3.2A. (C) Cumulative frequency graphs showing the distribution of fluorescence intensities captured from the same sample using four different methods. All methods report a similar distribution, except a fully automated run of ChainTracer without any user input.

Chapter 4

Identification of a non-motile subpopulation of *B. subtilis* with elevated P_{rpsD} activity

4.1 Introduction

Although possessing identical genomes, bacterial populations can generate a surprisingly diverse range of cell fates and developmental niches, even under well-mixed conditions. The Gram-positive soil bacterium *B. subtilis* is a popular model organism for studying bistable and heterogeneous gene expression, and many of its nutrient stress responses appear to follow a strategy of bet-hedging (Veening *et al.*, 2008b). A portion of a growing culture of *B. subtilis* may develop motility as a response to nutrient limitation cues, and may also become genetically competent or sporulate. The success of bistable differentiation is partly due the remainder of the population being free to continue growing if the decision to differentiate proves inappropriate.

The generation of such bistable populations has two requirements; a regulatory mechanism that is prone to noise, and a feedback loop capable of reinforcing the regulator into an activated state when it surpasses a threshold level. The subject of this chapter is the bistable generation of motile cells in a growing culture, which is regulated by the production of σ^{D}, the motility sigma factor. The *sigD* gene is located on the 26 kb *fla/che* transcript, and its location at the penultimate 3′ position exposes it to exonucleolytic degradation, as well as the possibility of premature transcription termination (Cozy and Kearns, 2010). Motility development is affected by the position of *sigD* in the transcript, as its position as the penultimate ORF makes σ^{D} levels prone to gene expression noise. A stable σ^{D}-ON state is achieved through a positive feedback loop, partly by transcription from a σ^{D}-dependent promoter located closer to *sigD*. Also contributing to the reinforcement of the positive feedback loop is the product of the last gene in the transcript, SwrB, which enhances transcription from the original σ^{A}-dependent (a housekeeping sigma factor) promoter for the entire *fla/che* operon (Kearns and Losick, 2005).

In this chapter, evidence is presented for a down-shift in biosynthesis among motile

cells by means of inhibiting transcription of ribosomal protein subunit genes. Initial evidence for a relationship between motility and ribosomal subunit synthesis was provided by collaborators, who carried out a transcriptome analysis which compared motile and non-motile subpopulations on *B. subtilis*. Though using FACS to separate cells is generally considered a single-cell level, in the case of *B. subtilis*, the prevalence of cell chain formation during exponential growth may be a confounding factor (a chain containing many cells can be counted as a single particle [cell]). To verify their claims using another technique sensitive to a single-cell level, proteolytically unstable GFP was used as a reporter of ribosomal subunit promoter activity (P_{rpsD}) in order to enable the use of fluorescence microscopy.

The prevalence of cell chains in growing cultures of *B. subtilis* also posed a challenge to analysing microscopy images—as such, a custom CellProfiler pipeline was used to detect single cells using nucleoid stains. This system is described in detail in section 2.4.1, and is used throughout this chapter (note that the methods in chapter 3 were developed and tested afterwards with intention of making them more widely available; see section 2.4 for a summary of microscopy analysis techniques used in this thesis).

4.2 Results and Discussion

4.2.1 Construction and measurement of heterogeneity from a P_{rpsD} reporter construct

To investigate whether ribosomal subunit promoters in motile cells are less active, the promoter of the ribosomal subunit gene showing the strongest down-regulation, *rpsD*, was used to create a reporter construct. For this purpose, a P_{rpsD}-*gfp* translational fusion was created in the *amyE* locus. To allow comparisons to be made to σ^D activity, a P_{hag}-*mCherry* reporter in the *aprE* locus was constructed in a comparable manner to a previously described P_{hag}-*gfp* construct (Kearns and Losick, 2005).

P_{rpsD}-*gfp* yielded a strong fluorescence signal that was comparable in all cells, whether they were motile or not (figure 4.1A). However, GFP is a very stable protein, and could together with the strong induction from ribosomal promoters mask differences and changes in P_{rpsD} activity. To improve the time resolution of GFP intensity measurements, an unstable variant of GFP was created by adding the SsrA tag as a C-terminal translational fusion (Locke and Elowitz, 2009).

SsrA is a C-terminal peptide tag used by bacteria to designate undesirable peptides to be degraded by the proteolytic ClpXP complex. In its original context, it is introduced by tmRNA in a process called trans-translation; tmRNA binds to a non-terminating ribosome and provides an ORF (*ssrA*) in order to release the peptide and target it for

degradation (Gottesman *et al.*, 1998). When used as a tool to shorten the half-lives of otherwise stable proteins, modified SsrA tags are often used. The three last residues of SsrA are recognised by the unfoldase component ClpX, and targeted mutagenesis has led to generation of a range of SsrA variants with varying efficacies (Andersen *et al.*, 1998).

In this chapter, two GFP-SsrA variants were created, in which the three C-terminal residues, "LAA" ($\mathrm{SsrA_{wt}}$), were replaced by "AAV" and "DAV" ($\mathrm{SsrA_{DAV}}$, $\mathrm{SsrA_{AAV}}$, or collectively $\mathrm{SsrA_{mut}}$). The effect of these tags on the half-life of GFP-SsrA fluorescence was measured and compared to the wild-type tag. To assay this, translation was blocked by addition of chloramphenicol, followed by repeated epifluorescence microscopy measurements of GFP fluorescence over 60 min. GFP signal from the strain carrying the GFP-$\mathrm{SsrA_{DAV}}$ mutant had a half-life of approximately 18 min, whereas GFP-$\mathrm{SsrA_{AAV}}$ exhibited a shorter half-life of approximately 10 min (figure 4.1F–G). This relationship was also reflected by the DAV mutant exhibiting stronger signal than a corresponding AAV mutant. Thus, these tags greatly reduce the half-life of GFP, which when untagged exhibited fluorescence that did not bottom out, but the decline indicated a calculated half-life of approximately 2 h (figure 4.1E). Conversely, the fusion of the wild type SsrA tag to GFP reduced the half-life even further, to approximately 5 min (figure 4.1H).

Andersen *et al.* (1998) observed a similar range of degradation rates when investigating the turnover of various *gfp-ssrA*$_{\mathrm{mut}}$ reporters (including "AAV"). In *E. coli* and *Pseudomonas putida*, the wild-type variety of SsrA always yielded the shortest half-life at 40 and 60 min for the two species. The AAV variant used in this study exhibited a half-life of 60 and 190 min in *E. coli* and *P. putida*, respectively, which is considerably longer than the 10 min that was observed in this study. However, the methods, species, GFP variants and conditions used to assay the half-lives differ, but measurements made with the same method are still comparable. Nonetheless these results agree with those of Andersen *et al.* (1998), in that the effects of different SsrA tags varied in efficacy, and that variations should be expected between species.

P_{rpsD}-*gfp-ssrA*$_{\mathrm{mut}}$ constructs exhibited a decrease in overall fluorescence intensity, but no obvious heterogeneous pattern related to motility was observed (figure 4.1B–C). However, when GFP was fused to a wild-type SsrA tag, clear heterogeneity in fluorescence signals between cells emerged (Figure 4.1D). To examine whether there was a relationship between motility and lower levels of ribosome synthesis, a strain carrying both the reporter constructs P_{rpsD}-*gfp-ssrA* and P_{hag}-*mCherry* was grown as a liquid culture and imaged at regular time intervals during exponential growth (figure 4.2A). The GFP and mCherry fluorescence signals of individual cells were measured and plotted against each other (figure 4.2D–G). The P_{rpsD}-*gfp-ssrA* construct produced visible signal, and it appeared that cells producing stronger GFP signal are predominantly found in the σ^{D}-OFF subpopulation. By classifying each cell by their σ^{D}

and P_{rpsD} states, it was possible to show that the σ^D-OFF subpopulation contained a larger proportion of cells in a P_{rpsD}-ON state, compared to the σ^D-ON subpopulation (figure 4.2D–G). Treating the data as binomial (i.e., how a population is distributed between two possible outcomes) is advantageous here, as it allows the use of Fisher's exact test. Furthermore, when expressed as discrete numbers, the fluorescence intensity data underlying the classifications do not follow a normal distribution, which precludes an unpaired Student's *t*-test. This result agrees with the findings of the transcriptome profiling of the σ^D-ON subpopulation (presented as preliminary data in section 1.5), as it shows that motile cells differ in their transcription levels for ribosomal protein subunit genes. Three other ribosomal subunit genes were tested for the presence of the same behaviour, and though their fluorescence intensities decreased with the presence/efficacy of an SsrA tag, a wild-type SsrA tag completely quenched any fluorescence[1].

4.2.2 ClpXP activity is homogeneous with respect to motility

A possible reason for the observation of P_{rpsD}/P_{hag} heterogeneity could be that the heterogeneous distribution of P_{rpsD} activity is caused by ClpXP activity differing between motile and non-motile cells. However, this is unlikely since differences between these populations were not observed when the activity of the P_{rpsD} promoter construct with slightly weaker SsrA tags was examined (figure 4.1B–C). To corroborate this, an IPTG-inducible $P_{hyper\text{-}spank}$-*gfp-ssrA* construct was created. When expressed inducibly, GFP-SsrA does not reach a concentration sufficient for yielding a detectable fluorescence signal. This can be demonstrated when comparing the GFP signal from the construct in a wild-type and Δ*clpX* background, which shows that its GFP signal is maintained low by constant degradation (figure 4.3). The relative GFP signal yielded by the inducible constructs also agreed with what was previously observed when P_{rpsD} was used to drive the reporters (no SsrA > DAV > AAV > wt SsrA) (figures 4.1 and 4.3).

Lastly, it can be precluded that SsrA itself does not impact motility development and the P_{hag} promoter, as figures 4.1A-D demonstrate that the range of P_{hag}-*mCherry* signal does not change with the presence of, or type of SsrA tag produced.

4.2.3 Autogenous regulation of rpsD by its product S4

Early studies on the P_{rpsD} promoter showed that a secondary structure in the mRNA is able to bind to the S4 protein, providing negative feedback on a translational level (Grundy and Henkin, 1991; 1992). As this was the only means of regulation of the P_{rpsD} known besides those relating to global GTP levels, this region was removed from the reporter construct in order to investigate whether it had an impact on the reciprocal

[1] Syvertsson, MRes project

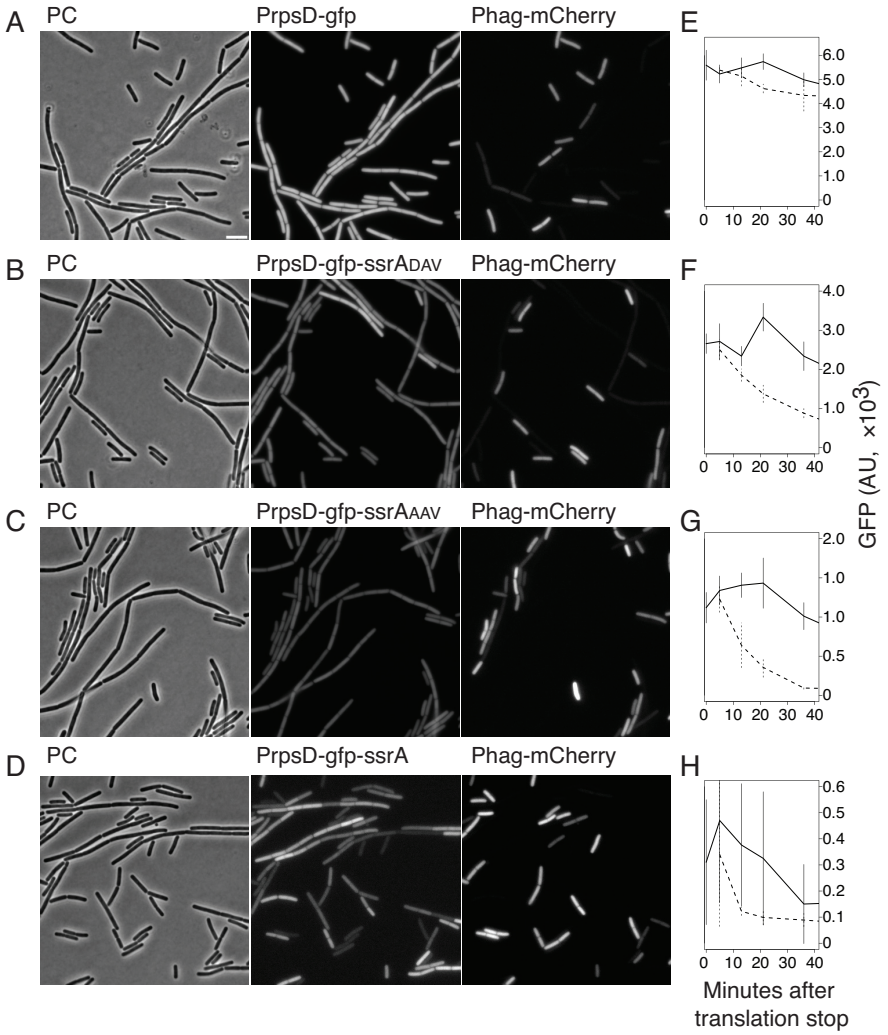

Figure 4.1: Effect of C-terminal fusion of SsrA tags to the P_{rpsD}-*gfp* construct. In A–D, each panel show phase contrast (PC) in the left panel, GFP in the middle and RFP to the right. Scale bar is 5 μm. Contrast levels are consistent in all RFP images, and GFP images in panels A–C. The contrast for GFP in panel D has been set separately to highlight heterogeneity. (A) P_{rpsD}-*gfp* (bSS429) is homogeneously fluorescent and shows no heterogeneity in fluorescence relating to σ^{D} activity (B–C) P_{rpsD}-*gfp-ssrA*$_{\mathrm{DAV}}$ (bSS492) or P_{rpsD}-*gfp-ssrA*$_{\mathrm{AAV}}$ (bSS432) (respectively) exhibit lower fluorescence intensities compared to the P_{rpsD}-*gfp* construct, which reflects their degradation rate. (D) P_{rpsD}-*gfp-ssrA* (bSS387) fluorescence is heterogeneous and appears to be related to motility, as σ^{D}-ON cells often have lower fluorescence intensities. Panels E–H show GFP half-life assays of the GFP variants shown in the adjacent microscopy images. Untreated control samples are shown as solid lines, and the effect of adding chloramphenicol to stop translation is shown as a dashed line. Error bars are 1 standard deviation. At least 10 cells were sampled in cases where a homogeneous fluorescence was observed. Under conditions where GFP signal was heterogeneous, every cell (both motile and non-motile) in an image were sampled to avoid sampling bias. (E) GFP (no SsrA tag) with a calculated half-life of 120 min (F) GFP-SsrA$_{\mathrm{DAV}}$, 18 min (G) GFP-SsrA$_{\mathrm{AAV}}$, 10 min (H) GFP-SsrA, 5 min.

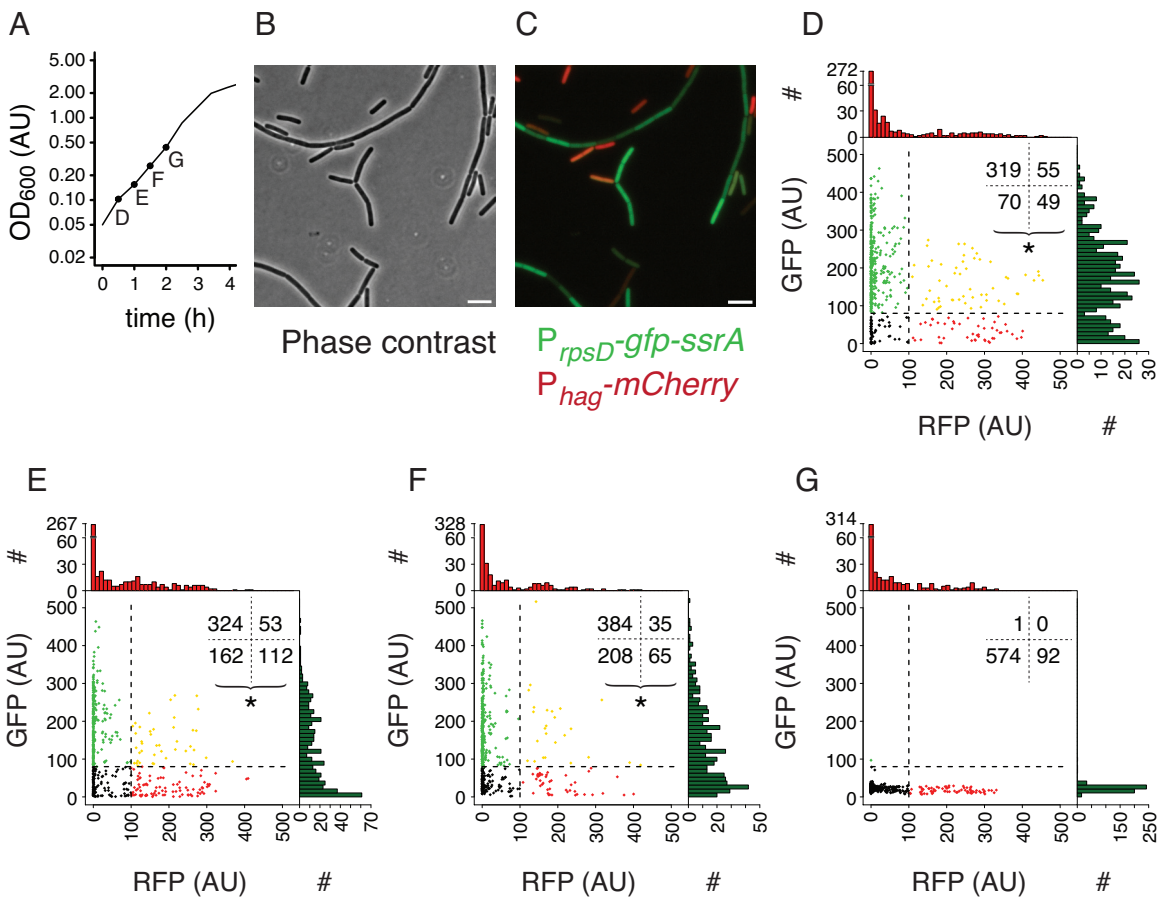

Figure 4.2: A transient difference in P_{rpsD} activity between σ^D-ON and OFF populations. (A) Growth curve showing time points where microscopy was carried out as black dots (letters indicate which figure panel corresponds to the time point). (B–C) Phase contrast, merged image showing GFP in green (showing activity of P_{rpsD}-*gfp-ssrA*) and RFP in red (P_{hag}-*mCherry*). Images were captured at time point 2. Scale bar is 5 μm. (D–G) Scatter-histogram plots of P_{rpsD}-*gfp-ssrA* and P_{hag}-*mCherry* activities of individual cells from the four time points indicated in panel A. Cells are subdivided into σ^D-ON and OFF populations on the x-axis, P_{rpsD}-ON and OFF on the y-axis. Threshold values for ON and OFF populations for the two reporters were determined based on their histograms. Numbers in the top right corners signify cell counts in the four quadrants. A difference in the proportions of P_{rpsD}-ON cells in σ^D-ON and OFF populations was tested for using a Fisher's exact test (*: $p < 0.005$). To aid visual interpretation of plots, the number of data points plotted per panel is limited to 493 by random sampling in cases where total observations exceed this size. Statistical tests were carried out before reducing the sample size.

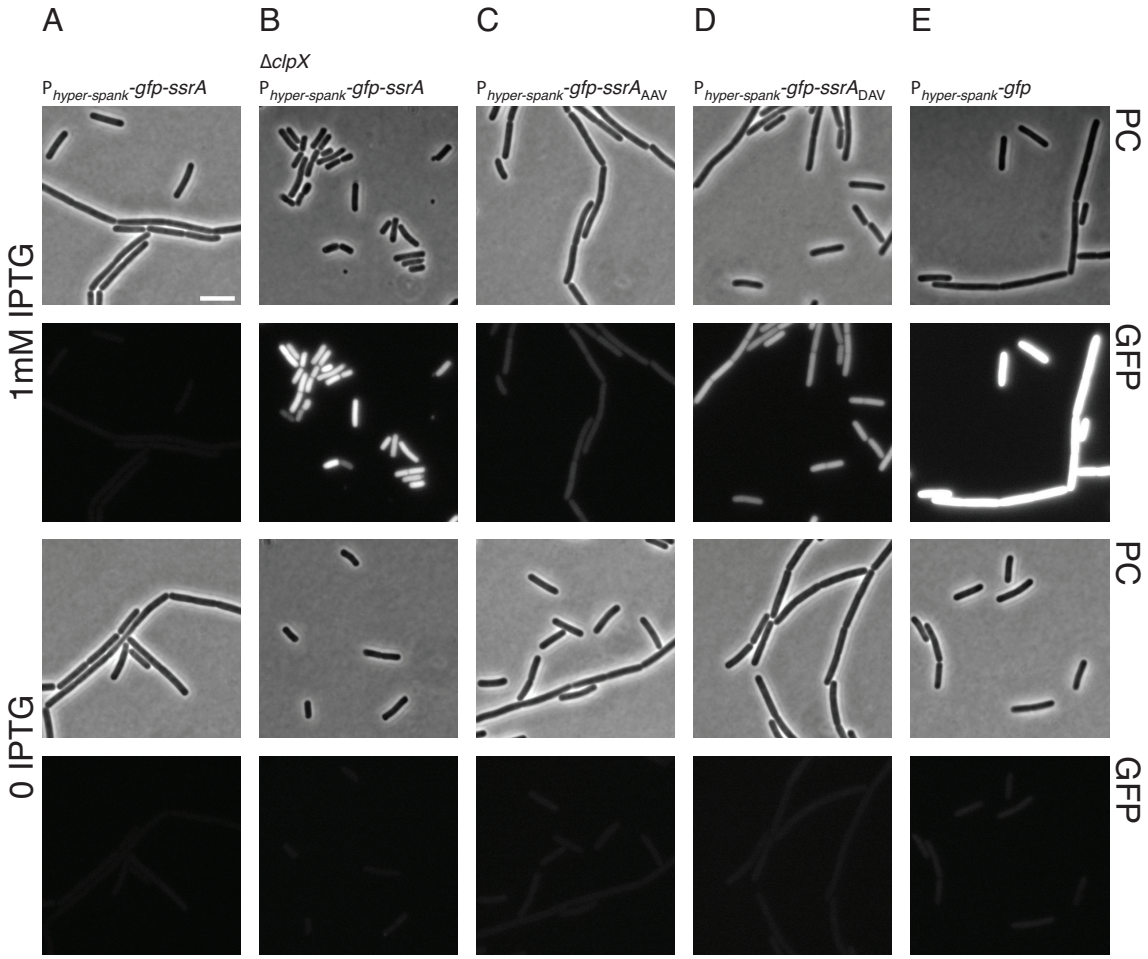

Figure 4.3: Effect of a C-terminal translational fusion of SsrA to GFP driven by $P_{hyper\text{-}spank}$. The top half of each panel shows the strain induced by 1 mM IPTG (PC and GFP), while the bottom shows it uninduced (PC and GFP). Scale bar is 5 μm. Strains used are (A) bSS135 ($P_{hyper\text{-}spank}$-*gfp-ssrA*) (B) bSS485 ($P_{hyper\text{-}spank}$-*gfp-ssrA* Δ*clpX*) (C) bSS162 ($P_{hyper\text{-}spank}$-*gfp-ssrA*$_{\mathrm{AAV}}$) (D) bSS161 ($P_{hyper\text{-}spank}$-*gfp-ssrA*$_{\mathrm{DAV}}$) (E) bSS136 ($P_{hyper\text{-}spank}$-*gfp*).

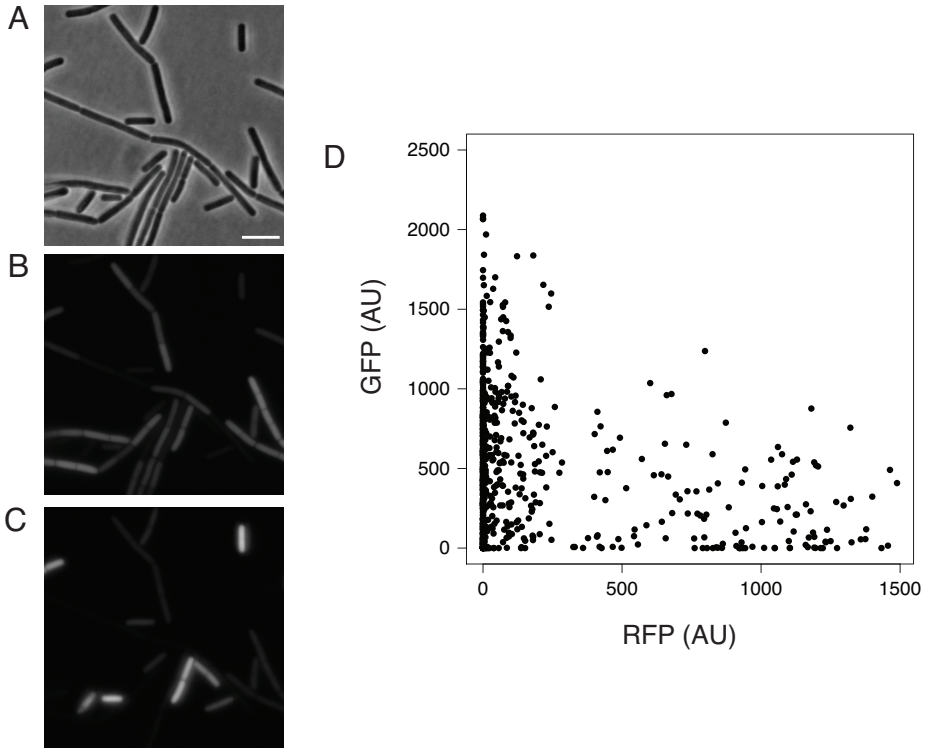

Figure 4.4: Heterogeneous GFP signal observed using a P_{rpsD}-*gfp-ssrA* construct which is not autogenously regulated. Panels A-C show microscopy images from strain bSS399 during exponential growth. The strain carries the constructs $P_{rpsD}\Delta f$-*gfp-ssrA* and P_{hag}-*mCherry*, the former lacking a secondary structure in the reporter gene mRNA responsible for binding and sequestration by the *rpsD* product S4. (A) Phase contrast, scale bar is 5 μm. (B) GFP (C) RFP (D) Single-cell measurements ($n = 887$) from the images in panels B and C, showing a heterogeneous and broadly reciprocal distribution for $P_{rpsD}\Delta f$-*gfp-ssrA* activity when compared to P_{hag}-*mCherry*.

relationship. The part of the untranslated region of the construct corresponding to positions +21 to +156 were removed by divergent PCR and re-ligation. The promoter activity was substantially increased, but heterogeneity did not appear to have been impacted, nor was the reciprocal relationship with the P_{hag} promoter (figure 4.4).

4.2.4 Relaxation of the stringent response does not affect motility induction

It appeared that the increased translational activity somehow regulating motility could be a possible explanation of the P_{hag}/P_{rpsD} relationship. The stringent response is a global regulatory response that limits nutrient spending on the levels of replication, transcription and translation. It is mediated through the small molecule (p)ppGpp and the drop in GTP levels that accompanies its production, which in turn effects a down-regulation of many ribosomal subunit genes, including *rpsD*. RelA is responsible for producing (p)ppGpp during the stringent response, and it was hypothesised that stochastically induced RelA activity in certain cells results in the induction of motility. A RelA-deficient *B. subtilis* strain exhibits a relaxation of the stringent response, and such strains do not produce (p)ppGpp in response to nutrient starvation (Wendrich and Marahiel, 1997). *relA* was deleted to investigate how the regulator affects the P_{hag}/P_{rpsD} relationship. In a strain lacking RelA, neither construct appears to have been significantly affected in terms of heterogeneity or reciprocity, as is shown in figure 4.5A–B.

4.2.5 σ^{D} and motility development do not regulate P_{rpsD} activity

Results showed that motile cells exhibit a lower P_{rpsD} activity in a wild-type background, which made motility regulation a logical venue of investigation. The key motility regulator is the sigma factor σ^{D}. Deletion of *sigD* in a double-labelled strain yielded the expected presence of cells growing exclusively in chains, first demonstrated by Márquez *et al.* (1990), as well as the associated inactivity of the P_{hag} promoter. However, neither the overall expression levels, nor the presence of a heterogeneous fluorescence intensity profile of P_{rpsD}-*gfp-ssrA* was disrupted by the absence of σ^{D} (figure 4.5C–D). This suggests that the motility signal transduction cascade does not regulate the heterogeneity in P_{rpsD} activity. This was confirmed when two other developmental regulators were tested.

SinR represses biofilm formation and indirectly activates motility development, thus reversing its impact on two systems in opposite directions upon biofilm development. In a Δ*sinR* strain, cells form bundled clumps of filamentous cells that have been previously reported by others (Bai *et al.*, 1993). In Δ*sinR*, P_{hag} is less active but

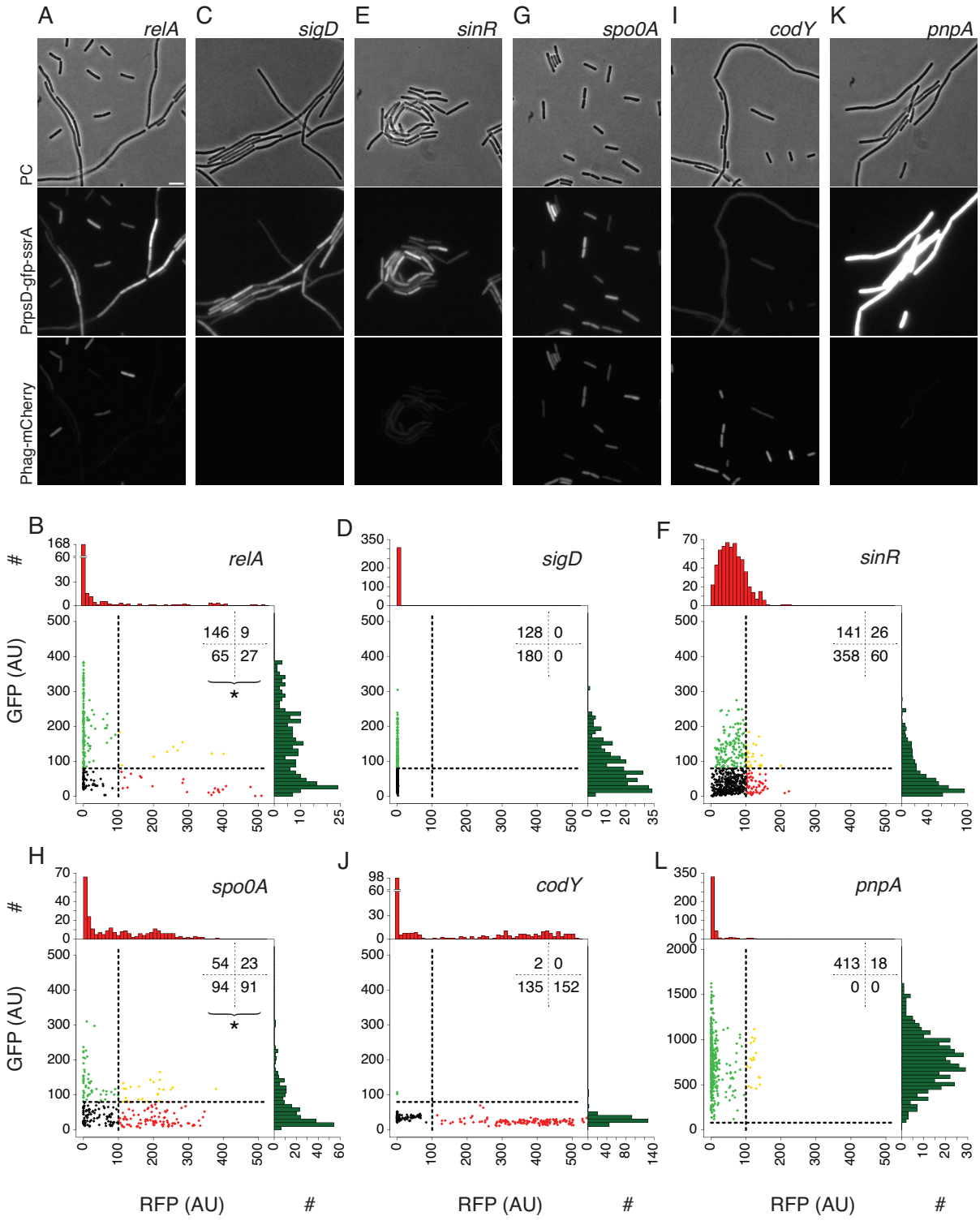

Figure 4.5: (See caption on following page)

Figure 4.5: (See figure on previous page) Regulatory pathways investigated as possible contributors to P_{hag}/P_{rpsD} reciprocity. The top row shows phase contrast images; middle row, P_{rpsD}-*gfp-ssrA* signal images; bottom row, P_{hag}-*mCherry* signal images. Scale bar is 5 μm. Each panel of microscopy images is paired with a scatter-histogram plot (A–B) A *relA* (bSS396) deletion does not visibly affect either system. (C–D) a *sigD* (bSS407) deletion affects motility and the P_{hag} promoter, but the heterogeneous phenotype of P_{rpsD} is unaffected. (E–F) A *sinR* (bSS475) deletion causes clumping cells, but does not affect P_{rpsD} activity, while the P_{hag} promoter appears to be induced at a low level below the motility cut-off. (G–H) A *spo0A* (bSS397) deletion negatively affects cell chaining, but the fluorescence activities of both reporter constructs are still present. (I–J) A *codY* (bSS395) deletion induces motility, and inhibits P_{rpsD} activity. (K–L) A *pnpA* (bSS394) deletion causes increased induction of P_{rpsD}, and inhibits motility

not completely turned off—as was the case for Δ*sigD*—and uniformly distributed across the population. Furthermore, cells were not visibly motile, as they were adhering to each other and forming clumps. Nevertheless, the P_{rpsD} activity in a Δ*sinR* strain appears unchanged from a wild-type background, and exhibits similar heterogeneity and range of fluorescence levels (figure 4.5E–F).

Spo0A regulates stationary phase processes, including motility (Fujita *et al.*, 2005). In a Δ*spo0A* strain, cell chaining appears to be inhibited, but P_{hag} activity is reported over the same range of fluorescence signal as in a wild-type background (figure 4.5G–H). While no comparison between σ^{D}-ON and OFF cells can be made in a Δ*spo0A* background, as the whole population is σ^{D}-ON, heterogeneity in P_{rpsD}-*gfp-ssrA* signal is present, indicating that Δ*spo0A* does not affect this phenotype.

4.2.6 CodY affects P_{rpsD}-gfp-ssrA

Ribosomal subunit promoters are regulated on the level of transcription initiation, and are sensitive to drops in the levels of initiating dNTPs (the +1 position of the transcript), as well as (p)ppGpp. Several ribosomal subunit promoters have a G at the +1 position, as opposed to the more common A found at most transcription start sites (Krásný *et al.*, 2008). GTP together with branched-chain amino acids (leucine, isoleucine or valine) binds to and activates CodY, which leads to repression of several responses cued by nutrient limitation. This also makes CodY responsive to drops in GTP associated with the stringent response. The protein also regulates motility by binding to P_{hag} and preventing transcription of the flagellin (*hag*) gene, and therefore CodY is a potential candidate for investigation (Bergara *et al.*, 2003).

Introducing Δ*codY* into a strain carrying P_{rpsD}-*gfp-ssrA* and P_{hag}-*mCherry* resulted in RFP signal levels in individual cells remaining the same as in a wild-type background, but an effect was seen on the P_{rpsD} promoter. Interestingly, P_{rpsD} activity decreased to a baseline level of fluorescence for both σ^{D}-ON and OFF populations,

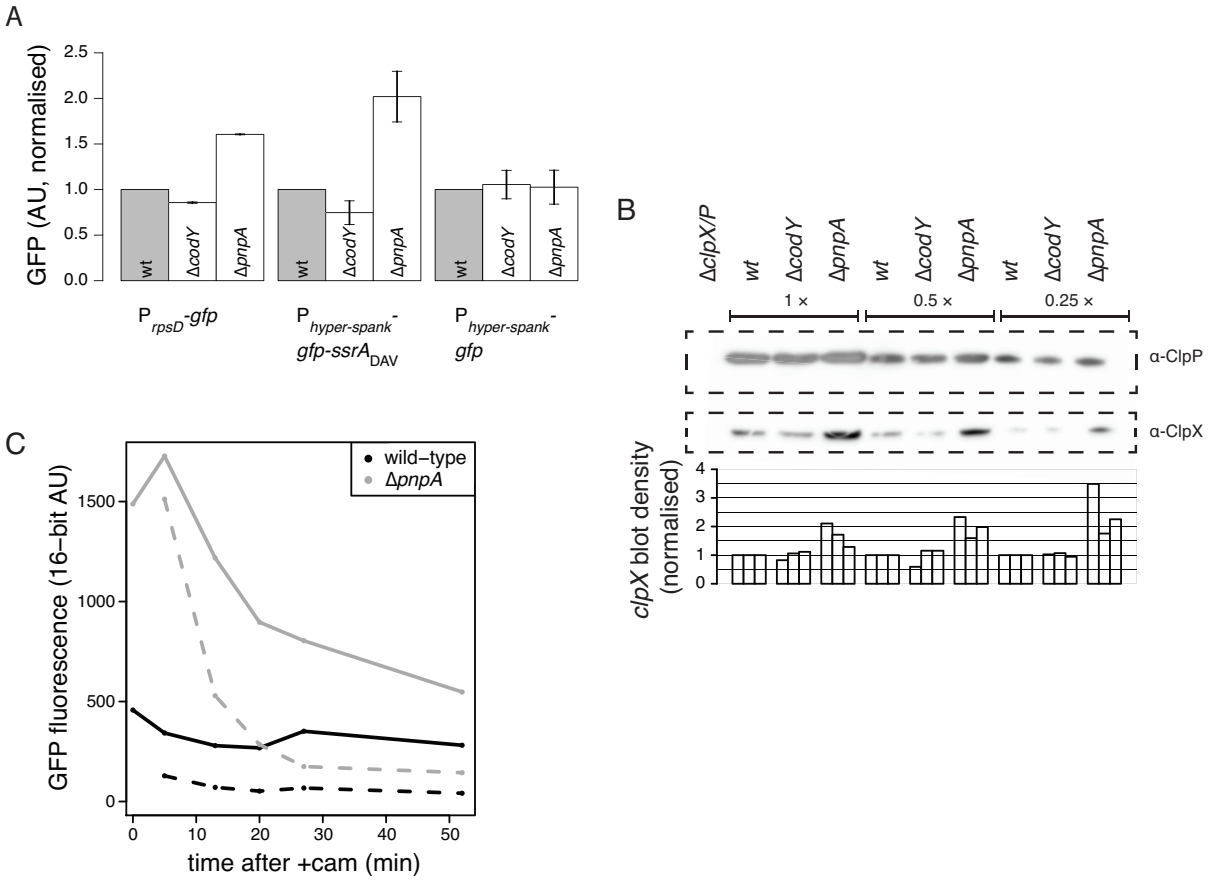

Figure 4.6: Specific effects of Δ*codY* and Δ*pnpA* on P_{rpsD} and ClpXP activity. (A) Effect of Δ*pnpA* and Δ*codY* on GFP signal from three constructs where changes to P_{rpsD} activity, ClpXP degradation, and overall expression levels affect signal strength. Each group of bars has been normalised to the wild-type measurement in panels A–B Strains used (in order of appearance in bar plot): bSS429, bSS459, bSS464, bSS161, bSS251, bSS479, bSS136, bSS250, bSS483. (B) Western blot showing ClpX and ClpP levels in wild-type, Δ*codY* (bSS283), and Δ*pnpA* (bSS340) strains, technical replicates are done as 1:2 and 1:4 dilutions of the first group of samples. The bar chart below blots show densitometry calculations of three independent repeats. Strains LH95 and BD2095 were used as ClpX and ClpP negative controls (respectively). (C) GFP half-life assay performed on a wild-type (bSS322) and Δ*pnpA* (bSS340) strain carrying P_{rpsD}-*gfp-ssrA*. Dashed lines represent fluorescence after the addition of chloramphenicol (+cam). The calculated half-lives were 9.6 and 20.9 min, respectively.

meaning that P_{rpsD} activities of the two subpopulations were indistinguishable (figure 4.5I–J). Using a non-SsrA tagged P_{rpsD} driven construct, $\Delta codY$ showed only a modest reduction in signal compared to a wild-type background. Conversely, the effect of $\Delta codY$ on an inducible but SsrA-tagged GFP also showed a similar reduction in signal, suggesting that $\Delta codY$ affects both P_{rpsD} activity and degradation. As a control, inducible but non-SsrA tagged GFP did not show an increase in signal when in a $\Delta codY$ background, ruling out that the deletion affects overall transcription and translation globally (figure 4.6A). Upon assaying the protein levels of ClpX and ClpP by western blotting, no significant changes in the protein concentration could be detected in $\Delta codY$ strains (figure 4.6B).

4.2.7 ΔpnpA affects both ClpX concentration and activity

One other potential regulator that was tested was PnpA, a polynucleotide phosphorylase which degrades RNA from the 3′ end. Interestingly, it has been shown that a *pnpA* deletion causes cell chaining, possibly suggesting a reduction in motility (Wang and Bechhofer, 1996).

The small protein ComS promotes competence by binding to the ComK antagonist MecA, which frees up ComK to induce a positive feedback loop that locks cells into the competence development pathway. The *comS* gene is located within the first ORF of the very long (25 kb) *srfAA-srfAB-srfAC-srfAD* transcript, which encodes for surfactin production. Translation of *comS* from this long mRNA requires PnpA (Luttinger *et al.*, 1996). The *sigD* gene is also translated from a very long (26 kb) mRNA. It is positioned near the 3′ end of the mRNA, and this position is an important factor in the heterogeneous induction of motility. This was shown by Cozy and Kearns (2010), who moved *sigD* further upstream within the *fla/che* transcript and caused an increase in the proportion of the population that becomes motile.

A $\Delta pnpA$ deletion was introduced to the double-labelled P_{rpsD}-*gfp-ssrA* and P_{hag}-*mCherry* strain bSS387. Indeed, $\Delta pnpA$ led to a reduction in motile cells, and the few cells that developed motility exhibited P_{hag} activity in the ranges found in a wild-type background (figure 4.5K–L).

Interestingly, GFP fluorescence from P_{rpsD}-*gfp-ssrA* in a $\Delta pnpA$ strain increased significantly compared to what was observed in a wild-type background. Furthermore, the distribution of GFP fluorescence intensities was more homogeneous than in a wild-type background. In infrequently observed cells that did develop motility in a $\Delta pnpA$ strain, the P_{hag}-*mCherry* signal appeared to stay in approximately the same range as observed in a wild-type background.

The effect of $\Delta pnpA$ appears to be non-specific, as the deletion is able to increase expression from a P_{rpsD}-*gfp* construct, as well as an inducible but SsrA-tagged construct.

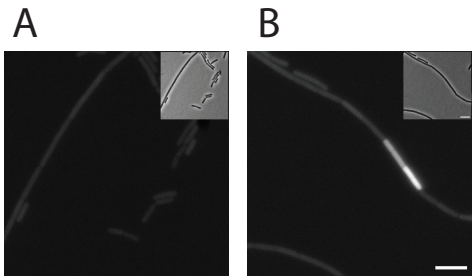

Figure 4.7: Phenotypical differences in P_{rpsD}-*gfp-ssrA* expression between laboratory strains 168CA (bSS107) and PY79 (bSS2). Using a variant of the P_{rpsD}-*gfp-ssrA* construct with non-superfolder GFP, the difference in promoter activity was very clear between (A) 168CA and (B) PY79, where the latter exhibited higher levels of fluorescence under the right conditions. Scale bars are 5 μm.

However, the contribution from inhibiting ClpXP degradation appears to show a greater relative change than that attributable to promotion of P_{rpsD} activity (figure 4.6A). The ClpXP levels in a Δ*pnpA* background was assayed through western blotting, and it appeared that a *pnpA* deletion caused an increase in ClpX levels, but did not affect ClpP levels (figure 4.6B). To reconcile the conflicting evidence of reduced ClpXP activity but increased ClpX levels, the half-life of GFP-SsrA signal was assayed in a Δ*pnpA* strain. The *in vivo* half-lives of GFP signal in a wild-type strain increased more than two-fold in a Δ*pnpA* background, from 9.6 to 20.9 min (figure 4.6C).

This evidence suggests that while there is more ClpX in a Δ*pnpA* strain, it does not improve the efficiency of the degradation machinery. Processes affecting RNA half-life and maturation are likely to yield global effects, and it is perhaps likely that the decrease in degradative efficiency is due to the overproduction of a ClpXP substrate normally present at lower levels, which overloads the machinery (Cookson *et al.*, 2011).

4.3 Laboratory strains 168CA and PY79 differ in P_{rpsD} activity

It was noticed that 168CA and PY79 often differed with respect to the frequency and GFP intensity of $P_{rpsD}{}^{\mathrm{HIGH}}$ cells (figure 4.7). The P_{rpsD}-*gfp-ssrA* construct yielded overall higher levels of fluorescence when expressed in a PY79 background, but it still exhibited a comparable heterogeneity. The two strains are known to differ genetically both on the level of single nucleotide polymorphisms and larger multi-locus deletions which do not exist on the PY79 chromosome when compared to 168CA (Zeigler *et al.*, 2008).

In this section of the thesis, a transposon-based screening approach is used to evaluate whether a genetic locus conferring the difference in this phenotype can be transferred by genetic congression. That is, if it can be transformed by linkage if a selected for transposon with its associated resistance cassette is placed close enough to

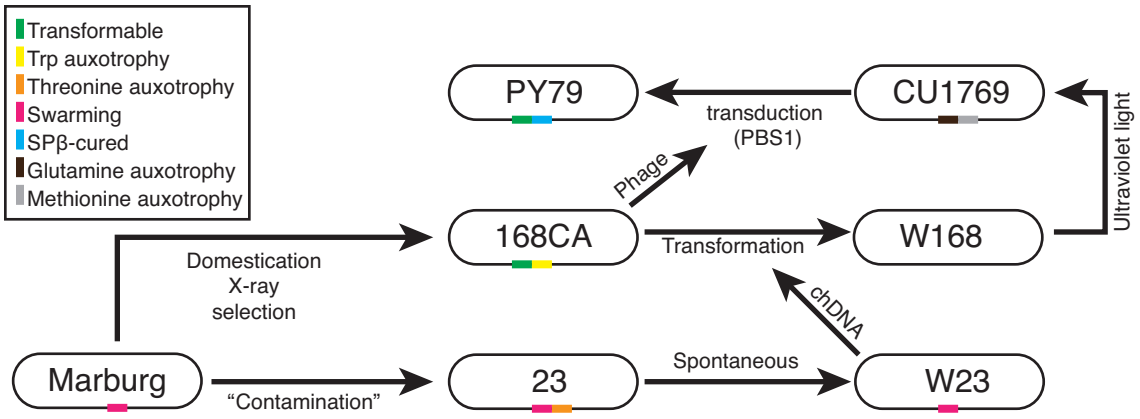

Figure 4.8: Reconstruction of steps taken to create the *B. subtilis* strains 168 and PY79. Diagram showing the domestication process of *B. subtilis* laboratory strains relevant to 168 and PY79. Arrows show treatments used to create a new strain from one already existing. Where a contribution is made by a third strain, it is indicated by an arrow pointing into the first. Relevant properties of strains are indicated by coloured markers. Adapted from Zeigler *et al.* (2008), where the domestication process of several more *B. subtilis* laboratory strains are reconstructed.

the putative locus responsible for the phenotype.

4.3.1 Historical context of B. subtilis laboratory strains

Wild-type isolates of *B. subtilis* are difficult to work with—they form large colonies and are often hard to transform by modern standards. The development of laboratory strains with natural competence allowed for researchers to more readily induce generous uptake and chromosomal integration of DNA. As a comparison, working with wild-type strains requires more laborious methods such as phage transduction. The process of domesticating *B. subtilis* for laboratory work began with the natural isolate *Marburg* (deposited as NCIB 3610), which through successive selection for more manageable colony phenotypes, as well as X-ray mutagenesis led to the creation of the widely used strain 168CA (figure 4.8). At some point during the domestication of the Marburg strain, a non-auxotrophic (tryptophan dependent) strain was created, designated 23, which later gave rise to the strain W23 after a spontaneous mutation. Transforming 168 with W23 to cure its tryptophan prototrophy yielded W168, which was further treated with ultraviolet light to create CU1769, which was finally transduced using PBS1 phage from 168CA to cure it of a SPβ prophage, yielding PY79 (figure 4.8).

4.3.2 Development of a microscopy-based screen to identify the underlying genotype of $P_{rpsD}{}^{\mathrm{HIGH}}$ variance

A technical limitation when making use of the P_{rpsD}-*gfp-ssrA* construct was that the presence of SsrA keeps fluorescence at low levels. This means that using higher-throughput methods such as FACS or even detection of the GFP signal using microplate reader systems is difficult, as the signal strength is close to the detection threshold of the apparatus. Attempts to culture strains carrying P_{rpsD}-*gfp-ssrA* under conditions that favour a $P_{rpsD}{}^{\mathrm{HIGH}}$ phenotype in small volumes in 96-well plates were unsuccessful, likely due to a decreased growth rate of cells in microtitre plates compared to that of a culture flask. The smallest culture volume that was successfully used to observe the $P_{rpsD}{}^{\mathrm{HIGH}}$ phenotype was 800 μl in 2 ml Eppendorf tubes secured along a shaking platform moving along the length of the tube. This protocol allowed for a satisfactory proportion of screened candidates to exhibit a $P_{rpsD}{}^{\mathrm{HIGH}}$ phenotype after 6–7 hours of incubation.

The screen consisted of three rounds of microscopy on separate occasions, where transposants are discarded if exhibiting a $P_{rpsD}{}^{\mathrm{HIGH}}$ phenotype. Screening 200 colonies carrying Tn*YLB-1* insertions from the transformation PY79 × 168CA-tp[2] yielded 5 colonies (table 4.1) that exhibited no $P_{rpsD}{}^{\mathrm{HIGH}}$ phenotype during three consecutive screening opportunities. However, when culturing these colonies under regular, non-high throughput conditions (125 ml culture flasks holding 10 ml media), all but one colony (*yddJ*::Tn*YLB-1*) still exhibited the phenotype that was screened against. This rate of false positives suggests that the screening method does not provide an appropriate substitute for the conditions that the $P_{rpsD}{}^{\mathrm{HIGH}}$ phenotype was observed under earlier in this chapter.

yddJ encodes for a protein of unknown function, putatively a lipoprotein. It is unlikely that the gene itself is directly related to the $P_{rpsD}{}^{\mathrm{HIGH}}$ phenotype, but rather that genetic linkage during transformation was responsible for the introduction of an element from the 168CA chromosome to PY79 that yielded the $P_{rpsD}{}^{\mathrm{HIGH-}}$ phenotype. Back-crossing the transposon from the Δ*yddJ* strain into the same PY79 carrying the P_{rpsD}-*gfp-ssrA* construct indicated a 12.5% linkage (3 in 24 colonies $P_{rpsD}{}^{\mathrm{HIGH-}}$). The feasibility of linkage as a cause of these observations is supported by genomic studies of *B. subtilis* laboratory strains, where Zeigler *et al.* (2008) report six larger regions to be different in PY79 when using 168CA as a reference. One of the reported alterations is a 20 kbp deletion spanning *ydcL* to *yddM* (23 genes), *yddJ* being the 5$^{\mathrm{th}}$ to last gene deleted.

This method of screening yielded inconsistent results, which could partly be

[2] transposon pool chDNA, see section 2.8.2 for details

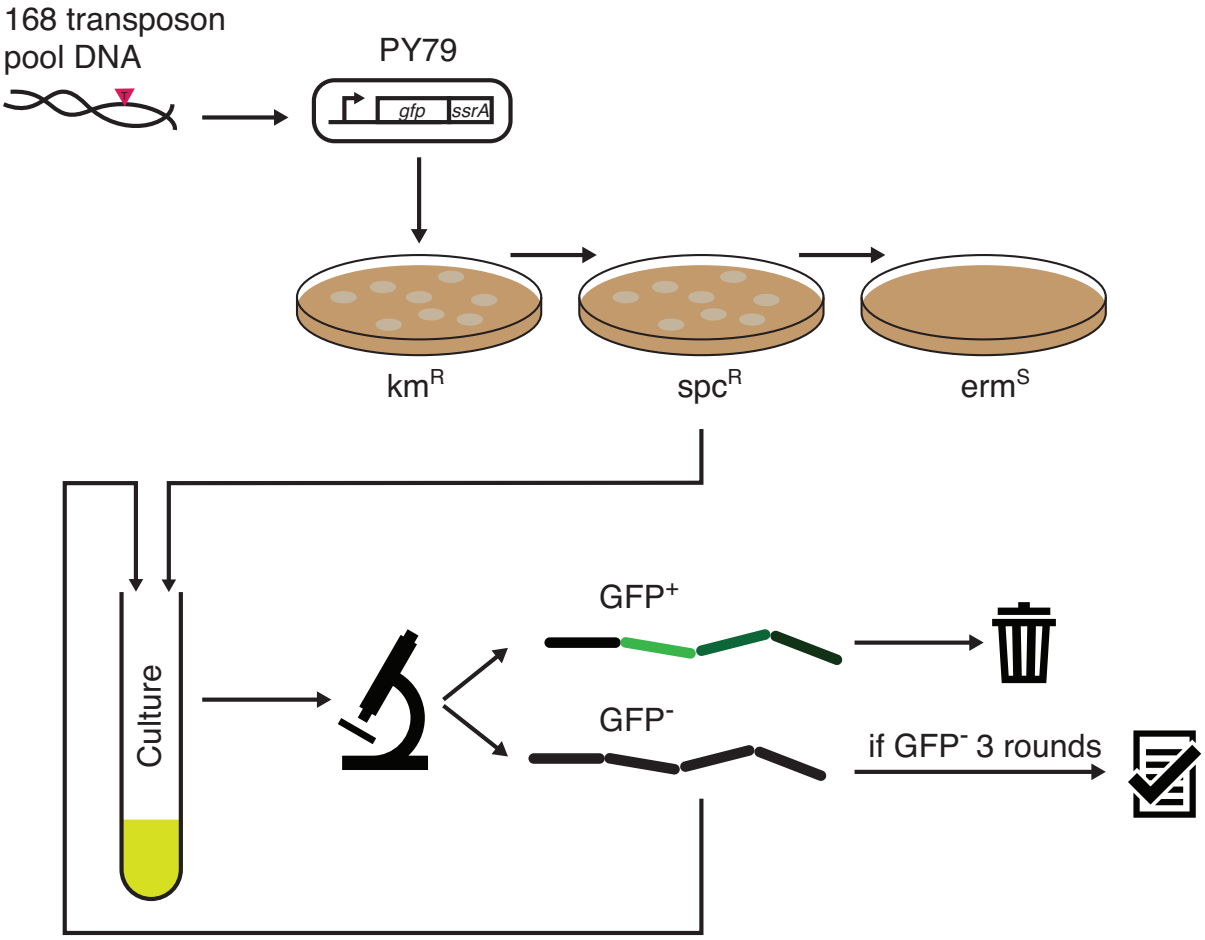

Figure 4.9: Overview of microscope-based screen for transposants with a GFP reduction from PY79 to 168CA levels. PY79 is transformed with 168CA transposant pool chDNA to introduce elements of the 168CA genome through linkage. Successful transposon insertions and retention of the reporter construct is selected for by screening for km, spc resistances and erm sensitivity. Each colony is grown in liquid media and screened for GFP signal using epifluorescence microscopy. If a colony is exhibiting the GFP^+ phenotype of PY79, it is discarded. Colonies that yield a GFP^- phenotype three times are kept for further analysis.

Gene name	Notes
yddJ	putative lipoprotein controlled by AbrB
ribT	riboflavin biosynthesis
ydiR	*Bsu*M DNA restriction system
rrnO-23S rRNA	rRNA
rrnO-16S rRNA	rRNA

Table 4.1: Transposants with a 168CA-like P_{rpsD}-*gfp-ssrA* phenotype. List of $P_{rpsD}{}^{\mathrm{HIGH}-}$ candidate colonies from a screen of transformants of PY79 × 168CA transposant pool chDNA (see section 2.8.2). Colonies listed have exhibited a $P_{rpsD}{}^{\mathrm{HIGH}-}$ phenotype on three independent observations.

blamed on the approach of not monitoring candidate transposants continuously, but instead making a single observation. Furthermore, it was clear that the $P_{rpsD}{}^{\mathrm{HIGH}}$ phenotype is much more sensitive to aeration and agitation of the culture media than anticipated, as the evaluation of a set of candidate transposants under normal conditions showed a return to the $P_{rpsD}{}^{\mathrm{HIGH}}$ phenotype. To summarise, conducting a microscopy-based transposon screen of this phenotype appears to be unfeasible due to technical limitations. However, as one of the finds corresponded to a known difference in PY79, some promise had been shown that the methodology was sound.

4.3.3 Use of a microplate-based screen to screen for changes in $P_{rpsD}{}^{\mathrm{HIGH}}$

The non-satisfactory outcome of microscopy-based screening for loss of the $P_{rpsD}{}^{\mathrm{HIGH}}$ phenotype showed that a microscopy-based approach was not suitable. Due to the relatively low fluorescence intensity of $P_{rpsD}{}^{\mathrm{HIGH}}$ cells, as well as the cell chaining phenotype of *B. subtilis*, microscopy is the only method that can be used to identify the phenotype. However, as mentioned previously, a microscopy-based screen was not suitable to conduct at the high-throughput scale required for a transposon mutagenesis screen. The fluorescence intensity of P_{rpsD}-*gfp-ssrA* is assumed to be too weak to be detected using a microplate reader, so the possibility of using a construct carrying a mutated SsrA tag was investigated.

To verify that a change in strain background would affect such a construct, P_{rpsD}-*gfp-ssrA*$_{\mathrm{DAV}}$ was introduced into 168CA and PY79 and observed under a microscope. Fluorescence levels could be observed to differ, but did not produce a signal strong enough for use with microplate methods. To further increase the signal strength, a modified variant of the P_{rpsD} promoter lacking negative feedback was used (Grundy and Henkin, 1991). The construct $P_{rpsD}\Delta f$-*gfp-ssrA*$_{\mathrm{DAV}}$ produced a measurable difference in fluorescence intensities when the two background strains were compared (figure 4.10A–C). The PY79 background did not produce a higher fluorescence under all conditions, but at some points of a growing culture, the two background strains were indistinguishable in terms of $P_{rpsD}\Delta f$-*gfp-ssrA*$_{\mathrm{DAV}}$ activity. This apparent variation in GFP signal between the two strains further highlighted the need for monitoring the phenotype over time, which can be accommodated at these levels of fluorescence by using a microplate reader (figure 4.10D–F).

Monitoring the fluorescence of 168CA and PY79 strains carrying a $P_{rpsD}\Delta f$-*gfp-ssrA*$_{\mathrm{DAV}}$ construct showed similar OD_{600} and absolute values of fluorescence intensities for the two strains throughout the growth of the culture (figure 4.10D–E). While the two strains could not be distinguished based on their growth curves or GFP measurements, adjusting GFP for OD_{600} yielded a clear difference between the

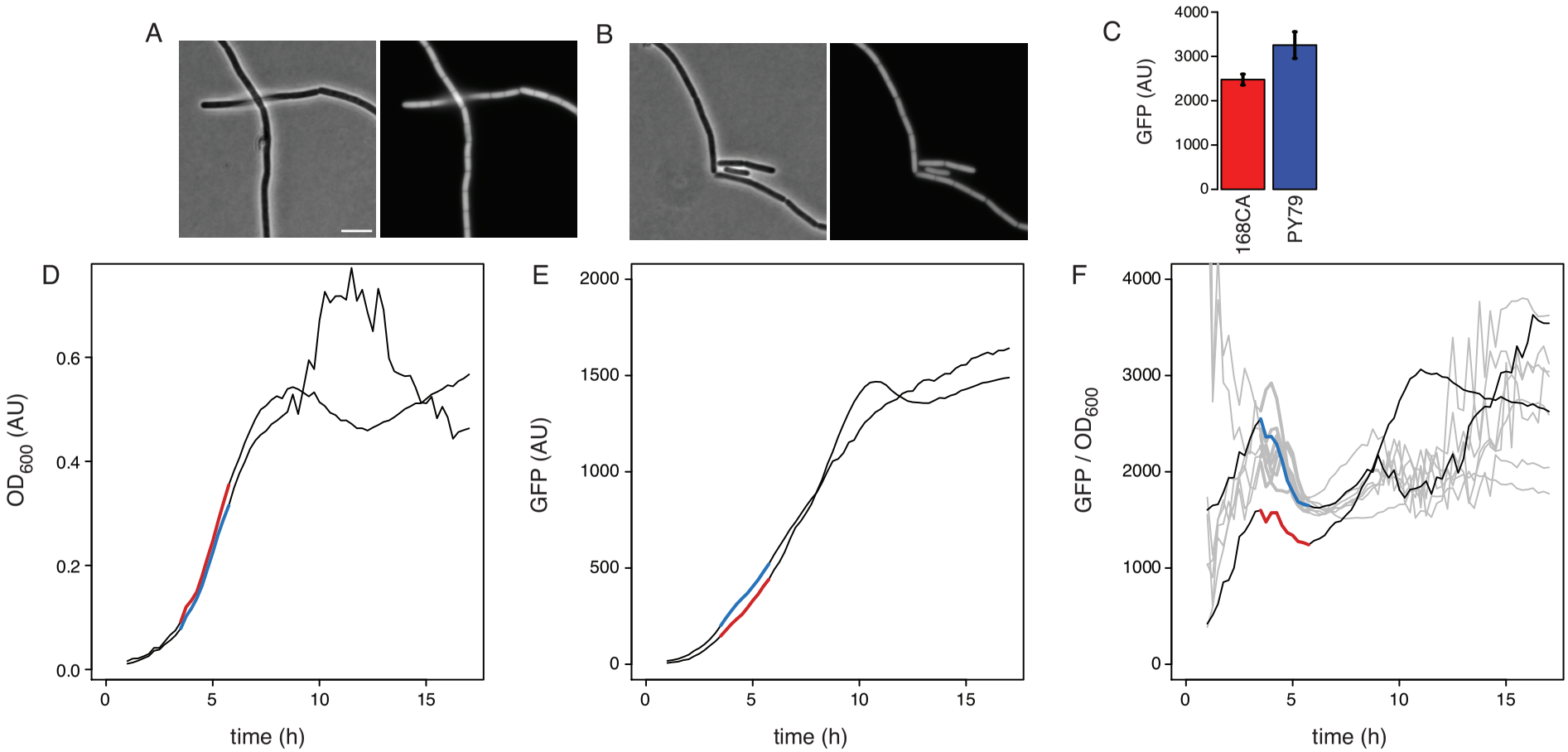

Figure 4.10: Comparison of PY79 and 168CA carrying $P_{rpsD}\Delta f$-*gfp-ssrA*$_{DAV}$. Panels A–B show phase contrast and GFP microscopy images of strains carrying the $P_{rpsD}\Delta f$-*gfp-ssrA*$_{DAV}$ construct. Scale bar is 5 μm. (A) bSS440, a 168CA strain. (B) bSS477, a PY79 strain. (C) Quantified fluorescence intensities from the images in panels A–B. Error bars show 1 standard deviation. Panels D–F show the behaviour of bSS440 (red lines) and bSS477 (blue lines) on a microplate assay. (D) Growth curves of the two strains, segments of the curve used for the screen are coloured. (E) The corresponding GFP measurements for the growth curves shown in panel D. (F) GFP measurements adjusted for OD_{600} levels. Grey curves in the background represent an example of negative results of the transformation bSS477 $\times$ 168CA-tp, which have failed to express a 168CA-like phenotype.

two strains (figure 4.10F). Upon screening transposants for 168CA-like GFP levels, it was clear that most candidates were adhering closer to the curve representing a PY79-like control, as most candidates in the screen were expected to report a negative result (figure 4.10F).

It was noticed that the growth curve of PY79 yields a significant increase in OD_{600} during the stationary phase (figure 4.10D). While this certainly is a striking difference between the two strains, it is likely an unrelated characteristic of PY79, and is probably a result of pellicle formation in the microplate wells. Therefore, the screen was focused on the exponential growth phase, which is marked with coloured thick lines in figure 4.10D–F. As the screen is performed by transferring portions of the 168CA genome to PY79 by linkage, it is however possible that the pellicle phenotype is linked to the $P_{rpsD}{}^{\mathrm{HIGH}}$ phenotype, and any positive candidates may exhibit a change in both phenotypes if such a locus can be transferred.

180 colonies were screened using this method, but no colonies exhibiting a 168CA-like phenotype could be identified. Some observations were made of colonies exhibiting a phenotype with increased GFP signal, but upon closer investigation those colonies were also growing slower, and lower OD_{600} values made them hard to compare to controls. This assay could be improved upon by simplifying the reporter constructs used. Omitting the SsrA tagging would remove any uncertainties whether that component of the construct changes the GFP signal produced throughout the course of a culture. Any candidate colonies identified to have an effect on P_{rpsD} could then be verified using a P_{rpsD}-*gfp-ssrA* using the proper conditions and methods required for expressing the $P_{rpsD}{}^{\mathrm{HIGH}}$ phenotype.

A more targeted approach to moving pieces of the 168CA genome to PY79 could also have been adopted. By using resistance markers on the 168CA chromosome, placed in each of the major six regions differing in PY79, these regions could have been effectively ruled out before a screen was started. Additionally, the effect of re-introducing a missing rRNA operon in PY79 (Zeigler *et al.*, 2008) could have also been investigated prior to starting a screen for the difference.

4.4 Conclusions

This chapter aimed to verify data obtained through transcriptome profiling, by means of fluorescence microscopy. Using this method the initial finding was reaffirmed, suggesting that motility development in *B. subtilis* is connected to lower levels of ribosomal subunit promoter activity.

Another focus was the investigation of possible regulators that could play a role in regulating the reciprocal relationship between motility and ribosomal subunit promoter

activity. However, no regulator with a clear and targeted effect on the phenotype could be identified from a list of likely candidates identified from previous literature. That is, CodY and PnpA were identified as regulators with noticeable effects on both populations, but their global impact on the rest of the cell suggests that the reciprocal regulation is a by-product of ribosomal synthesis regulation, which is achieved on a global level, through GTP levels, among other things.

A similar bistable behaviour relating to ribosome synthesis was reported by Rosenberg *et al.* (2012). Using similar translational GFP fusions, a small population of cells reporting higher P_{rrnO}-*gfp* (rRNA) activity in stationary phase cultures grown in rich media. A key finding in their paper was that expression levels of ribosomal promoters appeared to inversely correlate with variability. In the case of the P_{rpsD} promoter, it would seem plausible that the states of low activity could be a driver for future heterogeneity in times of scarcer. The group also identified σ^{G} as a regulator underlying a burst in transcription during the pre-spore stage of sporulation—though far away metabolically and growth stage-wise, it makes a case for sigma factors being investigated further in the case of the $P_{rpsD}{}^{HIGH}$ phenotype as a future direction of the project.

Broadly, the results in this chapter suggests that a subset of cells are undergoing a premature onset of stress responses, but which stress response leads to the other remains unclear. A puzzling question remains—what is the significance of the exponential growth phase timing at which this phenotype is observed? At this stage of a culture no kind of nutrient source is likely to have been depleted, and yet some cells undertake a response to a stimulus that is seemingly not present. It would appear that these cells are making a gamble, and determining how well it pays off will be a task for future efforts in the field.

Chapter 5

Use of microcolony time lapse movies to analyse gene expression

5.1 Introduction

Observations of single cells sampled from a liquid culture can be noise-prone, as many concurrent lineages are propagating in parallel. This may obscure patterns occurring on a single or few-cell level within a population of closely related cells. The technique of growing bacterial cells on a microscope slide and observing growth *in situ* addresses this issue, and has yielded insights into the role of small variations between closely related cells. One factor that has been possible to identify using this technique is the role of ageing, as defined by the presence of an old cell pole (a new cell pole is defined as one having arisen from the septation stage of cell division). Using custom segmentation and tracking software, one group has successfully demonstrated a successive decrease in growth rate in cells possessing older cell poles (Veening *et al.*, 2008c, Stewart *et al.*, 2005). In the same paper, a developmental crossroads during the course of the microcolony was also identified, where cells either enter diauxic growth, or slow down and become future spore-formers (Veening *et al.*, 2008c).

This chapter will use microcolony time lapse methods to explore the $P_{rpsD}{}^{HIGH}$ phenotype observed in liquid cultures that was described in chapter 4. The first goal will be to establish whether the phenotype is visible under microcolony growth conditions, which differ somewhat from liquid culture. A particularly relevant difference is the inability of cells to be truly motile, as immobilisation on the slide at the interface of solid media and the cover slip does not permit movement, which may be sensed through mechanisms relating to the transition of cells from swimming to biofilm formation (Cairns *et al.*, 2013). Together with an increase in local cell density, this may affect regulatory decisions to stay motile. As it had been established previously (chapter 4) that σ^D activity is related to $P_{rpsD}{}^{HIGH}$, this was a concern. If the microcolony environment is suitable for observing a $P_{rpsD}{}^{HIGH}$ phenotype, it would be interesting to determine whether the higher production rate of ribosomes (the $P_{rpsD}{}^{HIGH}$ phenotype) correlates with a faster growth rate.

5.2 Technical aspects and limitations

Lineage tracking of growing microcolonies is a common method of analysis, and has been employed by several groups who have created custom software for this task (Piersma *et al.*, 2013, Young *et al.*, 2011, Veening *et al.*, 2008c). Automated analysis of a microcolony time lapse consists of two steps. First, cells are distinguished from the background using a visual feature which the background does not possess, such as a fluorescence signal, or appearing darker when viewed under phase contrast illumination. Areas of the image selected through thresholding are further processed to create objects (ROIs) through segmentation, sometimes by using features such as shape to further partition objects into areas of an expected size. Secondly, in the tracking stage of analysis, associations are made between objects from adjacent frames by using their position and shape. By also detecting division events, daughter cells can be linked to a mother cell from a previous frame to create a lineage tree.

Microcolony time lapse data from *B. subtilis* grown in rich media presents several technical challenges that prevent the use of automated software to aid analysis. As described previously in chapter 3, the filamentous cell chain phenotype of *B. subtilis* is unsuitable for segmentation analysis to highlight cell ROIs, thus making it difficult to analyse by existing methods. A further consequence of cells growing in chains is that the expansion of a colony is unlikely to result in a microcolony with a compact shape that fits within the area captured by the microscope camera. This makes lineage tracking difficult, as cells will leave and enter the field of view, causing gaps in the data. Similarly, because cells are physically attached to each other during exponential growth, there exists a greater chance of cells being pushed underneath the monolayer of cells instead of being positioned next to other cells, thus being disqualified from analysis.

Lastly, DAPI has been used to detect cells (see chapter 4) in this thesis, but can however not be used with growing cells due to toxicity from direct association with DNA. Another possibility is using a membrane dye such as FM5-95—which, as opposed to Nile red, can be incorporated into growing cells without lethal effects. However, the availability of segmentation methods based on membrane dyes is limited, and red membrane dyes are incompatible with the mCherry fluorophor used to report σ^{D} activity throughout this thesis.

5.2.1 Autofluorescence in microcolonies

An advantage of sampling bacterial cells from a culture for live-cell microscopy is that cells can be washed in a low-autofluorescence solution (e.g., PBS) prior to immobilisation on a slide. This can greatly improve image quality in some cases, as some media are known to be autofluorescent (Hinterdorfer and van Oijen, 2009, p. 69).

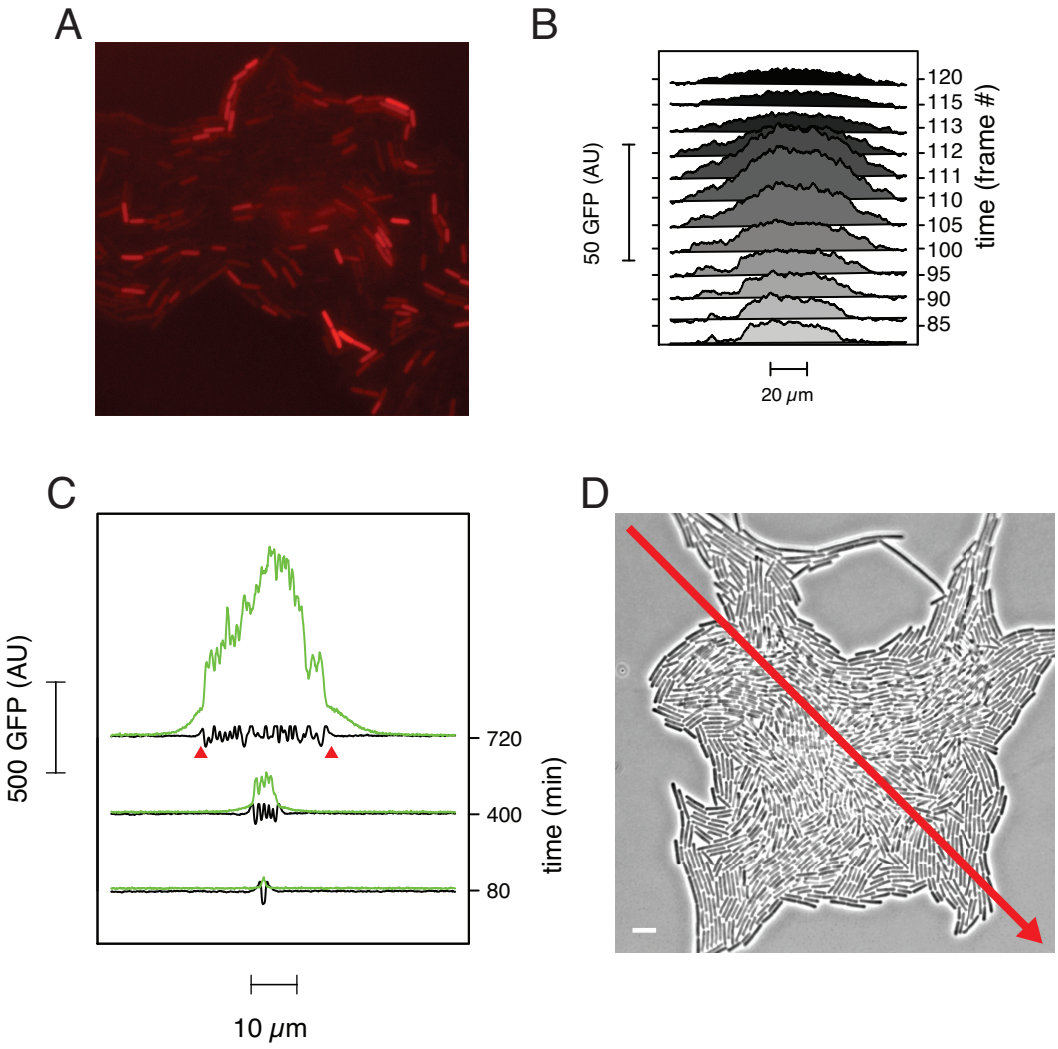

Figure 5.1: Autofluorescence in a *B. subtilis* microcolony. (A) Heterogeneous autofluorescence of 168CA visible using RFP filter settings. Background fluorescence was quantified to ∼130 AU, while cells with the stronger signal showed values of 100 AU above background levels. (B) Times series of line scans measuring the GFP intensity diagonally across GFP images of a stationary phase microcolony of 168CA. Autofluorescence starts increasing around frame 100, then drops suddenly after frame 111. Time between frames was 8 min. (C) Time series of line scans performed across a colony of AH7 ($P_{hyper\text{-}spank}$-*gfp*) producing a strong GFP signal. Black lines show the pixel intensity of the corresponding measurement made on the phase contrast image, which show cells as sub-baseline values (the edges of the colony are highlighted with red triangles). GFP signal is represented by the green line. Note that some signal is present outside of the perimeter of the microcolony as defined by its location as determined by phase contrast. (D) A frame from a microcolony time lapse experiment, showing the location of the line scan as a red arrowed line. Scale bar is 5 μm.

However, microcolonies are imaged *in situ*, and unwanted fluorescence can build up as a result of sheer cell density. To identify sources of autofluorescence in *B. subtilis* under the experimental conditions in this chapter, the lab strain 168CA was grown as a microcolony and imaged in the RFP and GFP channels used for all other strains. Autofluorescence in both channels was observed to be low and homogeneous for the time points that were to be studied, but it was also noted that stationary phase microcolonies exhibit a significant increase in autofluorescence in both channels. The autofluorescence visible using a GFP filter builds up and disappears relatively suddenly, while a more striking and heterogeneous pattern of autofluorescence is observed in the RFP channel (figure 5.1A–B).

As mentioned before, the time points where this behaviour was observed are beyond the scope of this study, but it should also be noted that heterogeneous autofluorescence has not been described in previous literature. As microcolony time lapse is a core method for studying developmental pathways on the single-cell level (Veening *et al.*, 2008c, Young *et al.*, 2011, Robert *et al.*, 2010), particular care should be taken to ensure that this heterogeneous pattern does not become a source of false positives. Though it is not relevant to this thesis to investigate the specific nature of the autofluorescence pattern, it can be noted that the late stationary phase timing and frequency of fluorescing cells is reminiscent of microcolonies tracked for sporulation studies (Veening *et al.*, 2009), which may provide a good starting point further studies of the phenomenon. The fluorescence may be due to a protein not produced during earlier growth phases, such as a nutrient stress related protein, or a stationary phase protein. Lastly, because this finding was unexpected and not pursued any further than to establish that it is manifested in 168CA without the need for a fluorescent protein reporter construct to be present, it is not known whether this phenomenon is dependent on the media used (CH), or any other unique properties of the experimental set-up. Needless to say, attempting to observe the phenotype under conditions used in sporulation studies would settle any concerns around the apparent autofluorescence of *B. subtilis* microcolonies.

Aside from unwanted fluorescence arising from the cells themselves, the use of GFP as a reporter can lead to extraneous signal when used at high levels. To visualise the extent to which GFP signal is visible as a halo outside of a colony, a strain carrying a $P_{hyper\text{-}spank}$-*gfp* construct was grown as a microcolony in the presence of IPTG. Line scans (see figure 5.1D for a visual aid) of both the phase contrast and GFP channel from the resulting time lapse series show that while stronger GFP values overlap with the location of the microcolony, but also that some weaker GFP signal is present past the perimeter of the microcolony (figure 5.1C). This source of error may be difficult to eliminate completely, but it can be partly controlled for by using background

measurements as blanks that can be subtracted from all other measurements. Normally, such measurements are sampled from an empty area of the microscopy image, but in this case it could be placed closer to the microcolony to reduce the effect of the GFP halo. Furthermore, care should be taken to preferentially sample cells growing close to the edge of the microcolony, as the halo effect is presumably greater in the centre of the area covered by cells.

5.3 P_{rpsD}-*gfp-ssrA* activity pulses in a growing microcolony

Microcolonies of strains carrying the two reporter constructs for P_{rpsD} and σ^D activity exhibited a striking pulsing phenomenon. The pulsing featured a heterogeneous and transient increase in GFP fluorescence, which eventually subsided to give way for a rise in mCherry fluorescence (figure 5.2). Occasional observations were also made of a second P_{rpsD}-*gfp-ssrA* pulse, albeit in a far smaller proportion of cells than in the first pulse.

To quantify the relationship between the P_{rpsD} and P_{hag} pulsing, a sample of single-cell intensity values of the double-labelled reporter strain was collected at regularly spaced time points (figure 5.3). The distinctive peaks identified by qualitatively assessing time lapse movies (figure 5.2) were reflected in quantitative analysis of both GFP and mCherry signals, and they were consistently appearing in the same order, with a P_{rpsD}-*gfp-ssrA* peak preceding a peak in P_{hag}-*mCherry* (spaced 200-300 min apart, observed over 3 repeats of the experiment depicted in figure 5.3).

5.4 Tracking a single cell throughout a microcolony

Having established that the reciprocal relationship between P_{rpsD}-*gfp-ssrA* and P_{hag}-*mCherry* is reproducible on the level of a microcolony, the next step was to analyse the behaviour of the reporters on a single-cell level. Several analyses are relevant to the $P_{rpsD}{}^{\mathrm{HIGH}}$ phenotype, and the first one was whether cells exhibiting the P_{rpsD} pulse were growing faster due to what can be assumed to be an elevated rate of ribosome synthesis. It was also interesting to know whether the intensity of the fluorescence signal from P_{rpsD}-*gfp-ssrA* could be used as a predictor for subsequent P_{hag}-*mCherry* intensity. That is, taking into account that previous results show a reciprocal relationship between the two promoters, does a particularly strong P_{rpsD}-*gfp-ssrA* signal prompt weaker P_{hag}-*mCherry* activity, for example?

Due to the technical limitations explained previously, a simplified approach to track individual cells was adopted, based on methods briefly described in previous literature (Elowitz and Leibler, 2000, Levine *et al.*, 2012). Instead of tracking a microcolony from its first frame and building a lineage tree, a single cell was tracked backwards in time from a late frame corresponding to the early stationary phase of the

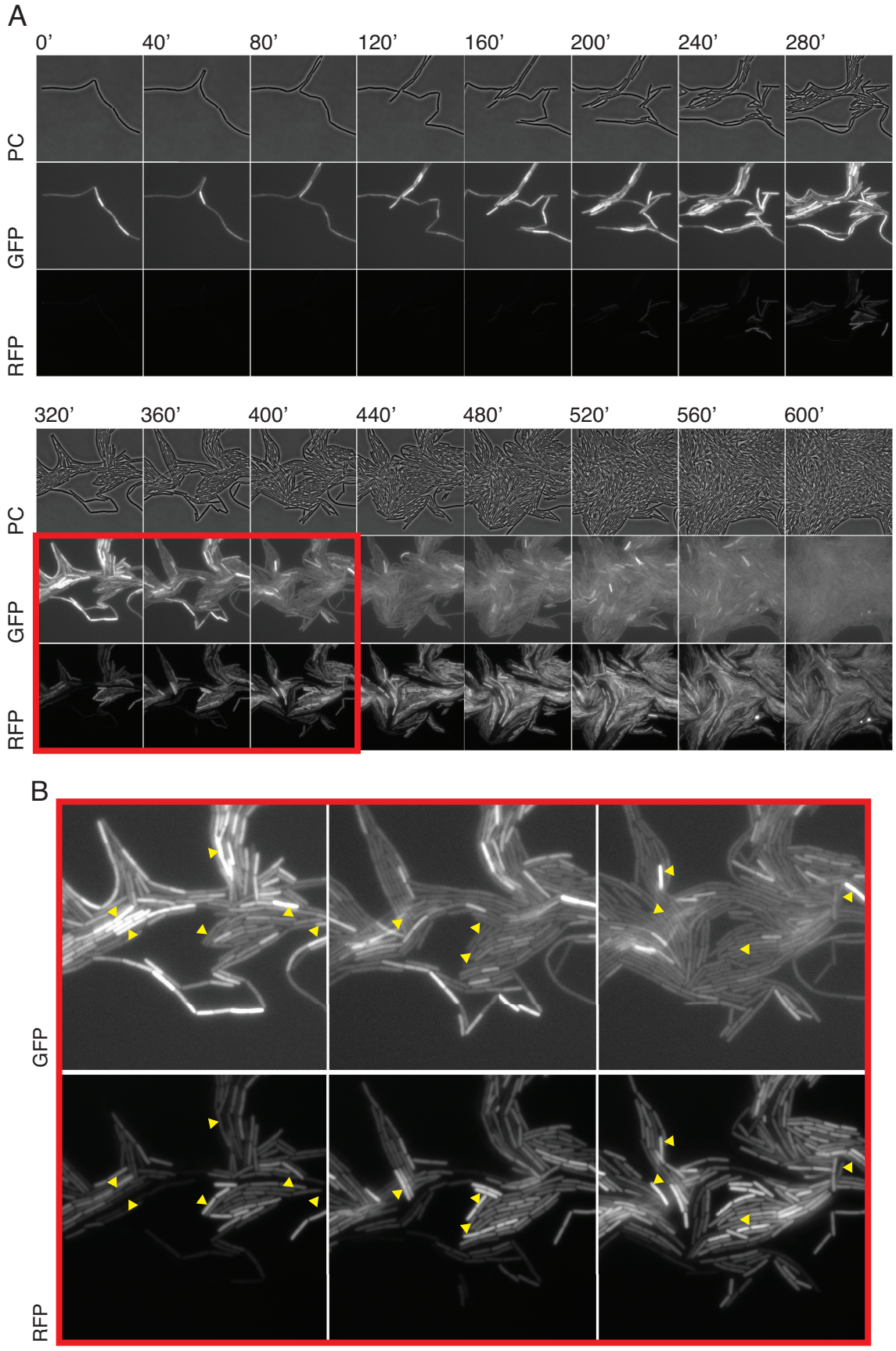

Figure 5.2: (See caption on following page)

Figure 5.2: (See figure on previous page) Coordinated pulsing of P_{rpsD} and P_{hag} in a growing microcolony of *B. subtilis*. (A) Montage showing a growing microcolony of the strain bSS387, carrying P_{rpsD}-*gfp-ssrA* and P_{hag}-*mCherry* reporter constructs. The panel is split into two parts, each showing the following rows of frames: phase contrast (top), GFP (middle) and RFP (bottom). Each frame is spaced 40 minutes (5 frames) apart. Frames highlighted by a red square are shown magnified in panel B. (B) Magnified section of montage where RFP and GFP signal is visible at the same time. Yellow triangles identify strongly fluorescing cells in either channel, highlighting that cells do not often exhibit high fluorescence intensity for both reporters at the same time. Scale bars are 5 μm.

colony (figure 5.5A). As a result, no lineage tree has to be computed. What is tracked is essentially a single branch in a lineage tree, without the need for collecting data on every cell, only to use one lineage in the subsequent analysis. Furthermore, this means that cells exhibiting phenotypes of interest in the later time points can be targeted specifically by using this method. In order to prevent tracking the same cell twice by including it more than one lineage that is being analysed, ROIs were compared carefully before a cell was selected for tracking. Figure 5.6B demonstrates the success of this technique, which shows few completely overlapping lines, which would be indicative of two closely related cells chosen at a late stage of the colony. Though purely by mathematical fact, this issue is unavoidable at earlier stages of the microcolony, as the experiment starts at a stage when only 1–4 cells are present. Cell lengths were also recorded, and the rate of length increase was used as an indicator of growth rate and to automatically detect cell division. This information allowed the duration and growth rate of each cell cycle to be extracted, as is demonstrated in figure 5.5B.

It was immediately obvious that the GFP-SsrA fluorescence intensity of single cells exhibits a pulsing dynamic that adheres closely to the timing of the cell cycle. As this was not clearly distinguishable from qualitative assessments of time lapses, this was an unexpected finding, and will be addressed in section 5.5.1.

The previously established succession of pulses, where P_{hag} follows P_{rpsD} could also be found in single cells over time meaning that the phenomenon is not just limited to the level of the microcolony. However, cells reporting activity from both promoters exist concurrently in liquid culture, which contradicts the observation in microcolonies, where the two reporters are not active at the same time. On the other hand, it can be noted that just like in a liquid culture, cells with high activity for both promoters are not frequently found (figure 5.6C).

Taking into account growth rate data and fluorescence intensity, it is possible to show that during cell cycles with high σ^{D} activity, cells grow slower (figure 5.7A). This relationship appears to be expected, as a slowdown in growth rate may be a result of decreased nutrient levels, which is a contributing environmental cue for motility

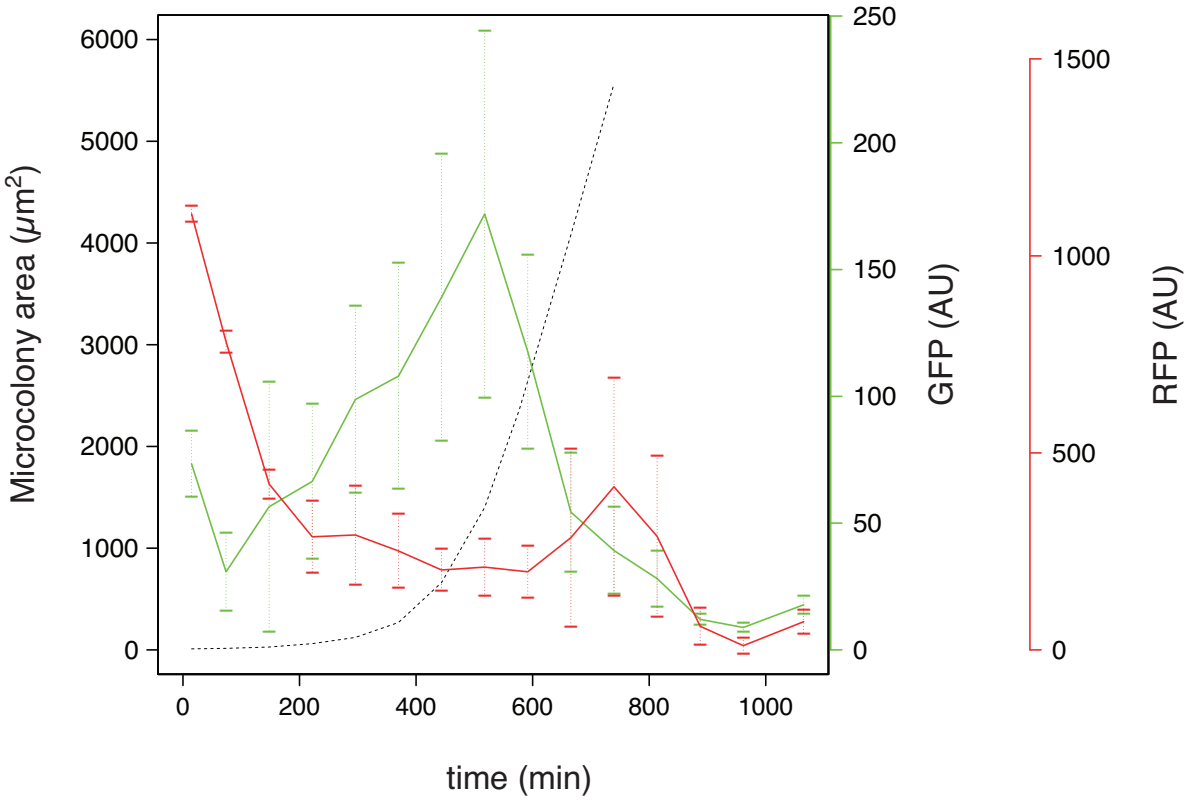

Figure 5.3: Quantification of a pulsing microcolony of *B. subtilis* in strain bSS387. Fluorescence intensities of cells sampled randomly every 5 frames from a time lapse movie of a microcolony growing on a CH agarose pad. The peak in GFP fluorescence (green line) can be observed here at 525 min, and a subsequent mCherry (red line) peak at 750 min. The initial peak in mCherry fluorescence (0–200 min) is likely to be a result of σ^{D} activity from culture conditions prior to spotting the sample on a slide. Error bars signify one standard deviation. A dotted black line represents an approximation of a growth curve based on the total area of pixels occupied by the microcolony. $n > 50$ per time point (or the whole colony at early time points where the population did not exceed 50)

development. Conversely, the relationship between growth rate and mean GFP per cell cycle shows that only faster growing cells enter the higher ranges of GFP fluorescence observed (figure 5.7B). Due to the link between growth rate and ribosome production, the notion that cells with high GFP fluorescence from the P_{rpsD}-*gfp-ssrA* reporter grow faster indirectly supports the claim that the GFP signal is a property of the P_{rpsD} promoter and not the dynamics of degradation by ClpXP.

Examining only the intensity levels and not taking into account the timing of events, it appears that the strength of GFP signal at any time during the lifetime of the cell is not a predictor of the strength of the subsequent RFP intensity peak. Figure 5.6D shows that the strength of intensity peaks from the two reporters (on a per cycle basis) do not correlate strongly.

5.5 Reciprocal relationship between P_{rpsD} and P_{hag} in Δ*codY* and Δ*pnpA* backgrounds

A previous chapter in this thesis investigated the impact of deleting of a set of possible regulators on the apparent reciprocal relationship between the reporters P_{rpsD} and P_{hag} (chapter 4). As the reciprocal relationship between the two promoters could be observed on the level of a microcolony (figure 5.3)), it was investigated whether it could be observed in two deletion backgrounds which had been previously demonstrated to lack the reciprocity. Growing Δ*codY* and Δ*pnpA* background strains as microcolonies showed some effects that were reflected by the behaviour of the same strains in liquid cultures.

A Δ*codY* strain was observed to lack any P_{rpsD}-*gfp-ssrA* activity when grown in a liquid culuture, but such a phenotype was observable when the strain grew as a microcolony (figure 5.4A). The intensity of the GFP signal in a Δ*codY* strain was however noticeable lower, which at least marks a change of P_{rpsD}-*gfp-ssrA* activity in the same direction as suggested by experiments done in liquid culture. Furthermore, the succession of pulses (P_{rpsD} before P_{hag}) remains the same in a Δ*codY* background.

In a Δ*pnpA* strain the increase in P_{rpsD}-*gfp-ssrA* activity is still present in microcolony experiments (figure 5.4B), but pulsing activity from the two constructs appears to be prolonged and overlapping somewhat. It also appears that the growth conditions in a microcolony allowed the repression of motility by the Δ*pnpA* background to be reversed in a subpopulation, which gave rise to a bimodal population during the observed P_{hag}-*mCherry* peak (figure 5.4B).

5.5.1 P_{rpsD}-gfp-ssrA pulsing with cell cycle periodicity

Quantifying fluorescence intensities over time for single cells revealed a distinctive pulsing in GFP signal which exhibited a periodicity close to the length of a cell cycle

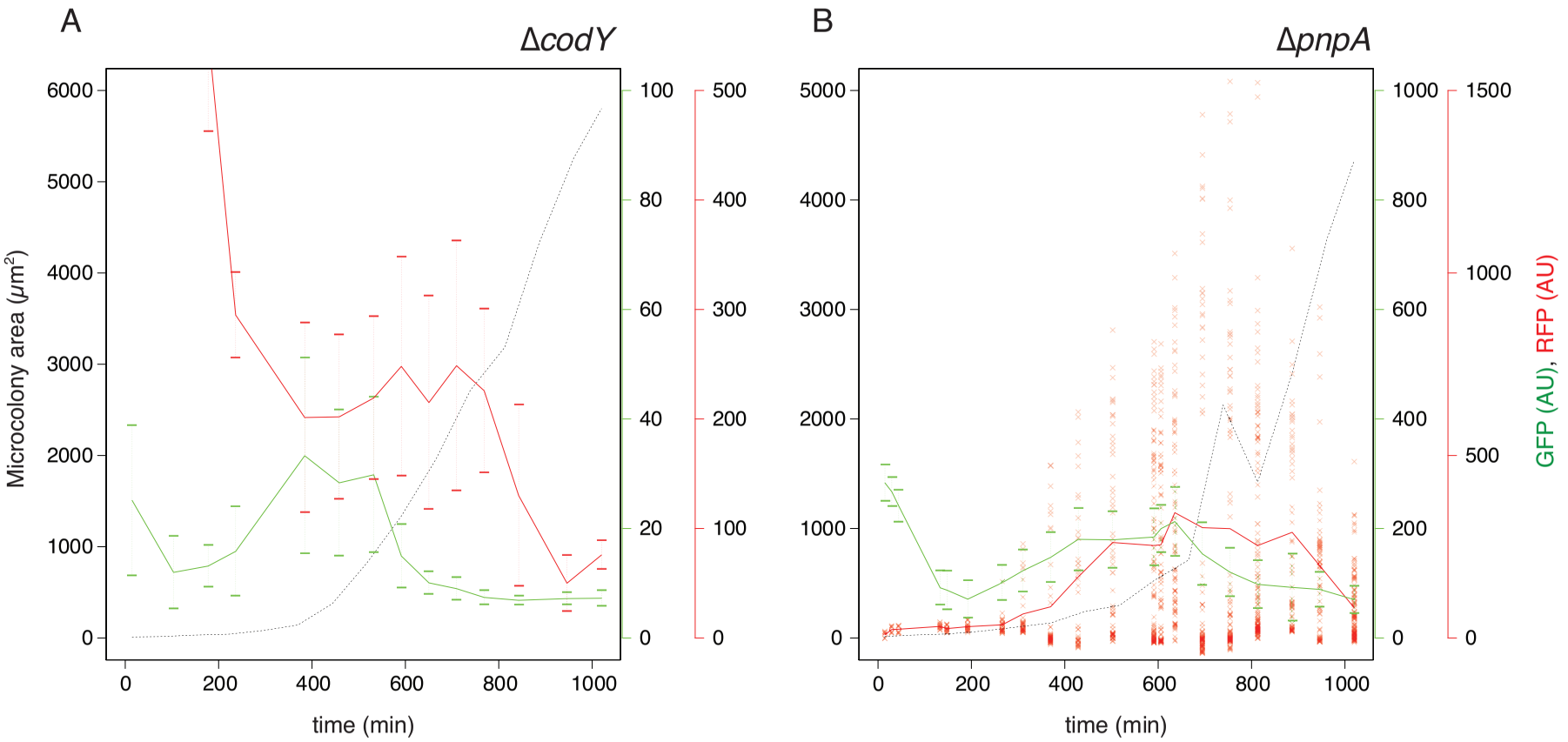

Figure 5.4: Effects of Δ*codY* and Δ*pnpA* backgrounds on pulsing behaviour of microcolonies carrying P_{rpsD}-*gfp-ssrA* and P_{hag}-*mCherry* constructs. (A) A Δ*codY* background strain (bSS395) shows a much lower level of GFP fluorescence, but a peak is present, albeit at a significantly lower level and shorter duration (∼200–600 min), while a P_{hag}-*mCherry* peak is still present. (B) A Δ*pnpA* background strain (bSS394) also replicates the reciprocal behaviour, but does not exhibit peaks with a clear separation. Because the microcolony appears to be bistable with respect to P_{hag}-*mCherry* activity, individual measurements are shown as red crosses instead of showing error bars. For comparison, a wild-type background strain is represented in figure 5.3. $n > 50$ per time point (or the whole colony at early time points where the population did not exceed 50)

(figure 5.5C).

The absolute rate of change of GFP signal (ΔGFP) was used to better visualise time points with the greatest P_{rpsD} activity. Averaged over the number of cell cycles tracked, the number of ΔGFP peaks were observed to be close to a 1:1 relationship to cell division in a strain carrying P_{rpsD}-*gfp-ssrA* (figure 5.8A).

To ascertain whether this periodicity was caused by the P_{rpsD} or SsrA component of the reporter construct, the two components were observed separately using constructs that were either inducible or non-SsrA tagged (table 5.1). No dramatic change in the ΔGFP peak:cell cycle ratio was observed when the reporter construct was changed to an inducible and ClpXP-degraded construct ($P_{hyper\text{-}spank}$-*gfp-ssrA*$_{\mathrm{DAV}}$) (figure 5.8A). However, by tracking a strain carrying an inducible and non-degraded GFP reporter construct, a similar periodicity was found (figure 5.8C). Lastly, an untagged (P_{rpsD}) construct was tracked, which also exhibited peaks with a similar periodicity (figure 5.8D). As cell cycle-level pulsing could not be attributed to a single component of the P_{rpsD}-*gfp-ssrA* reporter, it would seem that it likely arises as a consequence global expression levels in the cell.

Although this experiment has failed to yield any novel insight into a possible relationship between the cell cycle and P_{rpsD}, it has further reaffirmed that ClpXP degradation is a very constant process by showing that the SsrA tag does not affect the cell cycle pulsing. That is, if such cell cycle-related pulsing was a property of ClpXP activity, it would most likely be a contributor to heterogeneity.

In a paper where a similar reporter was used (proteolytically unstable GFP), Elowitz and Leibler (2000) created a system of three regulators repressing each other in a circular fashion. The outcome of the regulation was visualised by using GFP signal, which was tracked in a fashion similar to this chapter. A regular pulse was produced due to the artificial repression system, and control experiments showed that the reporter being induced alone did not give rise to pulses. One difference in experimental setups may provide a crucial clue to the origin of pulsing for P_{rpsD}-*gfp-ssrA*, which is the genetic location of the reporter construct. In Elowitz and Leibler's paper, the promoter is carried on a high copy number plasmid (Camps, 2010), while the construct studied in this chapter is carried on the chromosome. Plasmids replicate independently of the chromosome, and it would seem logical that they can be maintained at a more stable copy number to cell volume ratio. As a contrast, change to the ratio of a gene to cell volume is more dramatic—the gene copy number can only be 1 or 2 which is a doubling in relative terms, whereas cell volume increases continuously as the cell grows and approaches division. It is possible that this pulsing comes as a consequence of the chromosomal positioning of the reporter constructs, as all of the reporters used in this chapter have been located in the ectopic integration locus *amyE*. Experiments where the

Construct	Peaks/cycle
P_{rpsD}-*gfp-ssrA*	0.913 σ 0.096
$P_{hyper\text{-}spank}$-*gfp-ssrA*$_{\mathrm{DAV}}$	1.168 σ 0.235
$P_{hyper\text{-}spank}$-*gfp*	1.117 σ 0.147
P_{rpsD}-*gfp*	1.321 σ 0.223

Table 5.1: ΔGFP peaks in a growing microcolony. Figures show mean ΔGFP peaks per cell cycle, observed during the period of a microcolony where complete cell cycles could be recorded (peaks observed before and after the first and last divisions are not counted). σ signifies one standard deviation.

reporter is located closer *ter*, the terminus region for replication may be more preferable, as such genes exist on average in a lower copy number. Additionally, placing the construct on a high copy number plasmid may also be of interest, as mentioned previously.

5.6 Conclusions

Analysis of microcolony time lapses has yielded some new insight into the behaviour of the P_{rpsD}-*gfp-ssrA* and P_{hag}-*mCherry* reporter constructs. While it does not accurately reflect the conditions of a liquid culture, it does reproduce many of the reciprocal relationships between the two promoters. One thing that was not observed in a microcolony is the mix between P_{hag}-ON and P_{rpsD}-ON cells observable at the same time. This was an important aspect of the studies conducted in liquid culture, as one of the interesting observations made there was that motile and P_{rpsD}-OFF cells were present at the same time during early exponential phase. Cells are closely related and likely more synchronised in a microcolony than in a liquid culture, so it would be interesting in future studies to observe the behaviour of the two reporters in a more controlled chemostat system—if all cells can reach steady state, can the reciprocal relationship still be observed?

Lastly, the simplified approach adopted to cell tracking in this chapter still leaves some important questions unanswered. One particularly relevant question that was still remains would be what the fate of P_{rpsD}-ON cells is in later generations—do they grow faster, for longer, or is the effect transient? Answering such questions may not even be suitable for microcolony time lapse experiments, especially due to complications for analysis caused by the cell chaining phenotype which has shown to be so intimately related to the P_{rpsD}-ON phenotype. The emerging technique of microfluidic cell culture could be adopted to separate P_{rpsD}-ON and OFF cells into separate compartments and monitor their growth and numbers, for example (Mehling and Tay, 2014).

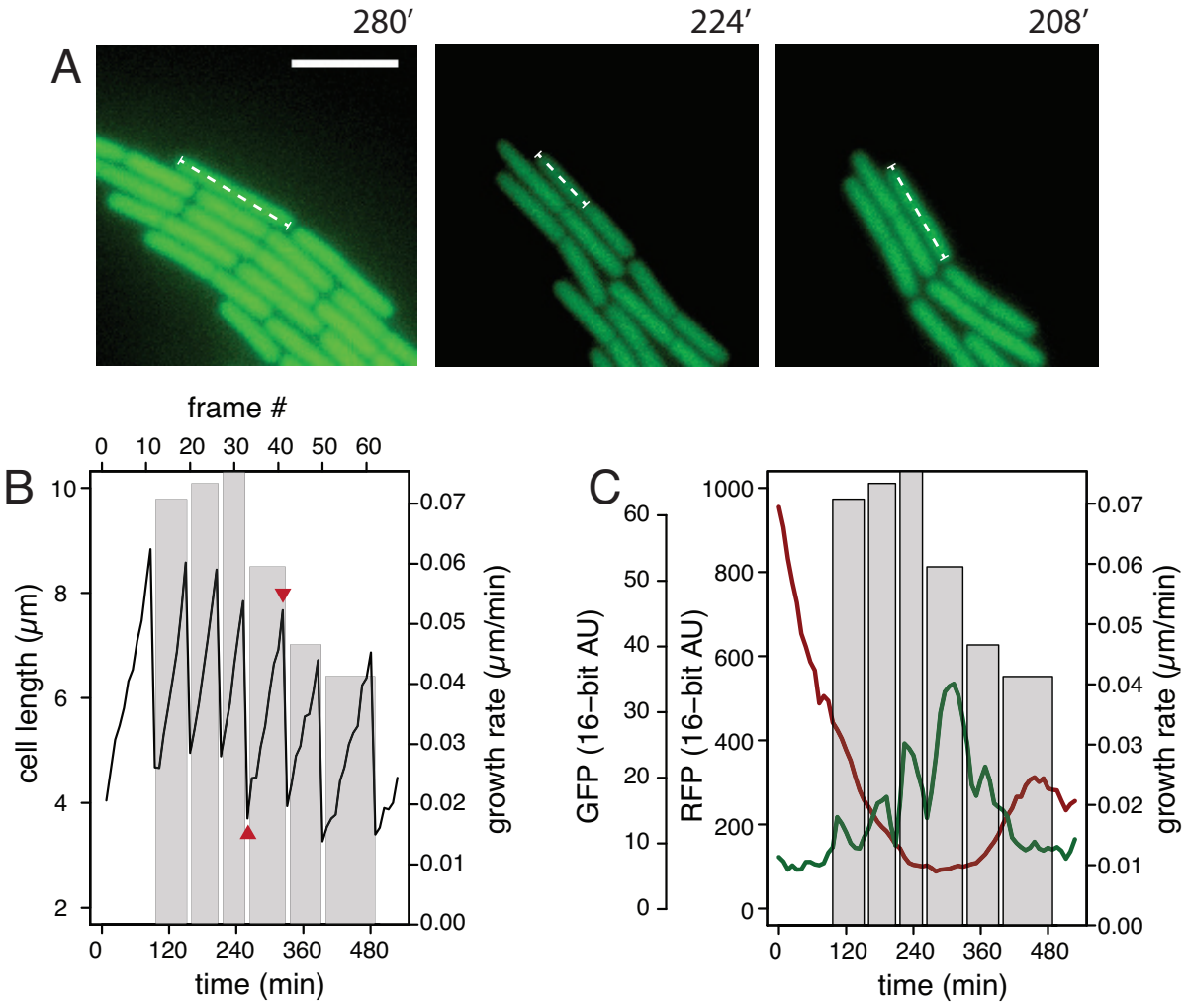

Figure 5.5: Tracking of a single cell throughout a microcolony. (A) ROI highlighting of a single cell throughout a microcolony. Single cells were tracked backwards, starting from a late time point. The dashed white shows at 208 min a cell that is about to divide, then at 224 min the cell after division, then at 280 min it is once again ready to divide. These time points represent the peaks and dips in the cell length graph shown in panel B. Scale bar is 5 μm. (B) Cell length data (black lines) is used to define the time from when a cell separates from its sibling to when it divides (events indicated by two red carets). The rate at which the cell length increases between these two points is used as the measure of growth rate, indicated by bars drawn behind each cell cycle. (C) Growth rate data overlaid with P_{hag}-*mCherry* and P_{rpsD}-*gfp-ssrA* fluorescence intensities.

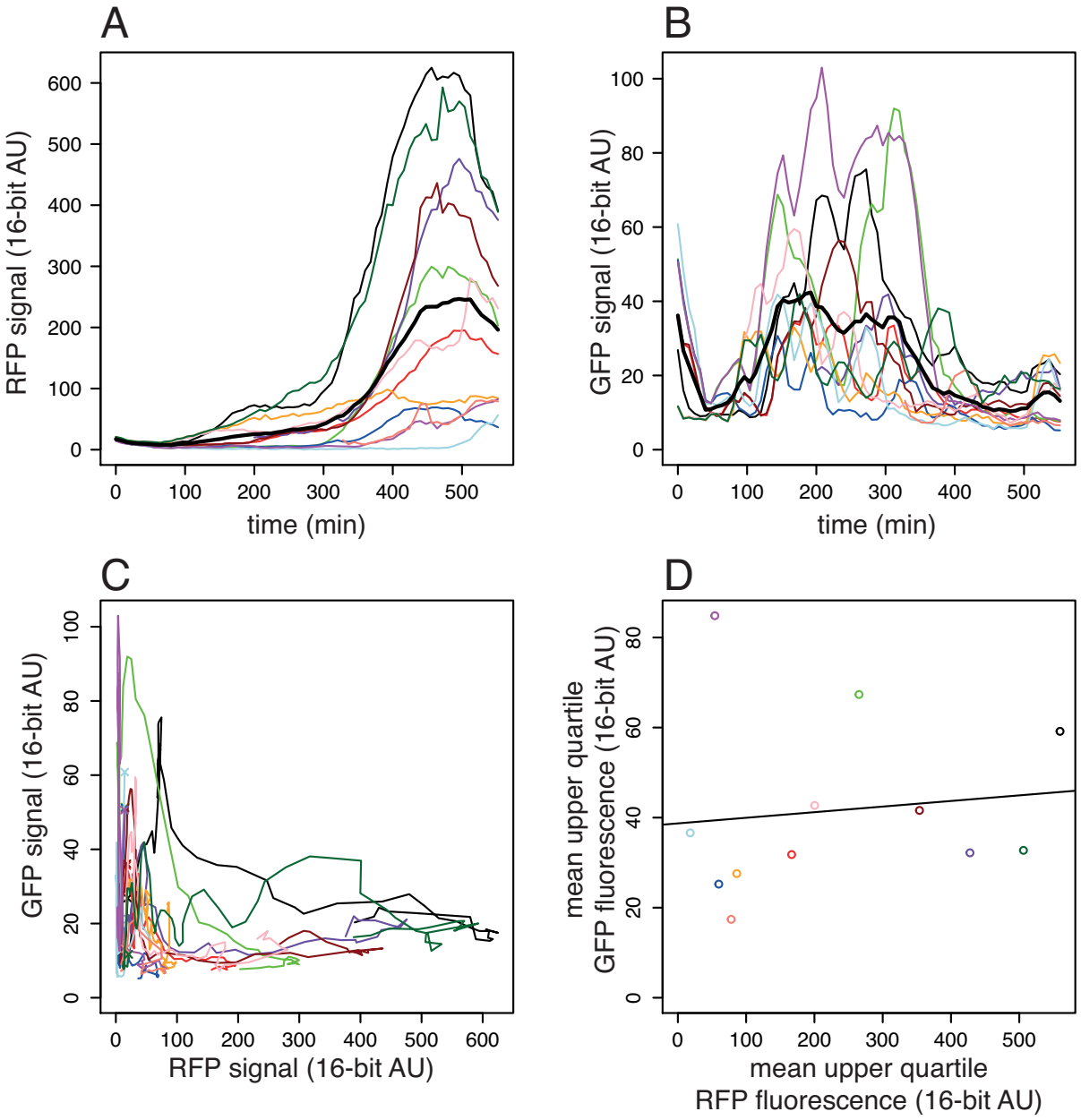

Figure 5.6: GFP and RFP measurements of several cells tracked throughout a microcolony of growing bSS387 (A–B) RFP and GFP fluorescence of 12 single cells tracked throughout a microcolony. Bold black lines indicate the mean fluorescence. (C) Change in RFP/GFP signal over time. Each cell is represented as a line, with segments of the line connecting data points indicating their RFP and GFP fluorescence intensities over time. A cross (×) indicates the first frame for the tracked cell. (D) fluorescence intensities of peaks (means of the 25% brightest values for GFP and RFP data points, per cell) for the two reporters are not strongly correlated ($r = 0.095$), suggesting that prior GFP intensity does not predict subsequent RFP intensity. 12 cell lineages were tracked from this microcolony.

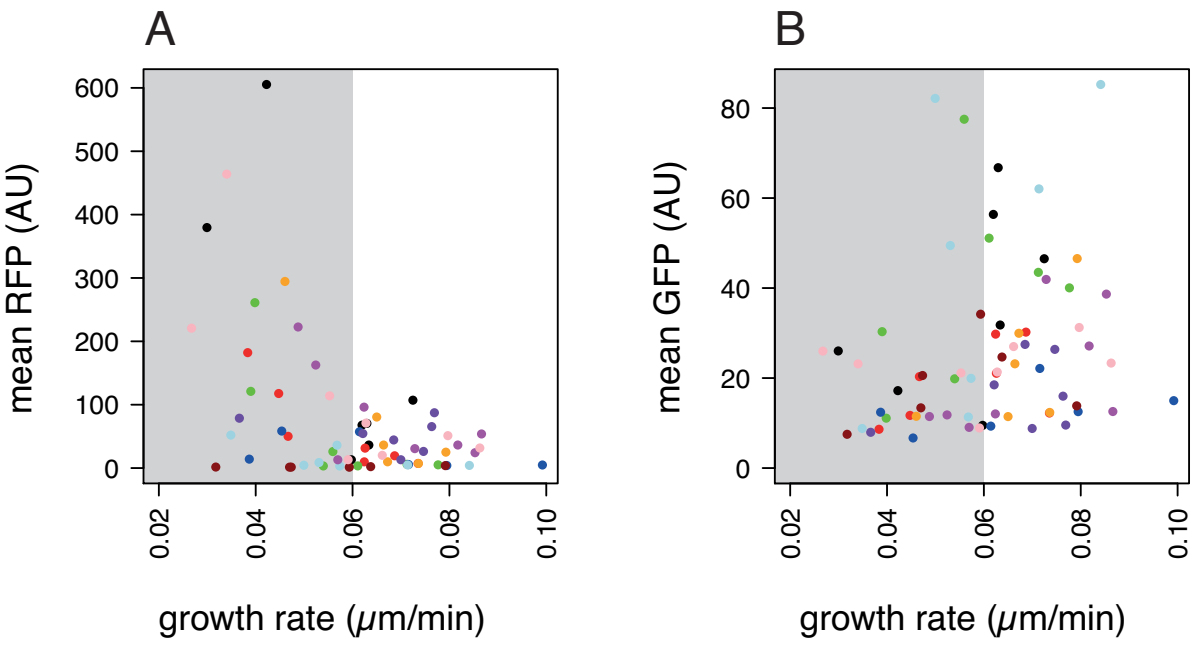

Figure 5.7: Correlating growth rate and cell cycle data to promoter activity in a microcolony. (A-B) average RFP and GFP signal (respectively) over cell cycles correlated to the corresponding growth rate. Each colour represents an individual cell in the microcolony. The shaded background shows the lower half of the range of growth rates observed. 69 cell cycles over 12 lineages were tracked.

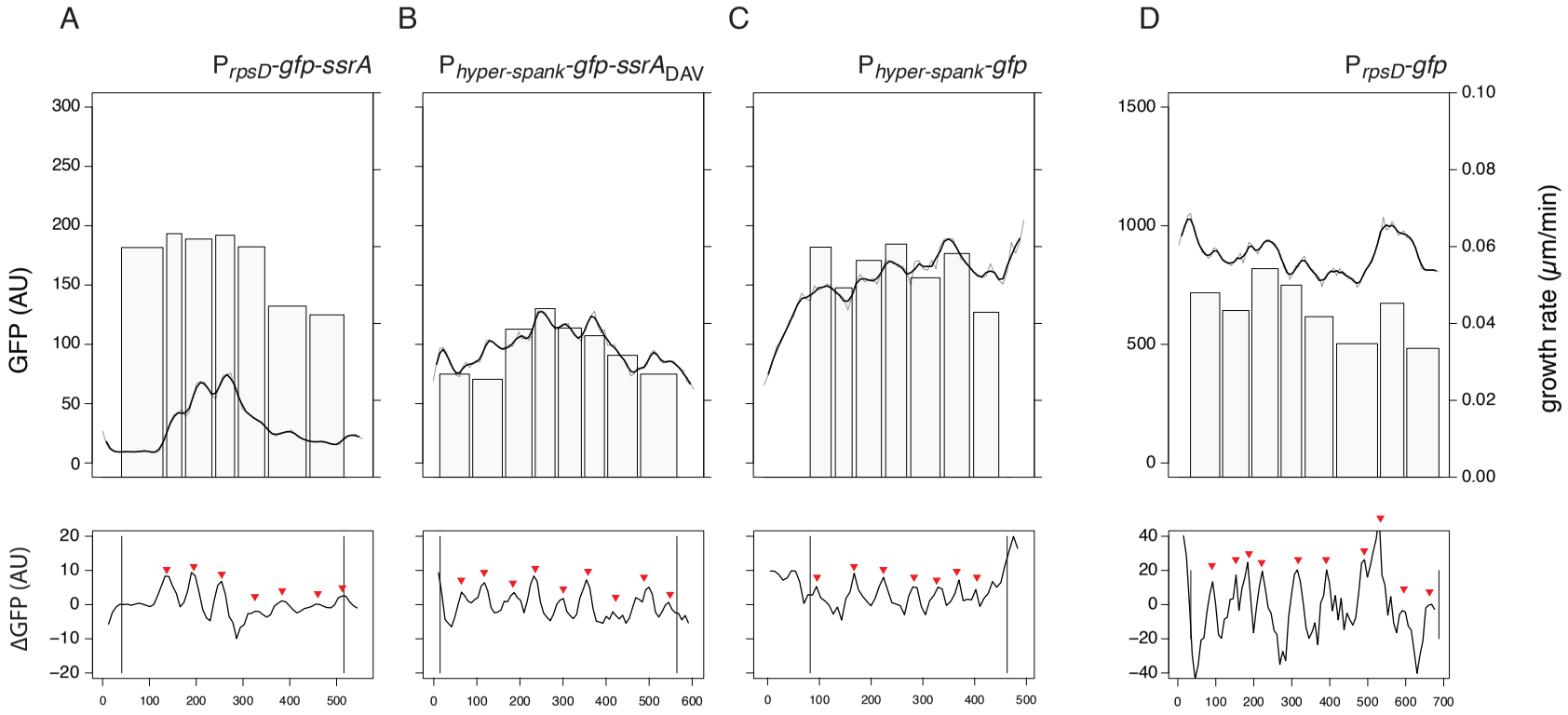

Figure 5.8: Measurements of GFP signal and ΔGFP calculations, correlated to cell cycles. The top part of each panel shows GFP signal as a grey line, and a corresponding smoothed line as a thicker black line (a moving average with a window spanning 3 measurements). Cell cycles are shown as grey bars, with widths corresponding to the duration of the cell cycle, and heights corresponding to growth rates (expressed as length increase in μm per second). The lower part of the panel shows the immediate change in GFP signal between two neighbouring measurements, plotted on the same time axis as the above plot. Peaks are indicated by red carets, and vertical black lines indicate the beginning and end of tracked cell cycles. Each panel represents a different reporter construct: (A) P_{rpsD}-*gfp-ssrA*. (B) $P_{hyper\text{-}spank}$-*gfp-ssrA*$_{\mathrm{DAV}}$, induced with 500 μM IPTG. (C) $P_{hyper\text{-}spank}$-*gfp*, induced with 10 μM IPTG. (D) P_{rpsD}-*gfp*.

Chapter 6

Role of chromosome positioning on reporter construct efficacy

6.1 Introduction

Prokaryotic genomes are often contained in a single or few circular chromosomes, and in the case of *B. subtilis*, a single chromosome of 4.2 Mb (Kunst *et al.*, 1997). The process of genome replication is best pictured as a circle with an origin or replication at the top from which the replication machinery proceeds bidirectionally, towards a terminating region at the bottom.

On the level of transcriptional units, genes often cluster together into operons where they are regulated together based on their functionality. For example, it is often the case that the genes required for completing a synthesis pathway of a metabolite can be found in the same operon. Inferences about the function of a gene can often be made from its neighbours, and can be useful as an early step in the process of identifying the function of an unknown gene (Snel *et al.*, 2000). Bias is also found with regard to the proximity to the origin of replication, as essential genes in *B. subtilis* and *E. coli* tend to cluster closer to the *oriC* region. Similarly, using codon optimisation as a measure of how highly expressed a gene is, half of the genes in the top 10% most optimised genes are located disproportionally close to the *oriC* region (Rocha, 2004, Sharp and Li, 1987). rRNA operons are highly expressed transcriptional units, and their transcripts account for upwards of 95% of RNA content in cells (Sonenshein *et al.*, 2002, p. 313–22). Keeping in line with the positioning of other highly expressed genes, 70% of identified rRNA operons in 68 species are reported by Rocha (2004) to lie near the origin. In *B. subtilis*, all 10 rRNA operons are located on the *oriC*-proximal half of the chromosome, out of which 8 are within 45° of *oriC* on the leading strand side (Sonenshein *et al.*, 1993, p. 686-8).

Another aspect of locus organisation within a chromosome is the orientation of the gene, that is, whether it is encoded on the leading or lagging strand of DNA replication. Analysis of the full *B. subtilis* genome sequence shows that the majority of genes on the right side of the chromosome are oriented in a clockwise direction, and vice-versa on the

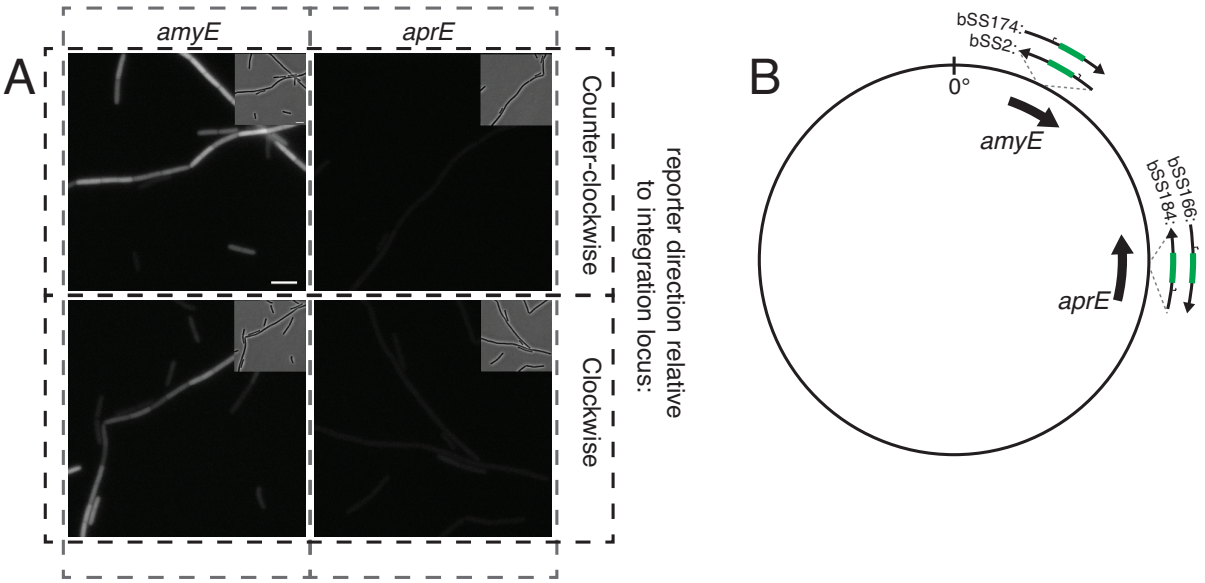

Figure 6.1: Expression of the same reporter construct differs between two commonly used integration loci. (A) GFP microscopy images of strains bSS2, bSS166, bSS174, bSS184 (left to right, top to bottom) carry identical P_{rpsD}-*gfp-ssrA* constructs and differ only in integration loci and their direction of transcription relative to the integration locus. No expression can be observed when the reporters are moved to the *aprE* locus. Phase contrast visible as insets, scale bar is 5 μm. (B) Diagram of the chromosome showing how the directions of transcription of reporters relate the direction of integration loci.

left side. In other words, genes are biased to be on the non-lagging strand of replication (Kunst *et al.*, 1997).

Within the *Bacillus* community, many standard cloning vectors exist for the purpose of introducing a genetic construct at an ectopic location on the chromosome. A common target that is non-detrimental to the health of the organism, is the gene encoding α-amylase (*amyE*), partly due to the ease of testing for Δ*amyE* by means of starch indicator plates and iodine staining. In many cases, there may be a need to use an additional integration site, where *aprE* (subtilisin E) or *lacA* (β-galactosidase) vectors are also used (Morimoto *et al.*, 2002, Härtl *et al.*, 2001). As a consequence of their locations in relation to the *ori* and *ter* sites of their chromosomes (28°, 94°, 299°, respectively), ectopic integration of genetic reporter constructs to the loci may result in a gene dosage that is inconsistent with the original context of the promoter under investigation.

Using a P_{rpsD}-*gfp-ssrA* construct with a less efficiently folding GFP variant, it was found that this construct was sensitive to the choice of integration locus (figure 6.1). The possibility that external read-through transcription would be a cause of the observation was eliminated by inverting the insert within the integration locus, and it was concluded that difference in average gene dosage of *amyE* compared to that of *aprE* was the cause of the difference in phenotypes.

6.1.1 Construction of a transposon-delivered reporter construct

The effect of locus positioning on the evidently sensitive P_{rpsD}-*gfp-ssrA* reporter construct raised the question of whether the differences in expression and noise could be quantified and mapped across the chromosome. To create a non-biased tool for surveying gene expression noise and strength across the *B. subtilis* chromosome, a transposon-based approach was adopted. The Mariner transposon system for *B. subtilis* is already widely used, and was chosen for this purpose (Le Breton *et al.*, 2006). The transposon is carried on a plasmid which replicates at 30 °C but is unstable higher temperatures, which together with selection pressure from kanamycin forces integration of the kanamycin-resistance cassette by the plasmid-encoded transposase (figure 6.2). An alternative and undesired single-crossover event is screened against after colony growth by additional erythromycin resistance, conferred by a cassette located outside the transposon.

To measure both gene expression strength and noise, β-Galactosidase and GFP were chosen as reporters for the two purposes, respectively. Single-cell measurements of GFP levels in a population can also be used to measure gene expression in addition to noise, and would therefore also serve to validate β-Gal measurements. To ensure minimal influence from transcriptional activity at integration sites, a transcriptionally isolated reporter construct was assembled. Construction was performed stepwise using two convergent terminator pairs from the *B. subtilis* chromosome, identified using the WebGeSTer database (Mitra *et al.*, 2010). The IPTG-inducible P_{spac} and the constitutively active (σ^A-dependent) P_{veg} promoters were chosen to drive expression of *lacZ* and *gfp*, respectively. To further ensure that the system is free of interference, the promoters were placed divergently, thus ensuring that upstream transcription does not proceed into a downstream reporter gene, which would have been a concern if the constructs were placed on the same strand. The original pMarB plasmid was modified to accommodate a unique restriction site within its transposon region, which originally consisted of two inverted repeats and a kanamycin resistance cassette. After inserting the reporter constructs and flanking terminator pairs into the modified pMarB, the resulting 13 kbp plasmid showed a significant reduction in successful transposants in the range of one order of magnitude (see table 6.1). The reason is likely to be the addition of material from the *B. subtilis* chromosome, provided by the two terminator fragments and P_{veg}. These regions of homology are sufficient for integration of the construct by natural competence rather than by means of transposon insertion, which would lead to a single crossover event (Khasanov *et al.*, 1992).

The culture of the strain bSS353 carrying pSS125 with the most favourable transposon insertion rate from table 6.1 was spread out on X-gal indicator plates to give an overview of β-Gal activities of individual colonies. A range of colonies with varying

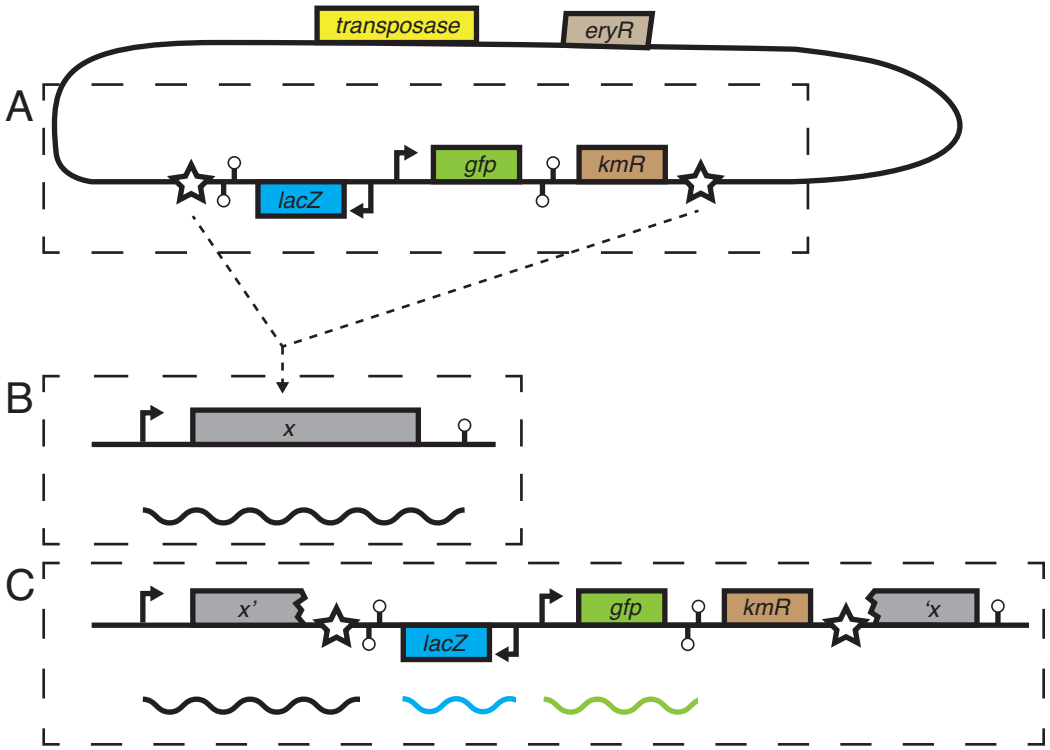

Figure 6.2: Overview of transposon-borne *lacZ/gfp* reporter construct. (A) The reporters *lacZ* and *gfp* (indicated as boxes) are located on the delivery plasmid within the transposon (bounded by inverted repeats, symbolised by stars), and are driven divergently from two different promoters, P_{spac} and P_{veg}, respectively (indicated by bent arrows). Convergent terminator pairs (indicated by lines with round ends at the end of genes) protect the construct from read-through from the chromosome. (B) A hypothetical gene (x) on the *B. subtilis* chromosome prior to insertion. mRNA is transcribed from the entire length of the operon. (C) The hypothetical gene x interrupted by insertion of the reporter construct. The transcript from the x promoter is interrupted by a terminator introduced by the transposon, mRNA from the reporter genes can be transcribed without interference from gene x.

	kan^R:erm^R ratio		
Plasmid	low	high	mean
pMarB ($n = 6$)	14:1	45:1	29.1:1
pSS125 ($n = 4$)	0.9:1	1.4:1	1.2:1

Table 6.1: Transposon insertion efficiency of reporter payload-carrying plasmid pSS125 compared to its parent plasmid, pMarB.

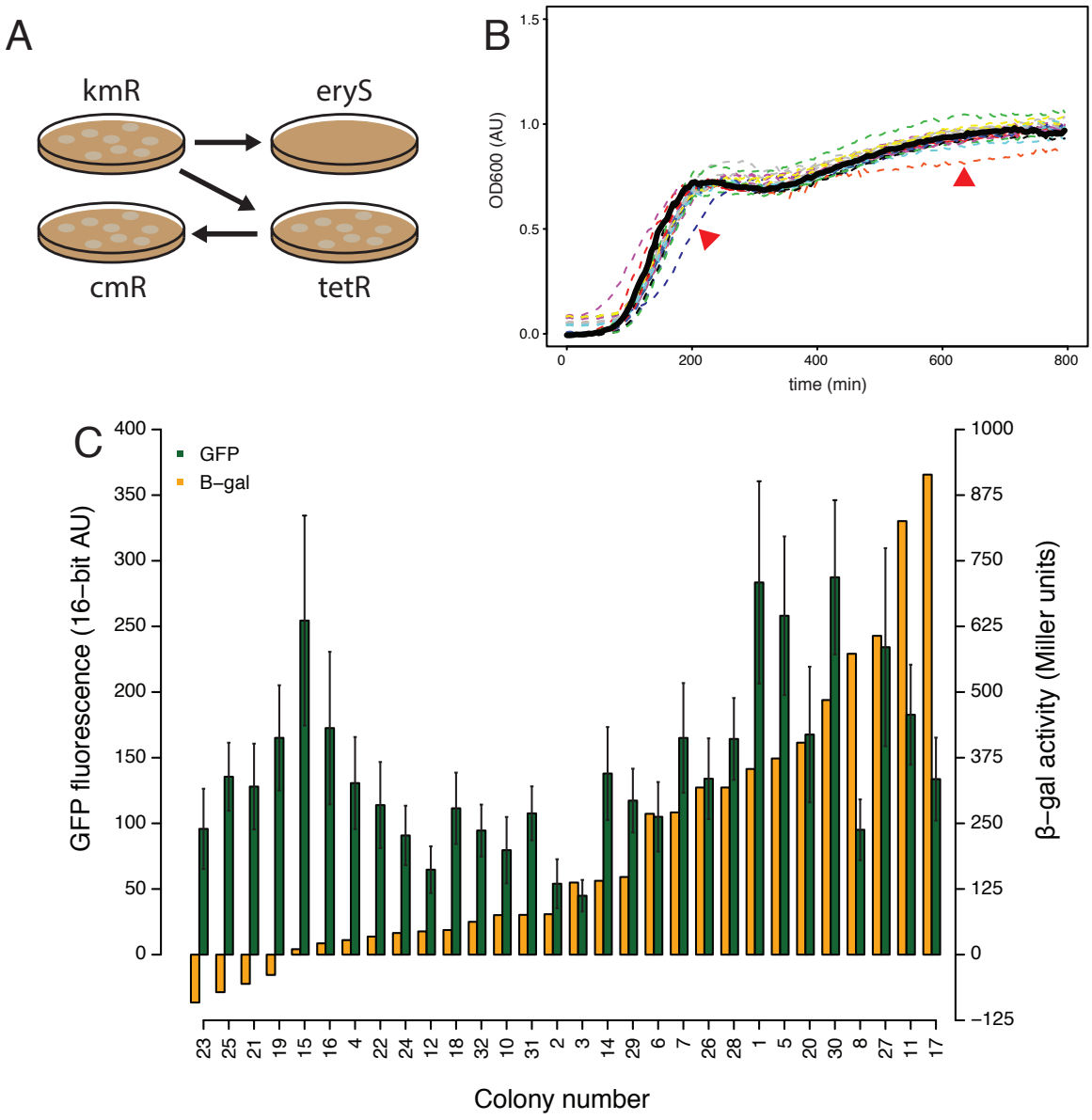

Figure 6.3: Selection of *lacZ/gfp* transposants for further study. (A) Colonies are screened for the appropriate resistance markers, corresponding to the transposon (kanamycin), *lacA* deletion background (tetracycline), presence of *lacI* (chloramphenicol), and absence of the whole pSS125 plasmid (erythromycin). (B) 32 colonies were assayed for normal growth. Two colonies stood out as growing slower (indicated by red triangles), and were discarded. (C) GFP and β-Galactosidase activities were assayed (in independent experiments). The correlation between GFP signal and β-Gal activity is $r = 0.37$.

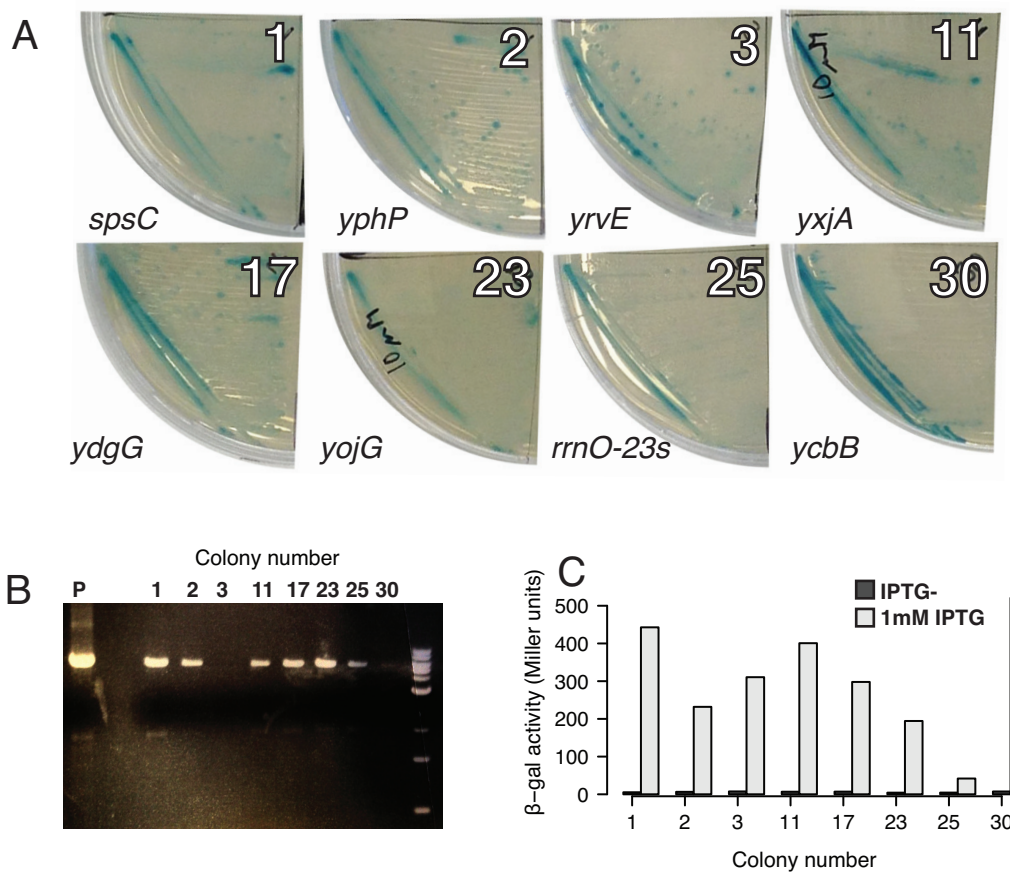

Figure 6.4: Verification of transposant colonies. 8 Candidate colonies chosen following an initial screen of GFP and β-Galactosidase activities, choices were made to give an even spread of values, which was expected to correlate to an even spread of insertion sites around the chromosome. (A) Colonies struck on X-gal plates with 10 μM IPTG (insertion sites indicated next to each plate) (B) Verification of intact transposons by PCR. pSS125 used as positive control (C) measurement of background β-Gal activity to control for read-through from neighbouring genes at the insertion site. Low activity of uninduced constructs indicate that no extraneous transcription is present and convergent terminators are working.

Colony	Locus name(s)	Genetic map	(predicted) gene function
1	*spsC*	332°	spore coat polysaccharide synthesis
2	*yphP*	196°	unknown; similar to unknown proteins from *B. subtilis*
3	*yrvE*	241°	unknown; similar to single-strand DNA-specific exonuclease
11	*yxjA*	342°	unknown; similar to pyrimidine nucleoside transport
17	*ydgG*	52°	unknown; similar to transcriptional regulator (MarR family)
23	*yojG*/*yojF*	181°	unknown
25	*rrnO*-23S	1°	ribosomal RNA-23S
30	*yckB*	31°	unknown; similar to amino acid ABC transporter (binding protein)

Table 6.2: Integration loci of strains carrying *lacZ/gfp* reporter transposon

intensity of blue colour was tested for the appropriate resistances (figure 6.3A). To ensure that the insertion of the transposon was as unintrusive as possible, screening was specifically done with preference for colonies that appeared unaffected by the transposon, with the assumption that such insertions have taken place in largely dispensable genes. 32 transposants were first chosen by the criteria that their colonies should be of a consistent [healthy] size, which serves to avoid selection of transposants with gene interruptions that causes a growth defect. To further eliminate colonies with a possible harmful insertion, the candidates were grown on a 96-well plate to assay growth, where two further colonies were eliminated (figure 6.3B). Finally, the remaining 30 colonies were assayed for GFP and β-Gal activities in independent experiments (figure 6.3C). For further study, 8 colonies were chosen based on their GFP and β-Gal activities; the two highest and 2 lowest readings for each assay, which was expected to be a rough predictor for the position of the transposon integration locus on the chromosome.

The 8 colonies were back-crossed into the original host strain bSS353 to ensure that only a single copy of the transposon existed on the chromosome, then tested for integrity of the transposon by PCR (figure 6.4B). The insertion sites of the 8 colonies were identified by arbitrary PCR that amplified DNA flanking either side of the transposon (see table 6.2 for locus names). Finally, to verify that the terminator pairs of the transposon were working, β-Gal activity was assayed with and without the presence of IPTG, showing that no production of *lacZ* was present under uninduced conditions (figure 6.4C). This should ultimately be assayed using a more specific technique such as Southern blotting, but as a convenient control it indicates that the terminator constructs are working properly.

6.2 Survey of expression and noise levels at different positions of the chromosome

To measure expression levels and noise across the *B. subtilis* chromosome, candidate transposants were cultured to a mid-exponential growth phase and samples were collected for both GFP and β-Gal assays. Previous attempts to culture the transposants separately for the two assays resulted in a low degree of correlation between GFP and β-Gal signal (figure 6.3C), showing that performing the assays on samples from the same culture was an important consideration.

Where samples were taken from the same culture for the purposes of both assays, the activity of both reporters correlated strongly (figure 6.5). One exception was the candidate transposant colony where the reporter construct was inserted in a rRNA gene (colony #25), where the β-Gal measurements did not correspond to the strong GFP levels. This may have been due to the *lacZ* construct facing against the direction of both

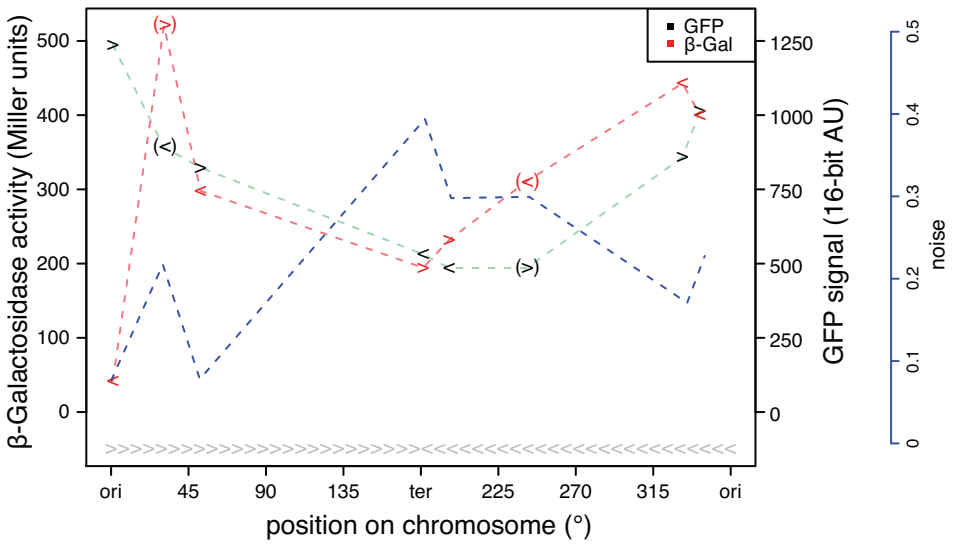

Figure 6.5: GFP and β-Galactosidase activity at 8 sites around the *B. subtilis* chromosome. Black arrowheads indicate mean GFP measurements of a sample of cells, and red arrowheads indicate a β-Gal measurement on the same culture. The direction of the arrowhead indicates the directionality of the reporter construct in relation to the direction of chromosome replication (indicated at the bottom by grey arrowheads as a visual aid). Phenotypic noise based on the standard deviation over the mean of GFP measurements is drawn as a blue dashed line.

replication and transcription from the very strong *rrnO* promoter.

Overall, the trend for change in promoter activity appears to increase with distance from the *ter* region, manifesting itself as a dip in promoter activity in the middle of the GFP and β-Gal graphs in figure 6.5. This is likely due to the overall lower copy number of loci on the *B. subtilis* chromosome near *ter*. Phenotypic noise measurements (defined as the standard deviation divided by the mean) were possible to calculate for the GFP assay, due to measurements being made on the single-cell level. Noise in GFP signal appears to increase with proximity to the *ter* region, which may suggest that the finite-number effect affects this reporter system (Kærn *et al.*, 2005).

This tool needs further validation, and first of all it would be of great interest to make a marker frequency analysis. This can be done by Q-PCR, and is a frequently used tool in studies on replication initiation, where an *ori:ter* ratio can be used a measure of over-initiation (Scholefield *et al.*, 2011).

Some similar tools based on the Tn *YLB-1* Mariner transposon have been developed recently, where the transposon instead carries an outward-facing promoter, or a promoterless copy of *lacZ* (Pozsgai *et al.*, 2011, Mulder and Schumann, 2013). However, these tools report expression in the context of the genes where the transposon is inserted, as opposed to targeting more specifically the positioning and topology of the insertion site neighbourhood. It is hoped that by taking this screen to a larger scale,

exceptions to the general trend visualised in figure 6.5 can be found, which would be good starting points for future investigations on chromosome topology and epigenetics.

Finally, relating back the observation that sparked interest in this experiment, it is evident that a difference attributable to chromosome positioning, as has been demonstrated using two widely used reporters. Perhaps another important takeaway result is also that if one is to study heterogeneity in gene expression, making the effort to find a good alternative locus of integration close to the native position of the promoter of interest could be worth considering. Though just as importantly, placement of the construct at locus with higher average gene dosage may be done to increase expression levels, which in the case of P_{rpsD}-*gfp-ssrA* appeared to be necessary to overcome degradation of the reporter. Moreover, since *amyE* is found in a part of the chromosome with low noise, these results show that the placement of the reporter on the chromosome has not contributed undue amounts of noise, which may give rise to artificial heterogeneity. Either way, finding a largely dispensable gene for such a purpose in the appropriate neighborhood should be an easy task, as finding out which are the essential and non-essential genes is no area that has lacked for attention from researchers over the years (Commichau *et al.*, 2013).

Chapter 7

Final Discussion

The main aim of this thesis was to establish whether there existed a difference in ribosome production in motile *B. subtilis* when compared to the non-motile subpopulation. Several hurdles had to be overcome to study this hypothesis.

First, the cell chaining phenotype of non-motile *B. subtilis* posed a problem when using conventional means of analysing microscopy images of cells carrying promoter-fluorophore reporter constructs. The heterogeneity of the fluorescence signals observed was also a complicating factor here, as the signal peaks themselves could not be reliably used to identify cells. Here, the approach of using nucleoid staining was adopted, which could be applied homogeneously across a sample, which facilitated analysis of samples in a semi-automated fashion. Although incompatible with the reporter constructs used in this thesis, the problem was also tackled in an alternative way by using membrane staining. This resulted in the development and testing of a method for the ImageJ plugin ObjectJ, and it is hoped that it will be of help to others in the field. This method is not only limited to intensity measurements of cytoplasmic fluorophors, but also for cell lengths, which is an important indicator in studies on cell division (Rico *et al.*, 2010).

Secondly, in order to avoid having to make time series measurements to detect changes in a very strong GFP signal, the reporter was made proteolytically unstable so that any observed signal would reflect current promoter activity. Similar observations have been made recently, where rRNA promoter activity of cells could be observed to follow similar patterns with regard to nutrient availability and the subsequent sporulation (Rosenberg *et al.*, 2012). While this revealed a striking degree of heterogeneity, it also added the concern that the addition of the SsrA tag to the reporter would have an impact on heterogeneity. Controlling for this possibility proved difficult, as no other tested promoter could reach the levels of activity exhibited by P_{rpsD}, and signal would thus be quenched to a homogeneous level by the fusion of the SsrA tag to the reporter. Further support for homogeneous substrate processing by ClpXP is made by observing constructs where the SsrA tag had been mutated to decrease degradation

rates. Such constructs were homogeneously fluorescent for all promoters tested.

Finding regulators underlying this reciprocal relationship between ribosome synthesis and motility proved difficult. Most regulators known to affect motility did not have any impact on the relationship. Δ*pnpA* did affect both reporters, but it was subsequently shown that the effect of this deletion was non-specific, and that it mostly affected the rate of degradation of substrate by ClpXP. One regulator that did affect both phenotypes was *codY*, which upon deletion increased motility and decreased the prevalence of P_{rpsD}-ON cells, but its deletion does however have a very global impact on nutrient regulation. Due to the difficulty of identifying a regulator responsible for both phenotypes on a less global scale, it is proposed that extrinsic noise in nutrient uptake and processing is an underlying upstream cause of the differentiation. The lack of success when screening for a possible regulator also reflected the theory of a more global upstream cause, such as nutrient availability underlying the phenotypes. A state of lower nutrient availability could be a plausible trigger for motility development, and exclusion from the levels of resources required for P_{rpsD} to overcome SsrA-mediated GFP degradation. Investigating such a hypothesis could be a direction of future work, where separated populations could have their metabolomes compared, which may even be possible to achieve on a single-cell level with the right apparatus (Zenobi, 2013). If the hypothesis of alternative metabolic pathway utilisation is worth pursuing, it would also be interesting to identify possible metabolic pathways to close off in order to elicit or inhibit the $P_{rpsD}{}^{\mathrm{HIGH}}$ phenotype. The role of CodY in metabolic control together with CcpA encourages this idea, and it is very possible that this was the case in the Δ*codY* strain tested (Wünsche *et al.*, 2012, Sonenshein, 2007).

Because SsrA-tagging only proved to work with one promoter (P_{rpsD}), the question whether the relationship extends to ribosomal subunit synthesis in general can be questioned. Earlier attempts to create fusions with other ribosomal subunit promoters (P_{rpsJ}, P_{rpsF}, P_{rpsO}) identified at the onset of the project[1], showed that P_{rpsD} is likely the only ribosomal promoter that is able to produce GFP at rates high enough to overcome SsrA tagging. Ribosomal RNA promoters may be desirable to test, as their strength is similar to their protein subunit counterparts, but were not in scope for this study as the original transcriptional profiling data carried information on open reading frames. If the P_{rpsD}-ON phenotype can be extended to ribosome synthesis in general, perhaps the P_{hag}-ON/P_{rpsD}-OFF phenotype could be applied to other bistable cell fates related to nutrient starvation. Naturally, the other cell fates are not as readily observable as motility, which is most likely why this relationship was identified in the first place—perhaps microcolony time lapse experiments reveal such a relationship in the future? Chai *et al.* (2011) investigate a similar system, where metabolism decides

[1] S. Syvertsson, MRes project (2010)

whether a cell should enter a biofilm matrix-producing cell fate, but no ties to ribosomal activity are made.

From the onset of this project until its completion, next-generation sequencing technologies have made great improvements in efficiency, and have as a consequence come within budget range of more and more institutes. RNA sequencing is now eclipsing DNA microarrays as the method of choice for transcriptome profiling, and as such it would be interesting to see if the project could go back to its beginnings and see how a motile subpopulation distinguishes itself on the transcriptional level. Furthermore, as RNA sequencing requires substantially smaller RNA samples, these methods may even be able to approach the question from the other side and profiling the transcriptome of cells exhibiting an $P_{rpsD}{}^{\mathrm{HIGH}}$ phenotype.

The use of microcolony time lapse microscopy is an important tool in the study of differentiation in bacteria, and its use in this thesis helped show that the reciprocal relationship between P_{hag} and P_{rpsD} persists on the microcolony level. It did however fail to replicate the observation of motile cells being present at the same time as when the greatest P_{rpsD} activity was observed, which was a key observation in liquid cultures. It is speculated that this could be due to cells being more synchronised in a microcolony than they are in liquid culture. By taking more care to rigorously standardise culture conditions by the use of a chemostat, liquid cultures could be observed in steady state at varying growth rates, and it would then be interesting to see if the ratios of ON/OFF subpopulations for the two promoters could be manipulated accordingly.

Furthermore, observations of changes in GFP signal in microcolonies also revealed a striking cell cycle-dependent oscillation in intensity. Such pulses are likely caused by varying gene dosage as a consequence of the chromosomal positioning of the reporter construct. Also, further supporting data suggesting that ClpXP degrades substrate with a homogeneous efficacy across the population, experiments show that the fluctuations occur independently of SsrA-tagging.

On designing some experiments, it became clear that some constructs used in this study were sensitive to the position of the integration locus on the *B. subtilis* chromosome. This posed the question of whether the gene dosage effect of a continuously replicating circular chromosome provides a predictably ramped increase of baseline expression levels the closer to *ori* a gene gets. It has already been shown that the correct positioning is crucial for a regulatory gene which relies on noise to generate a bistable outcome Chai *et al.* (2011). However, no large-scale map of overall expression levels has been produced before, and this work has developed the tools to conduct such a survey in a non-biased manner.

Appendix A

Data analysis scripts

This section contains scripts and pipelines written in R, Python, bash, or CellProfiler for the purpose of data analysis are outlined as schematic figures in the main thesis to aid interpretation (figures 2.1 and 2.2). Their full code is reproduced in this appendix, and where appropriate the schematic figure accompanying the code is referred to in their descriptions. A short description of code function is also available as comments in the code itself, indicated with the hash symbol (#).

A.1 CellProfiler pipeline

The following configuration file represents a typical CellProfiler pipeline suitable for capturing data from the GFP channel using outlines of DAPI-stained nucleoids. Each unindented line signifies the start of a new *module*, a discrete step in the analysis. In this pipeline, the modules with self-explanatory titles (e.g., "LoadImages") represent a step in the analysis pipeline.

```
1  LoadImages:[module_num:1|svn_version:\'11587\'|variable_revision_number:11|show_window:
       True|notes:\x5B\x5D]
2      File type to be loaded:individual images
3      File selection method:Text-Exact match
4      Number of images in each group?:2
5      Type the text that the excluded images have in common:Do not use
6      Analyze all subfolders within the selected folder?:None
7      Input image file location:Default Input Folder\x7CNone
8      Check image sets for missing or duplicate files?:Yes
9      Group images by metadata?:Yes
10     Exclude certain files?:No
11     Specify metadata fields to group by:Image,col,time,Strain
12     Select subfolders to analyze:
13     Image count:2
14     Text that these images have in common (case-sensitive):DAPI
15     Position of this image in each group:1
16     Extract metadata from where?:File name
17     Regular expression that finds metadata in the file name:^(?P<Strain>.*)\\.col(?P<col
           >\x5B0-9\x5D{1,2})\\.t(?P<time>\x5B0-9\x5D{1})\\.(?P<Image>\x5B0-9\x5D{1,2}).*
```

```
    Type the regular expression that finds metadata in the subfolder path:.*\x5B\\\\/\
        x5D(?P<Date>.*)\x5B\\\\/\x5D(?P<Run>.*)$
    Channel count:1
    Group the movie frames?:No
    Grouping method:Interleaved
    Number of channels per group:3
    Load the input as images or objects?:Images
    Name this loaded image:DNA
    Name this loaded object:Nuclei
    Retain outlines of loaded objects?:No
    Name the outline image:LoadedImageOutlines
    Channel number:1
    Rescale intensities?:Yes
    Text that these images have in common (case-sensitive):w2GFP
    Position of this image in each group:2
    Extract metadata from where?:File name
    Regular expression that finds metadata in the file name:^(?P<Strain>.*)\\.col(?P<col
        >\x5B0-9\x5D{1,2})\\.t(?P<time>\x5B0-9\x5D{1})\\.(?P<Image>\x5B0-9\x5D{1,2}).*
    Type the regular expression that finds metadata in the subfolder path:.*\x5B\\\\/\
        x5D(?P<Date>.*)\x5B\\\\/\x5D(?P<Run>.*)$
    Channel count:1
    Group the movie frames?:No
    Grouping method:Interleaved
    Number of channels per group:3
    Load the input as images or objects?:Images
    Name this loaded image:green
    Name this loaded object:Nuclei
    Retain outlines of loaded objects?:No
    Name the outline image:LoadedImageOutlines
    Channel number:1
    Rescale intensities?:No

ApplyThreshold:[module_num:2|svn_version:\'6746\'|variable_revision_number:5|show_window
    :True|notes:\x5B\x5D]
    Select the input image:green
    Name the output image:ThreshGreen
    Select the output image type:Grayscale
    Set pixels below or above the threshold to zero?:Below threshold
    Subtract the threshold value from the remaining pixel intensities?:Yes
    Number of pixels by which to expand the thresholding around those excluded bright
        pixels:0.0
    Select the thresholding method:Otsu Global
    Manual threshold:0.0
    Lower and upper bounds on threshold:0.000000,1.000000
    Threshold correction factor:1
    Approximate fraction of image covered by objects?:0.01
    Select the input objects:None
    Two-class or three-class thresholding?:Three classes
    Minimize the weighted variance or the entropy?:Entropy
    Assign pixels in the middle intensity class to the foreground or the background?:
        Background
    Select the measurement to threshold with:None

EnhanceOrSuppressFeatures:[module_num:3|svn_version:\'10591\'|variable_revision_number
    :2|show_window:True|notes:\x5B\x5D]
    Select the input image:DNA
    Name the output image:FilteredBlue
    Select the operation:Enhance
```

```
    Feature size:20
    Feature type:Speckles
    Range of hole sizes:1,10

IdentifyPrimaryObjects:[module_num:4|svn_version:\'10826\'|variable_revision_number:8|
    show_window:True|notes:\x5B\x5D]
    Select the input image:FilteredBlue
    Name the primary objects to be identified:Nuclei
    Typical diameter of objects, in pixel units (Min,Max):15,20
    Discard objects outside the diameter range?:No
    Try to merge too small objects with nearby larger objects?:No
    Discard objects touching the border of the image?:Yes
    Select the thresholding method:Otsu Global
    Threshold correction factor:1.5
    Lower and upper bounds on threshold:0.0,1.0
    Approximate fraction of image covered by objects?:0.01
    Method to distinguish clumped objects:Shape
    Method to draw dividing lines between clumped objects:Intensity
    Size of smoothing filter:10
    Suppress local maxima that are closer than this minimum allowed distance:7
    Speed up by using lower-resolution image to find local maxima?:Yes
    Name the outline image:PrimaryOutlines
    Fill holes in identified objects?:Yes
    Automatically calculate size of smoothing filter?:Yes
    Automatically calculate minimum allowed distance between local maxima?:Yes
    Manual threshold:0.0
    Select binary image:None
    Retain outlines of the identified objects?:Yes
    Automatically calculate the threshold using the Otsu method?:Yes
    Enter Laplacian of Gaussian threshold:0.5
    Two-class or three-class thresholding?:Three classes
    Minimize the weighted variance or the entropy?:Weighted variance
    Assign pixels in the middle intensity class to the foreground or the background?:
        Background
    Automatically calculate the size of objects for the Laplacian of Gaussian filter?:
        Yes
    Enter LoG filter diameter:5
    Handling of objects if excessive number of objects identified:Continue
    Maximum number of objects:500
    Select the measurement to threshold with:None

ExpandOrShrinkObjects:[module_num:5|svn_version:\'10830\'|variable_revision_number:1|
    show_window:True|notes:\x5B\x5D]
    Select the input objects:Nuclei
    Name the output objects:ShrunkenNuclei
    Select the operation:Shrink objects by a specified number of pixels
    Number of pixels by which to expand or shrink:1
    Fill holes in objects so that all objects shrink to a single point?:No
    Retain the outlines of the identified objects for use later in the pipeline (for
        example, in SaveImages)?:Yes
    Name the outline image:ShrunkenNucleiOutlines

MeasureObjectSizeShape:[module_num:6|svn_version:\'1\'|variable_revision_number:1|
    show_window:True|notes:\x5B\x5D]
    Select objects to measure:ShrunkenNuclei
    Calculate the Zernike features?:No

MeasureObjectIntensity:[module_num:7|svn_version:\'10816\'|variable_revision_number:3|
```

```
    show_window:True|notes:\x5B\x5D]
    Hidden:2
    Select an image to measure:ThreshGreen
    Select an image to measure:DNA
    Select objects to measure:ShrunkenNuclei

ExportToSpreadsheet:[module_num:8|svn_version:\'10880\'|variable_revision_number:7|
    show_window:True|notes:\x5B\x5D]
    Select or enter the column delimiter:Comma (",")
    Prepend the output file name to the data file names?:Yes
    Add image metadata columns to your object data file?:Yes
    Limit output to a size that is allowed in Excel?:No
    Select the columns of measurements to export?:No
    Calculate the per-image mean values for object measurements?:No
    Calculate the per-image median values for object measurements?:No
    Calculate the per-image standard deviation values for object measurements?:No
    Output file location:Default Output Folder\x7CNone
    Create a GenePattern GCT file?:No
    Select source of sample row name:Metadata
    Select the image to use as the identifier:None
    Select the metadata to use as the identifier:None
    Export all measurements, using default file names?:Yes
    Press button to select measurements to export:
    Data to export:Do not use
    Combine these object measurements with those of the previous object?:No
    File name:DATA.csv
    Use the object name for the file name?:Yes

OverlayOutlines:[module_num:9|svn_version:\'10672\'|variable_revision_number:2|
    show_window:True|notes:\x5B\x5D]
    Display outlines on a blank image?:No
    Select image on which to display outlines:DNA
    Name the output image:OrigOverlay
    Select outline display mode:Grayscale
    Select method to determine brightness of outlines:Max of image
    Width of outlines:4
    Select outlines to display:ShrunkenNucleiOutlines
    Select outline color:Red

SaveImages:[module_num:10|svn_version:\'10822\'|variable_revision_number:7|show_window:
    True|notes:\x5B\x5D]
    Select the type of image to save:Image
    Select the image to save:OrigOverlay
    Select the objects to save:None
    Select the module display window to save:None
    Select method for constructing file names:Sequential numbers
    Select image name for file prefix:DNA
    Enter file prefix:
    Do you want to add a suffix to the image file name?:No
    Text to append to the image name:
    Select file format to use:png
    Output file location:Default Output Folder\x7CNone
    Image bit depth:8
    Overwrite existing files without warning?:No
    Select how often to save:Every cycle
    Rescale the images? :No
    Save as grayscale or color image?:Grayscale
    Select colormap:bone
```

```
175     Store file and path information to the saved image?:No
176     Create subfolders in the output folder?:No
```

A.1.1 Auxiliary setup scripts

`winprep.sh` is a bash shell script that takes an input `pipeline.cp` and identifies the TCF parameter (30–46), creating three new `pipeline.cp` files with different user-specified TCF values (47–63). A `.bat` file is also created with instructions for executing the analysis pipelines using CellProfiler running in batch mode on a Windows computer (lines 52–60).

```
1  # Please update config.sh with your own paths before running!
2  . config.sh
3  cd $imgroot_mac
4  touch dapifinder.log
5  # Checks for pre-existing runs already executed before starting
6  if [ -f cp_batch.bat -o -d out1/ -o -d out2/ -o -d out3/ -o -d pipeline1/ -o -d
       pipeline2/ -o -d pipeline3/ ]; then
7      MESS="winprep.sh has already been run on "$imgroot_mac", please remove cp_batch.bat,
           as well as the out*/ and pipeline*/ folders"
8      echo $MESS
9      DATE=$(date); echo "[ "$DATE" ]: ERROR: "$MESS >> dapifinder.log
10     exit
11 fi
12 tcf=($1 $2 $3)
13 tcflen=${#tcf[*]}
14 MESS="winprep.sh started for TCF = ["$1" , "$2" , "$3"]"
15 DATE=$(date); echo "[ "$DATE" ]: "$MESS >> dapifinder.log
16 # Check that pipeline.cp exists
17 if [ ! -f pipeline.cp ]; then
18     MESS="pipeline.cp is missing from "$imgroot_mac
19     echo $MESS
20     DATE=$(date); echo "[ "$DATE" ]: ERROR: "$MESS >> dapifinder.log
21     exit
22 fi
23 # check that only three arguments are given
24 if [ "$tcflen" != "3" ]; then
25     MESS="Please enter three TCFs, like: user$ ./winprep.sh 1.0 1.1 1.2"
26     echo $MESS
27     DATE=$(date); echo "[ "$DATE" ]: ERROR: "$MESS >> dapifinder.log
28     exit
29 fi
30 # Get the line number for the threshold correction factor
31 ipo=`grep 'IdentifyPrimaryObjects' -n pipeline.cp`
32 ipoPos=`echo $ipo | sed -n "s/[:IdentifyPrimaryObjects].*//p" | wc -c`
33 ipoLine=`echo ${ipo:0:$ipoPos-1}`
34 nlines=`cat pipeline.cp | wc -l`
35 taillines=`expr $nlines - $ipoLine`
36 cf=`tail -n $taillines pipeline.cp | grep 'correction factor' -n`
37 cfPos=`echo $cf | sed -n "s/[:Threshold correction factor].*//p" | wc -c`
38 cfLine=`echo ${cf:0:$cfPos-1}`
39 tcfline=`expr $cfLine + $ipoLine`
40 # Check that the pipeline actually has a tcf to change
```

```
if [ "$tcfline" == "" ]; then
    MESS="Couldn't find any tcf to change"
    echo $MESS
    DATE=$(date); echo "[ "$DATE" ]: ERROR: "$MESS >> dapifinder.log
    exit
fi
# Insert desired tcf values into the pipeline, make
# input and output folders, and a .bat file to for
# running cell profiler under windows
touch cp_batch.bat
i=1
while [ $i -lt 4 ]; do
    echo "making pipeline number $i"
    outname="pipeline"$i".cp"
    outdir="pipeline"$i
    sed -E "$tcfline s/[0-9]{1,2}(\.)?([0-9]{0,2})?/${tcf[$i-1]}/" pipeline.cp >
        $outname
    mkdir $outdir
    mv $outname $outdir
    cpout=$imgroot"out"$i"/"
    echo "\"$cppath\" -c -r -p $imgroot$outdir/$outname -i $imgroot -o $cpout" >>
        cp_batch.bat
    mkdir "out"$i
    i=$((i + 1))
done
MESS="winprep.qc: Done. No errors reported..."
DATE=$(date); echo "[ "$DATE" ]: "$MESS >> dapifinder.log
```

`qcprep.sh` is a shell script that sets up files for quality control checking by the user. It iterates through the files created by a finished CellProfiler pipeline, takes output images—as and creates "montages" of three images joined. It uses the short Python script `makeMontage.py` to join the images (shown immediately after). The script also generates a `preferredTCF.csv` file for the user to specify their preferences for each image within a montage (image 1–3, or 0 if none are satisfactory).

```
. config.sh
shdir=$(pwd)
cd $imgroot_mac
touch dapifinder.log
# Check that qcprep hasn't been run already
if [ -d montages/ ]; then
    MESS="Output folder montages/ already exists. Has qcprep already been run?"
    echo $MESS
    DATE=$(date); echo "[ "$DATE" ]: ERROR: "$MESS >> dapifinder.log
    exit
fi
# Check that winprep and cellprofiler has been run first
if ! [ -d out1/ -a -d out2/ -a -d out3/ ]; then
    MESS="Looks like winprep.sh hasn't been run yet"
    echo $MESS
    DATE=$(date); echo "[ "$DATE" ]: ERROR: "$MESS >> dapifinder.log
    exit
fi
# Find how many files there are to go through...
```

```
pngs=(`find out*/ -maxdepth 1 -name "*.png"`)
npngs=${#pngs[@]}
let modpngs=$npngs%3
let nimages=$npngs/3
# Rudimentary check to see if all folders have same number of pngs, and to make sure folders are populated
if [ $npngs -eq 0 -o $modpngs != 0 ]; then
    MESS="Something wrong with cellProfiler output. Possible that winprep was run, but cellprofiler hasn't"
    echo $MESS
    DATE=$(date); echo "[ "$DATE" ]: ERROR: "$MESS >> dapifinder.log
    exit
fi
# Pad output data with zeroes for easier sorting
i=1
while [ $i -le 3 ]; do
    cd out$i
    . $shdir/zeropadder.sh
    cd $imgroot_mac
    i=$((i + 1))
done
# make a list of file names
cd out1/
ls *.png > orignames.txt
mv orignames.txt ../
cd $imgroot_mac
# Get labels for montages from script folder
cp $shdir/labels/*.png .
mkdir montages
echo "there are" $nimages "files to check..."
i=1
# Make montages of the three qc images for each microscopy image
while [ $i -le $nimages ]; do
    mkdir $i"/"
    origname=`sed -n "$i p" orignames.txt`
    echo $origname
    cp out1/$origname $i/1_$origname
    cp out2/$origname $i/2_$origname
    cp out3/$origname $i/3_$origname
    python $shdir/makeMontage.py one.png $i/1_$origname two.png $i/2_$origname three.png $i/3_$origname $i".png"
    mv $i".png" montages/
    rm -rf $i/
    i=$((i + 1))
done
# Cleanup
rm orignames.txt
rm one.png
rm two.png
rm three.png
echo $nimages >> nimages.txt
Rscript $shdir/makeTemplate.r
rm nimages.txt
MESS="qcprep.sh: Done. Made $nimages montages in $imgroot_mac/montages"
echo $MESS
DATE=$(date); echo "[ "$DATE" ]: "$MESS >> dapifinder.log
echo "Preferred images can be indicated in preferredTCF.csv, indicate [1-3] or 0 (skip image)."
```

```
# Took this bit from http://29a.ch/2009/5/14/concatenating-images-using-python
from PIL import Image
import sys

if not len(sys.argv) > 3:
    raise SystemExit("Usage: %s src1 [src2] .. dest" % sys.argv[0])

images = map(Image.open, sys.argv[1:-1])
w = sum(i.size[0] for i in images)
mh = max(i.size[1] for i in images)

result = Image.new("RGBA", (w, mh))

x = 0
for i in images:
    result.paste(i, (x, 0))
    x += i.size[0]

result.save(sys.argv[-1])
```

The R script `collatePrefs.R` iterates through the user-specified preferences from `preferredTCF.csv` and merges results from the most successful TCF values per microscopy image.

```
setwd("/Volumes/RED/cp/")
prefs = read.csv("preferredTCF.csv")
data1 = read.csv("out1/DefaultOUT_ShrunkenNuclei.csv")
data2 = read.csv("out2/DefaultOUT_ShrunkenNuclei.csv")
data3 = read.csv("out3/DefaultOUT_ShrunkenNuclei.csv")
nImages = max(prefs$imageNumber)
pickedData = NULL
for (i in 1:nImages) {
  cImage = prefs$preference[i]
  if (cImage == 1) {x = data1[data1$ImageNumber == i,1:ncol(data1)]}
  if (cImage == 2) {x = data2[data2$ImageNumber == i,1:ncol(data2)]}
  if (cImage == 3) {x = data3[data3$ImageNumber == i,1:ncol(data3)]}
  if (cImage > 0) {
    pickedData = rbind(pickedData,x)
  }
}
write.csv(pickedData,file="filteredData.csv")
```

A.2 Analysis of cell lineages in microcolony

A sample of the output can be seen in figure 5.8.

```
setwd("/Volumes/RED/0_catalogued microscope sessions/187_6sep2013_dv/ah7-8/")
rd = read.csv("ah7-8_cropped.csv")
```

```
# firstrun = 1 if wanting to detect cell cycles completely automatically
firstrun = 1
# show_lengths = 1 if needing cell lengths as visual aid (quality control purposes)
show_lengths = 0
man = NULL # manually corrected cell cycle boundaries
if (firstrun != 1) {man = read.csv("man_peak_dips.csv")}
nCells = max(rd$cell)
cells = c(1:nCells)
gyl = c(0,300) # GFP y axis limits
# ma() calculates a moving average with n specifying the window of measurements used
# e.g. 3 means there's one measurement on each side of the centre being averaged
ma <- function(x,n=3){filter(x,rep(1/n,n), sides=2)}
# Specify interval between each frame
res = 8
# Moving average window size. Set manually after looking at graphs
# background blanking red and green values into 'r' and 'g' here
g = NULL
for (i in 1:length(rd$green)) {
  g = c(g,rd$green[i] - rd$green[rd$cell == 0 & rd$frame == rd$frame[i]])
}
# convert frames into minutes (frame 1 = 0 mins)
times = (1:max(rd$frame)-1)*res
xl = c(0,max(times))
# diff() will output a vector of length n-1, so it needs a different times vector
difftimes = times-(res/2)
difftimes = difftimes[-1]
# all.* will hold data relating to whether a cell division has been detected
all.dips = NULL
all.peaks = NULL
all.cellIDs = NULL
for (i in cells) {
  picks = which(rd$cell == i)[order(rd$frame[rd$cell == i])]
  green = g[picks]

  pdf(file=paste("cell ",i,", int plot for thesis.pdf",sep=""))
  par(mfrow=c(1,1),mar=c(4,9,5,7))
  # Get current frames used and work out times. Could vary from cell to cell
  # depending on quality of raw data
  cFrames = rd$frame[picks]
  cTimes = (cFrames-1) * res
  plot(cTimes,green,type='l',lty=1,col=rgb(100,100,100,175,maxColorValue=255),
    lwd=1,xlim=xl,xaxt='n',ylim=gyl,yaxt='n',ylab="GFP (16-bit AU)",xlab="")
  lines(cTimes,ma(green),col=1,lwd=2)
  axis(side=2)
  # cell lengths for the current values being plotted
  cLens = rd$length[picks]
  # cell length graph that growth rate bars are based on. set show_lengths = 1 to draw
  if (show_lengths == 1) {
    par(new=TRUE)
    # colour scheme for every 5 points red, a nice visual aid if doing manual entry
    lencols = rep(1,100)
    lencols[seq(0,100,5)] = 2
    plot(cFrames,cLens,type='o',col=lencols,xlim=c(0,max(cFrames)),
         yaxt='n',xaxt='n',xlab="",ylab="")
    axis(side=3,at=seq(0,max(cFrames),10))
    mtext(side=3,"frame #",line=3)
  }
  # Automatic attempt to detect cell divisions
```

```
# 'dips' are drops in the cell length graph. ie a cell has just divided
# 'peaks' are the preceding peaks in cell length just before division
dips = FALSE # First observation can never be a dip
peaks = rep(FALSE,length(cFrames))
# Walk thorugh every frame and ask if the cell length just dropped
for (j in 2:length(cLens)) {
  division = FALSE
  # padding with a factor of 1.25 to cut down on false positives
  if (cLens[j]*1.25 < cLens[j-1]) {
    division = TRUE
    peaks[j-1] = TRUE
  }
  dips = c(dips,division) # intervals = positions with dips
}
all.dips = c(all.dips,which(dips))
all.peaks =  c(all.peaks,which(peaks))
all.cellIDs =  c(all.cellIDs,rep(i,length(which(dips))))
# After running this script once, automatically detected peaks and dips are
# found in "auto_peaks.csv". Manually remove false positives there and save as
# man_peak_dips.csv, set firstrun to 0 on line 4
if (i %in% man$cell) {
  dips = rep(FALSE,length(times))
  peaks = rep(FALSE,length(times))
  peaks[man$peak[man$cell == i]] = TRUE
  dips[man$dip[man$cell == i]] = TRUE
}
# Calculate data for growth rate bars
widths = which(dips)[1] # First bar is empty and its width is always space up until first dip
rates = 0 # We don't calculate the first rate, because there won't be a dip/peak pair
# growth rate bars are based on the slope of a line fitted to length data from dip to peak
for (j in 1:length(which(dips))) {
  if (which(dips)[j] < max(which(peaks))) {
    cDip = which(dips)[j]
    cPeak = which(peaks)[which(peaks) > cDip][1]
    nDip = which(dips)[j+1]
    if (is.na(nDip)) {nDip = length(cLens)} # Exception for last observation
    fitLine = lm(cLens[cDip:cPeak] ~ cTimes[cDip:cPeak])
    slope = as.numeric(fitLine$coefficients[2])
    rates = c(rates,slope,0) # 0 is for the frame it takes to drop length b/c division
    widths = c(widths,cPeak-cDip,nDip-cPeak) # division always takes 1 frame, so "res" minutes here
  }
}
# Plot growth rate bars
par(new=TRUE)
barplot(rates,widths*res,xlim=xl,space=0,ylim=c(0,0.1),
        col=rgb(100,100,100,10,maxColorValue=255),yaxt='n')
axis(side=4)
mtext(side=4,line=2,"growth rate (um/min)")
dev.off()

# Write a new pdf file with the delta GFP measurements in it
pdf(file=paste("cell ",i,", delta plot for thesis.pdf",sep=""))
par(mfrow=c(1,1),mar=c(15,9,10,7))
plot(difftimes,diff(ma(green)),type='l',lty=1,lwd=0,xaxt='n',
     ylab="",xlab="",ylim=c(-20,20),xlim=xl)
```

```
116   # Visual aids for outer boundaries of complete tracked cell cycles
117   lines(rep(widths[1]*res,2),c(-20,20))
118   lines(rep(sum(widths[1:length(widths)-1])*res,2),c(-20,20))
119   lines(c(0,max(difftimes)),c(0,0))
120   dev.off()
121 }
122 write.csv(data.frame(cell=all.cellIDs,peaks=all.peaks,dips=all.dips),"auto_peaks.csv")
```

A.3 Analysis and compilation of expression data from GFP and β-Gal assays

A CellProfiler analysis pipeline and an automated analysis script written in R were used to facilitate faster analysis of data. The below script takes input from CellProfiler (line 2) in the form of rows in a `preferredTCF.csv` spreadsheet with metadata relating to colony ID, time points (if applicable). A separate file contains manually entered β-gal assay results per colony and time point, as well as chromosomal locus and orientation of the transposon (line 3). Lines 23-47 generate plot symbols according to the position of chromosome and internal orientation of the reporters in the transposable element, resulting in arrows indicating the direction of *lacZ* and *gfp* in relation to the direction of DNA replication. Values for GFP observations, their error bars and noise (defined as $\frac{\sigma}{\mu}$) are drawn iteratively in lines 59–89. The portion of the plot corresponding to β-gal assay readings are plotted in lines 90-106.

```
1  setwd("/Volumes/RED/0_catalogued microscope sessions/177_6jul2013/")
2  rd = read.csv("filteredData.csv")
3  rd2 = read.csv("tndata.csv")
4  times = unique(rd$Metadata_time)
5  strains = unique(rd$Metadata_col)
6  ncols = length(strains)
7  par(mfrow=c(2,1),mar=c(4,4,1,8))
8  # vector for gfp means, sdev, n and time
9  means = NULL
10 sds = NULL
11 ncells = NULL
12 tp = NULL
13 # Generates statistics and metadata on gfp measurements
14 for (i in times) {
15   for (j in strains) {
16     x = rd$Intensity_MedianIntensity_ThreshGreen[rd$Metadata_time == i & rd$Metadata_col
         == j & rd$AreaShape_Area > 25 & rd$Intensity_MeanIntensity_DNA > 0.005] *
        (2^16-1)
17     means = c(means,mean(x))
18     sds = c(sds,sd(x))
19     ncells = c(ncells,length(x))
20     tp = c(tp,i)
21   }
22 }
23 gpchs = NULL
```

```
lpchs = NULL
# Generate a string array with arrow-like plot symbols to show ORF direction
for (i in 1:ncols) {
  pastTer = FALSE
  if (rd2$pos[i] > 180) {pastTer = TRUE}
  if (rd2$gfpdir[i] == 1) {
    gpch = ">"
    lpch = "<"
  }
  if (rd2$gfpdir[i] == 1) {
    gpch = ">"
    lpch = "<"
  }
  if (rd2$gfpdir[i] == 0 & pastTer == FALSE) {
    gpch = "<"
    lpch = ">"
  }
  if (rd2$gfpdir[i] == 0 & pastTer == TRUE) {
    gpch = "<"
    lpch = ">"
  }
  gpchs = c(gpchs,gpch)
  lpchs = c(lpchs,lpch)
}
xl = c(0,360)
for (i in times) {
  plot(0,0,pch="",
       ylim=c(0,max(means)*1.20),
       xlim=xl,
       ylab="bgal units",
       xlab="position on chromosome",
       yaxt='n',xaxt='n')
  # replication direction memory aid
  points(seq(0,360,length.out=50),rep(0,50),pch=c(rep(">",25),rep("<",25)),
         col="gray")
  # Plotting gfp values in this loop
  for (j in strains) {
    # j follows col number, colseq follows its pos in data frame
    colseq = which(rd2$col == j)[1]
    # Make a line connecting the dots under all plot points first
    if (j == 1) {
      lines(sort(rd2$pos)[1:ncols],means[which(rd2$time == i)][order(rd2$pos)[1:ncols]],
            col=rgb(0,255,0,100,maxColorValue=255))
    }
    cgfp = means[which(rd2$time == i & rd2$col == j)]
    csd = sds[which(rd2$time == i & rd2$col == j)]
    cpos = rd2$pos[colseq]
    # Draw error bars for GFP values
    lines(rep(cpos,2),c(cgfp+csd,cgfp-csd),
          col=rgb(0,0,0,100,maxColorValue=255),lwd=3)
    # Draw the actual points
    points(cpos,cgfp,col="green",pch=gpchs[colseq])
    # Axis and labels drawn only once
    if (j == 1) {
      axis(side=4,col="green",las=2)
      mtext("gfp units",side=4,line=2.5)
      axis(side=1,at=seq(0,360,45),labels=c("ori",seq(45,135,45),"ter",seq(225,315,45),"
          ori"))
```

```
81      }
82    }
83    # Draw noise values and associated axis info
84    noise = sds[which(rd2$time == i)]/means[which(rd2$time == i)]
85    par(new=TRUE) # new y axis needs new plot
86    plot(0,0,pch="",bty='n',yaxt='n',xaxt='n',xlab="",ylab="",ylim=c(0,max(sds/means)),
          xlim=xl)
87    lines(sort(rd2$pos[1:ncols]),noise[order(rd2$pos[1:ncols])])
88    mtext("noise",side=4,line=7)
89    axis(side=4,col="black",line=5,las=2)
90    # Set up plot for the bgal values
91    par(new=TRUE) # new y axis needs new plot
92    plot(0,0,pch="",bty='n',yaxt='n',xaxt='n',xlab="",ylab="",ylim=c(min(rd2$bgal)*0.9,max
          (rd2$bgal)*1.2),xlim=xl)
93    # Draw a lone for bgal
94    lines(sort(rd2$pos)[1:ncols],rd2$bgal[which(rd2$time == i)][order(rd2$pos)[1:ncols]],
95          col=rgb(0,0,255,100,maxColorValue=255))
96    # plot actual points for bgal
97    for (j in strains) {
98      colseq = which(rd2$col == j)[1]
99      cpos = rd2$pos[colseq]
100     points(cpos,rd2$bgal[which(rd2$time == i & rd2$col == j)],col="blue",pch=lpchs[
            colseq])
101     # Draw axis and labels once for bgal
102     if (j == 1) {
103       axis(side=2,col="blue",las=2)
104     }
105   }
106 }
```

A.4 Analysis of microplate data

Raw data from the microplate assay was processed using the below R script for quick identification of colonies deviating from the negative control phenotype (see figure 4.10D for a representative output graph). Lines 15–24 allow for repeats of individual colonies to be performed, which was typically set to $n = 2$ in this screen, taking into account the need for a high-throughput method, but also the reduction of noise required to identify colonies of interest with greater confidence. The value plotted is the mean fluorescence of all repeats divided by their mean OD_{600} values, both after blanking against the wells containing uninoculated media; these calculations are represented on lines 52–63 and 69–84. A GFP^- control of 168CA was included but not used in calculations, and instead plotted along other samples to give an indication of what GFP^- strains look like while evaluating results (see dashed line in figure 4.10D). A sample of the output can be seen in figure 4.10D-F.

```
1 setwd("/Volumes/RED/1_misc datasets/023_rpsd-screen_round4_29-30jul2013/")
2 f = t(read.csv("fluo.csv"))
3 a = t(read.csv("abs.csv"))
```

```
# workaround to determine number of timepoints.
# For some reason the column with the time stamps varies.
# check by doing f[,9] to make sure it's in column 9...
timecol = 8
tz = which(as.character(f[,timecol]) == "0")
nTimes = tz[2]-tz[1]
# Specify a cutoff if early or late readings are looking weird,
# e.g: cutoff = c(5,nTimes-5) trims 5 off each side
cutoff = c(15,nTimes-30)
# -1 = blank (media), -2 = optional second blank (non-fluo strain)
# 1-n = samples, 0 = empty
yl.a = c(0.01,0.75) # y axis limits
yl.f = c(0,2000)
yl.adj = c(0,4000)
cutoff2 = 5
# enter microplate layout here
pLayout = matrix(c(0,0,0,0,0,0,0,0,0,0,0,0,
                   0,-1,1:9,0,
                   0,-1,1:9,0,
                   0,10:19,0,
                   0,10:14,20:24,0,
                   0,25:34,0,
                   0,35:44,0,
                   0,0,0,0,0,0,0,0,0,0,0,0),
                 ncol=12,byrow=TRUE)
                 # Colours to use for each sample
cols = c(1,"firebrick3","dodgerblue3",rep("gray",90))
# specify samples to isolate on graph, else leave empty or NULL
isolate = c(2,3,20)
nums = NULL
lets = NULL
samples = sort(unique(as.numeric(pLayout)))
if (is.null(isolate) == FALSE) {samples = samples[which(samples %in% isolate)]}
lcount = 1
ncount = 1
# make vectors with microplate letters and numbers for picking up data
for (i in pLayout) {
  lets = c(lets,LETTERS[lcount]) # Count along the well labels
  nums = c(nums,ncount)
  lcount = lcount + 1
  if (lcount > 8) {
    ncount = ncount + 1
    lcount = 1
  }
}
blanks.f = NULL
blanks.a = NULL
last = 3+nTimes # position of last measurement
times = as.numeric(f[,timecol][4:last])
times.orig = times
# Process the blank measurements
for (i in which(pLayout == -1)) {
  ccol.f = which(f[1,] == lets[i] & f[2,] == nums[i])
  cmeas.f = as.numeric(f[,ccol.f][4:last])
  blanks.f = cbind(blanks.f,cmeas.f)
  ccol.a = which(a[1,] == lets[i] & a[2,] == nums[i])
  cmeas.a = as.numeric(a[,ccol.a][4:last])
  blanks.a = cbind(blanks.a,cmeas.a)
```

```
}
blank.f = rowMeans(blanks.f)
blank.a = rowMeans(blanks.a)
par(mfrow=c(1,3),mar=c(2,4,1,1))
plot(0,0,pch="",xlim=c(0,max(times.orig)),ylim=yl.a,
     xlab="time (h)",ylab="OD600 (AU)",yaxt='n')
axis(side=2,las=2)
text(0,0,"abs")
plot(0,0,pch="",xlim=c(0,max(times.orig)),ylim=yl.f,
     xlab="time (h)",ylab="GFP (16-bit AU)",yaxt='n')
axis(side=2,las=2)
text(0,0,"raw fluo")
plot(0,0,pch="",xlim=c(0,max(times.orig)),ylim=yl.adj,
     xlab="time (h)",ylab="GFP/OD (AU) ",yaxt='n')
axis(side=2,las=2)
text(0,0,"abs adjusted fluo")
alltimes = times[cutoff2:length(times)]
times = times[cutoff[1]:cutoff[2]]
for (i in samples[samples > 0]) {
  cWells = which(pLayout == i)
  cVals.f = NULL
  cVals.a = NULL
  # Find the current sample being processed
  for (j in cWells) {
    ccol.f = which(f[1,] == lets[j] & f[2,] == nums[j])
    cmeas.f = as.numeric(f[,ccol.f][4:last])
    cVals.f = cbind(cVals.f,cmeas.f)
    ccol.a = which(a[1,] == lets[j] & a[2,] == nums[j])
    cmeas.a = as.numeric(a[,ccol.a][4:last])
    cVals.a = cbind(cVals.a,cmeas.a)
  }
  # Apply blanks to samples
  cVals.f.blanked = rowMeans(cVals.f) - blank.f
  cVals.a.blanked = rowMeans(cVals.a) - blank.a
  for (k in 1:3) {
    par(mfg=c(1,k))
    # load values for the three different plots
    if (k == 1) {odAdjust = cVals.a.blanked
                 yl = yl.a
    }
    if (k == 2) {odAdjust = cVals.f.blanked
                 yl = yl.f
    }
    if (k == 3) {odAdjust = cVals.f.blanked/cVals.a.blanked
                 yl = yl.adj
    }
    # Plot it
    plot(0,0,pch="",xlim=c(0,max(times.orig)),ylim=yl,
         xlab="",ylab="",bty='n',yaxt='n',xaxt='n')
    odAdjust.orig = odAdjust[cutoff2:length(odAdjust)]
    odAdjust = odAdjust[cutoff[1]:cutoff[2]]
    thincol = 1
    if (i > 3) {thincol = "gray"} # if sample isn't one of the controls
    lines(alltimes,odAdjust.orig,col=thincol)
    lines(times,odAdjust,col=cols[i],lwd=2)
    if (k == 3) {
      fit = lm(odAdjust ~ times)
      slopes = c(slopes,fit$coefficients[2])
```

```
      colIDs = c(colIDs,i)
    }
  }
}
write.csv(data.frame(colony = colIDs,slope = slopes),file="slopes.csv")
```

A.5 GFP-SsrA half-life assay

A description of the process is available in the corresponding materials and methods section (2.7)).

```
rd = read.csv("fluorescence.csv")
logplot = 0
if (logplot == 1) {lp = 'y'}
if (logplot == 0) {lp = ''}
nImages = max(rd$image)
nStrains = max(rd$strain)
nTreats = max(rd$treatment)
nTimes = max(rd$time)
times = unique(rd$times)[0:nTimes]
strains = as.character(unlist(unique(rd$strains))[1:nStrains])
# Blank measuremenents picked out
allBlanks = NULL
for (i in 1:nImages) {
  allBlanks = c(allBlanks,rd$mean[rd$image == i][1])
}
# Each cell measurement is blanked against a background measurement
bk = NULL
for (i in 1:nImages) {
  x = rd$mean[rd$image == i] - allBlanks[i]
  bk = c(bk,x)
}
means = NULL
sds = NULL
treatList = NULL
strainList = NULL
timeList = NULL
timeMins = NULL
for (i in 1:nStrains) {
  for (j in 1:nTreats) {
    for (k in 1:nTimes) {
      x = bk[which(rd$strain == i & rd$treatment == j & rd$time == k)]
      x = x[x > 0]
      if (length(x) != 0) {
        means = c(means,mean(x))
        sds = c(sds,sd(x))
        strainList = c(strainList,i)
        treatList = c(treatList,j)
        timeList = c(timeList,k)
        timeMins = c(timeMins,times[k])
      }
    }
  }
}
```

```
# This data frame holds the processed data
plot_df = data.frame(means = means,
                     stdev = sds,
                     strainID = strainList,
                     treatementID = treatList,
                     timepoint = timeList,
                     timeMins = timeMins)
yl = c(0,max(plot_df$means))
if (logplot == 1) {yl = c(min(plot_df$means),max(plot_df$means))}
xl = c(1,max(times))
lltys = 1:2
par(mfrow=c(1,1),mar=c(4,4,3,1))
plot(0,0,ylim=yl,xlim=xl,pch="",log=lp,main="sfGFP half-life assay",
     ylab="GFP fluorescence (16-bit AU)",xlab="time (min)")
for (i in 1:nStrains) {
  for (j in 1:nTreats) {
      x = plot_df$timeMins[plot_df$strainID == i & plot_df$treatementID == j]
      y = plot_df$means[plot_df$strainID == i & plot_df$treatementID == j]
      lines(x,y,col=i,lty=lltys[j],lwd=3,type='b')
  }
}
legend("topright",legend=c(strains,"control","cam","fitted line","fitted line (blanked)"
    ),col=c(1:nStrains,1,1,1,1),
       pch=c(rep(15,4),31,31,31,31),lty=c(0,0,0,0,1,2,1,2),lwd=c(0,0,0,0,3,3,1,1))
regressions_b = NULL
c_slope_points = c(6,6,6,6) # Up to which points to use for working out half-life for
    Control. Per strain
slope_points = c(3,4,6,4) # See previous, but for cam-treated
for (i in 1:nStrains) {
  # Pick out the positions in processed data frame corresponding to this strain/
      treatement combination
  c_picks = which(plot_df$strainID == i & plot_df$treatementID == 1 & plot_df$timepoint
      <= c_slope_points[i])
  picks = c( # The first timepoint is shared between control and cam sample
    which(plot_df$strainID == i & plot_df$treatementID == 1 & plot_df$timepoint == 1),
    which(plot_df$strainID == i & plot_df$treatementID == 2)[1:slope_points[i]-1])
  # Fit a line according to the cam-treated sample, draw line
  cLine_cm = lm(plot_df$means[picks]~plot_df$timeMins[picks])
  cSlope_cm = cLine_cm$coefficients[2]
  abline(cLine_cm,col=i)
  # Again, for control
  cLine_c = lm(plot_df$means[c_picks]~plot_df$timeMins[c_picks])
  cSlope_c = cLine_c$coefficients[2]
  abline(cLine_c,col=i)
  # Calculate the "blanked" fitted line
  cLine_b = cLine_cm # cLine_b will hold the equation for a "blanked" line
  cLine_b$coefficients[2] = cSlope_cm - cSlope_c
  abline(cLine_b,lty=2,col=i)
  halflife_b = (cLine_b$coefficients[1]/2)/(-cLine_b$coefficients[2])
  percDrop = plot_df$means[plot_df$strainID == i & plot_df$treatementID == 2 & plot_df$
      timepoint == slope_points[i]]/plot_df$means[plot_df$strainID == i & plot_df$
      treatementID == 1 & plot_df$timepoint == slope_points[i]]
  regressions_b = c(regressions_b,halflife_b)
}
output_df = data.frame(halflife = regressions_b
                       ,strain = strains
                       )
write.csv(file="halflives2.csv",output_df)
```

Appendix B

Additional and alternative versions of figures

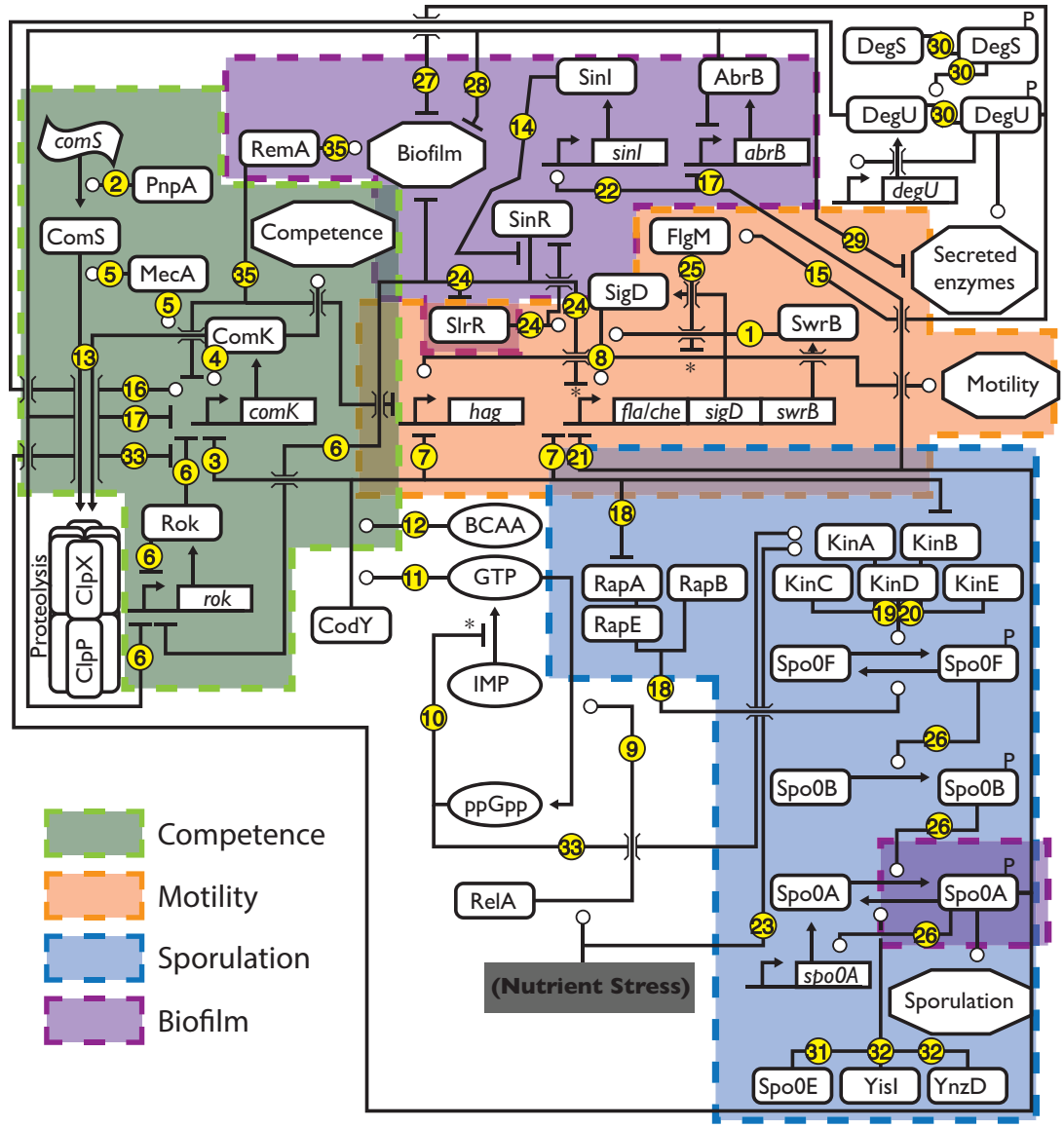

Figure B.1: Overview of relationships between regulators considered in this thesis (with references). See figure legend of figure 1.4 for an explanation of symbols used. Numbered references in yellow circles correspond to the following citations: 1—Kearns and Losick (2005); 2—Luttinger *et al.* (1996); 3—Serror and Sonenshein (1996); 4—Maamar and Dubnau (2005); 5—Turgay *et al.* (1997); 6—Hoa *et al.* (2002); 7—Bergara *et al.* (2003); 8—Márquez-Magaña and Chamberlin (1994); 9—Wendrich and Marahiel (1997); 10—Gallant *et al.* (1971); 11—Ratnayake-Lecamwasam *et al.* (2001); 12—Shivers and Sonenshein (2004); 13—Turgay *et al.* (1998); 14—Bai *et al.* (1993); 15—Hsueh *et al.* (2011); 16—Hamoen *et al.* (2000); 17—Hamoen *et al.* (2003a); 18—Molle *et al.* (2003b); 19—Jiang *et al.* (2000); 20—McLoon *et al.* (2011); 21—Molle *et al.* (2003a); 22—Fujita *et al.* (2005); 23—Kolodkin-Gal *et al.* (2013); 24—Chai *et al.* (2010); 25—Caramori *et al.* (1996); 26—Burbulys *et al.* (1991); 27—Marlow *et al.* (2013); 28—Hamon *et al.* (2004); 29—Veening *et al.* (2008a); 30—Mukai *et al.* (1990); 31—Ohlsen *et al.* (1994); 32—Perego (2001); 33—Tojo *et al.* (2013); 34—Mirouze *et al.* (2012); 35—Winkelman *et al.* (2013)

Bibliography

Agilent Technologies, Inc. (2013), 'Quikchange primer design'.
URL: *http://www.genomics.agilent.com/primerDesignProgram.jsp*

Anagnostopoulos, C. and Spizizen, J. (1961), 'Requirements for Transformation in Bacillus Subtilis', *Journal of Bacteriology* **81**(5), 741–746.

Andersen, J. B., Sternberg, C., Poulsen, L. K., Bjørn, S. P., Givskov, M. and Molin, S. (1998), 'New unstable variants of green fluorescent protein for studies of transient gene expression in bacteria.', *Applied and environmental microbiology* **64**(6), 2240–2246.

Angelini, T. E., Roper, M., Kolter, R., Weitz, D. A. and Brenner, M. P. (2009), 'Bacillus subtilis spreads by surfing on waves of surfactant.', *Proceedings of the National Academy of Sciences of the United States of America* **106**(43), 18109–18113.

Autret, S., Levine, A., Vannier, F., Fujita, Y. and Séror, S. J. (1999), 'The replication checkpoint control in Bacillus subtilis: identification of a novel RTP-binding sequence essential for the replication fork arrest after induction of the stringent response.', *Molecular Microbiology* **31**(6), 1665–1679.

Bai, U., Mandic-Mulec, I. and Smith, I. (1993), 'SinI modulates the activity of SinR, a developmental switch protein of Bacillus subtilis, by protein-protein interaction.', *Genes & development* **7**(1), 139–148.

Balaban, N. Q., Gerdes, K., Lewis, K. and McKinney, J. D. (2013), 'A problem of persistence: still more questions than answers?', *Nature Reviews Microbiology* **11**(8), 587–591.

Belas, R. (2013), 'When the swimming gets tough, the tough form a biofilm', *Molecular Microbiology* **90**(1), 1–5.

Berg, H. C. (2003), 'The rotary motor of bacterial flagella', *Annual Review of Biochemistry* **72**(1), 19–54.

Bergara, F., Ibarra, C., Iwamasa, J., Patarroyo, J. C., Aguilera, R. and Márquez-Magaña, L. M. (2003), 'CodY is a nutritional repressor of flagellar gene expression in Bacillus subtilis.', *Journal of Bacteriology* **185**(10), 3118–3126.

Bigger, J. W. (1944), 'Treatment of staphylococcal infections with penicillin by intermittent sterilisation', *The Lancet* **244**(6320), 497–500.

Blair, K. M., Turner, L., Winkelman, J. T., Berg, H. C. and Kearns, D. B. (2008), 'A Molecular Clutch Disables Flagella in the Bacillus subtilis Biofilm', *Science* **320**(5883), 1636–1638.

Boorsma, A., Foat, B. C., Vis, D., Klis, F. and Bussemaker, H. J. (2005), 'T-profiler: scoring the activity of predefined groups of genes using gene expression data', *Nucleic Acids Research* **33**(Web Server), W592–W595.

Branda, S. S., González-Pastor, J. E., Ben-Yehuda, S., Losick, R. and Kolter, R. (2001), 'Fruiting body formation by Bacillus subtilis.', *Proceedings of the National Academy of Sciences of the United States of America* **98**(20), 11621–11626.

Britton, R. A., Eichenberger, P., González-Pastor, J. E., Fawcett, P., Monson, R., Losick, R. and Grossman, A. D. (2002), 'Genome-Wide Analysis of the Stationary-Phase Sigma Factor (Sigma-H) Regulon of Bacillus subtilis', *Journal of Bacteriology* **184**(17), 4881–4890.

Browning, D. F. and Busby, S. J. W. (2004), 'The regulation of bacterial transcription initiation', *Nature Reviews Microbiology* **2**(1), 57–65.

Burbulys, D., Trach, K. A. and Hoch, J. A. (1991), 'Initiation of sporulation in B. subtilis is controlled by a multicomponent phosphorelay', *Cell* **64**(3), 545–552.

Cairns, L. S., Hobley, L. and Stanley-Wall, N. R. (2014), 'Biofilm formation by Bacillus subtilis: new insights into regulatory strategies and assembly mechanisms', *Molecular Microbiology* **93**(4), 587–598.

Cairns, L. S., Marlow, V. L., Bissett, E., Ostrowski, A. and Stanley-Wall, N. R. (2013), 'A mechanical signal transmitted by the flagellum controls signalling in Bacillus subtilis', *Molecular Microbiology* **90**(1), 6–21.

Camps, M. (2010), 'Modulation of ColE1-like plasmid replication for recombinant gene expression.', *Recent patents on DNA & gene sequences* **4**(1), 58–73.

Cano, R. J. and Borucki, M. K. (1995), 'Revival and identification of bacterial spores in 25-to 40-million-year-old Dominican amber', *Science* **268**(5213), 1060–1064.

Caramori, T., Barilla, D., Nessi, C., Sacchi, L. and Galizzi, A. (1996), 'Role of FlgM in sigma D-dependent gene expression in Bacillus subtilis', *Journal of Bacteriology* **178**(11), 3113–3118.

Carpenter, A. E., Jones, T. R., Lamprecht, M. R., Clarke, C., Kang, I. H., Friman, O., Guertin, D. A., Chang, J. H., Lindquist, R. A., Moffat, J., Golland, P. and Sabatini, D. M. (2006), 'CellProfiler: image analysis software for identifying and quantifying cell phenotypes.', *Genome biology* **7**(10), R100.

Chadsey, M. S. and Hughes, K. T. (2001), 'A multipartite interaction between Salmonella transcription factor σ28 and its anti-sigma factor FlgM: implications for σ28 holoenzyme destabilization through stepwise binding', *Journal of Molecular Biology* **306**(5), 915–929.

Chai, Y., Beauregard, P. B., Vlamakis, H., Losick, R. and Kolter, R. (2012), 'Galactose Metabolism Plays a Crucial Role in Biofilm Formation by Bacillus subtilis', *mBio* **3**(4), e00184–12–e00184–12.

Chai, Y., Norman, T. and Kolter, R. (2010), 'An epigenetic switch governing daughter cell separation in Bacillus subtilis', *Genes & development* **24**(8), 754–765.

Chai, Y., Norman, T., Kolter, R. and Losick, R. (2011), 'Evidence that metabolism and chromosome copy number control mutually exclusive cell fates in Bacillus subtilis.', *The EMBO Journal* .

Chiang, C.-F., Okou, D. T., Griffin, T. B., Verret, C. R. and Williams, M. N. V. (2001), 'Green Fluorescent Protein Rendered Susceptible to Proteolysis: Positions for Protease-Sensitive Insertions', *Archives of Biochemistry and Biophysics* **394**(2), 229–235.

Chu, F., Kearns, D. B., McLoon, A., Chai, Y., Kolter, R. and Losick, R. (2008), 'A novel regulatory protein governing biofilm formation in Bacillus subtilis', *Molecular Microbiology* **68**(5), 1117–1127.

Chudakov, D. M., Matz, M. V., Lukyanov, S. and Lukyanov, K. A. (2010), 'Fluorescent Proteins and Their Applications in Imaging Living Cells and Tissues', *Physiological Reviews* **90**(3), 1103–1163.

Claverys, J.-P. and Martin, B. (2003), 'Bacterial "competence" genes: signatures of active transformation, or only remnants?', *Trends in microbiology* **11**(4), 161–165.

Commichau, F. M., Pietack, N. and Stülke, J. (2013), 'Essential genes in Bacillus subtilis: a re-evaluation after ten years', *Molecular BioSystems* .

Cookson, N. A., Mather, W. H., Danino, T., Mondragón-Palomino, O., Williams, R. J., Tsimring, L. S. and Hasty, J. (2011), 'Queueing up for enzymatic processing: correlated signaling through coupled degradation', *Molecular systems biology* **7**.

Cormack, B. P., Valdivia, R. H. and Falkow, S. (1996), 'FACS-optimized mutants of the green fluorescent protein (GFP)', *Gene* **173**(1 Spec No), 33–38.

Cozy, L. M. and Kearns, D. B. (2010), 'Gene position in a long operon governs motility development in Bacillus subtilis', *Molecular Microbiology* **76**(2), 273–285.

Dagert, M. and Ehrlich, S. D. (1979), 'Prolonged incubation in calcium chloride improves the competence of Escherichia coli cells', *Gene* **6**(1), 23–28.

Daniel, R. A., Haiech, J., Denizot, F. and Errington, J. (1997), 'Isolation and characterization of the lacA gene encoding beta-galactosidase in Bacillus subtilis and a regulator gene, lacR.', *Journal of Bacteriology* **179**(17), 5636–5638.

Das, T., Sharma, P. K., Busscher, H. J., van der Mei, H. C. and Krom, B. P. (2010), 'Role of Extracellular DNA in Initial Bacterial Adhesion and Surface Aggregation', *Applied and environmental microbiology* **76**(10), 3405–3408.

de Jong, I. G., Beilharz, K., Kuipers, O. P. and Veening, J.-W. (2011a), 'Live Cell Imaging of Bacillus subtilis and Streptococcus pneumoniae using Automated Time-lapse Microscopy', *Journal of Visualized Experiments* (53), e3145.

de Jong, I. G., Haccou, P. and Kuipers, O. P. (2011b), 'Bet hedging or not? A guide to proper classification of microbial survival strategies', *BioEssays* **33**(3), 215–223.

Driks, A. (2011), 'Tapping into the biofilm: insights into assembly and disassembly of a novel amyloid fibre in Bacillus subtilis', *Molecular Microbiology* **80**(5), 1133–1136.

D'Souza, C., Nakano, M. M. and Zuber, P. (1994), 'Identification of comS, a gene of the srfA operon that regulates the establishment of genetic competence in Bacillus subtilis.', *Proceedings of the National Academy of Sciences of the United States of America* **91**(20), 9397–9401.

Dubnau, D. (1991), 'Genetic competence in Bacillus subtilis.', *Microbiological reviews* **55**(3), 395–424.

Elowitz, M. B., J, L. A., Siggia, E. D. and Swain, P. S. (2002), 'Stochastic Gene Expression in a Single Cell', *Science* **297**(5584), 1183–1186.

Elowitz, M. B. and Leibler, S. (2000), 'A synthetic oscillatory network of transcriptional regulators', *Nature* **403**(6767), 335–338.

Epstein, A. K., Pokroy, B., Seminara, A. and Aizenberg, J. (2011), 'Bacterial biofilm shows persistent resistance to liquid wetting and gas penetration.', *Proceedings of the National Academy of Sciences of the United States of America* **108**(3), 995–1000.

Eswaramoorthy, P., Duan, D., Dinh, J., Dravis, A., Devi, S. N. and Fujita, M. (2010), 'The Threshold Level of the Sensor Histidine Kinase KinA Governs Entry into Sporulation in Bacillus subtilis', *Journal of Bacteriology* **192**(15), 3870–3882.

Feucht, A. and Lewis, P. J. (2001), 'Improved plasmid vectors for the production of multiple fluorescent protein fusions in Bacillus subtilis', *Gene* **264**(2), 289–297.

Finkel, S. E. and Kolter, R. (2001), 'DNA as a Nutrient: Novel Role for Bacterial Competence Gene Homologs', *Journal of Bacteriology* **183**(21), 6288–6293.

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

Gallant, J., Irr, J. and Cashel, M. (1971), 'The mechanism of amino acid control of guanylate and adenylate biosynthesis.', *The Journal of biological chemistry* **246**(18), 5812–5816.

Geiger, T. and Wolz, C. (2014), 'Intersection of the stringent response and the CodY regulon in low GC Gram-positive bacteria.', *International journal of medical microbiology : IJMM* .

Gerlach, R. G. and Hensel, M. (2007), 'Protein secretion systems and adhesins: the molecular armory of Gram-negative pathogens', *International journal of medical microbiology : IJMM* **297**(6), 401–415.

Gottesman, S., Roche, E., Zhou, Y. and Sauer, R. T. (1998), 'The ClpXP and ClpAP proteases degrade proteins with carboxy-terminal peptide tails added by the SsrA-tagging system.', *Genes & development* **12**(9), 1338–1347.

Griffith, K. L. and Grossman, A. D. (2008), 'Inducible protein degradation in Bacillus subtilisusing heterologous peptide tags and adaptor proteins to target substrates to the protease ClpXP', *Molecular Microbiology* .

Grilly, C., Stricker, J., Pang, W. L., Bennett, M. R. and Hasty, J. (2007), 'A synthetic gene network for tuning protein degradation in Saccharomyces cerevisiae', *Molecular systems biology* **3**.

Grundy, F. J. and Henkin, T. M. (1991), 'The rpsD gene, encoding ribosomal protein S4, is autogenously regulated in Bacillus subtilis.', *Journal of Bacteriology* **173**(15), 4595–4602.

Grundy, F. J. and Henkin, T. M. (1992), 'Characterization of the Bacillus subtilis rpsD regulatory target site.', *Journal of Bacteriology* **174**(21), 6763–6770.

Gur, E., Ottofueling, R. and Dougan, D. A. (2013), Machines of Destruction – AAA+ Proteases and the Adaptors That Control Them, *in* 'Regulated Proteolysis in Microorganisms', Springer Netherlands, Dordrecht, pp. 3–33.

Guttenplan, S. B., Shaw, S. and Kearns, D. B. (2012), 'The cell biology of peritrichous flagella in Bacillus subtilis', *Molecular Microbiology* **87**(1), 211–229.

Hahn, J., Roggiani, M. and Dubnau, D. (1995), 'The major role of Spo0A in genetic competence is to downregulate abrB, an essential competence gene.', *Journal of Bacteriology* **177**(12), 3601–3605.

Haijema, B. J., Hahn, J., Haynes, J. and Dubnau, D. (2001), 'A ComGA-dependent checkpoint limits growth during the escape from competence.', *Molecular Microbiology* **40**(1), 52–64.

Haldenwang, W. G. (1995), 'The sigma factors of Bacillus subtilis.', *Microbiological reviews* **59**(1), 1–30.

Hamoen, L. W., Kausche, D., Marahiel, M. A., van Sinderen, D., Venema, G. and Serror, P. (2003a), 'The Bacillus subtilis transition state regulator AbrB binds to the -35 promoter region of comK.', *FEMS microbiology letters* **218**(2), 299–304.

Hamoen, L. W., Smits, W. K., de Jong, A., Holsappel, S. and Kuipers, O. P. (2002), 'Improving the predictive value of the competence transcription factor (ComK) binding site in Bacillus subtilis using a genomic approach', *Nucleic acids research* **30**(24), 5517–5528.

Hamoen, L. W., Van Werkhoven, A. F., Venema, G. and Dubnau, D. (2000), 'The pleiotropic response regulator DegU functions as a priming protein in competence development in Bacillus subtilis.', *Proceedings of the National Academy of Sciences of the United States of America* **97**(16), 9246–9251.

Hamoen, L. W., Venema, G. and Kuipers, O. P. (2003b), 'Controlling competence in Bacillus subtilis: shared use of regulators', *Microbiology* **149**(1), 9–17.

Hamon, M. A., Stanley, N. R., Britton, R. A., Grossman, A. D. and Lazazzera, B. A. (2004), 'Identification of AbrB-regulated genes involved in biofilm formation by Bacillus subtilis.', *Molecular Microbiology* **52**(3), 847–860.

Handke, L. D., Shivers, R. P. and Sonenshein, A. L. (2008), 'Interaction of Bacillus subtilis CodY with GTP', *Journal of Bacteriology* **190**(3), 798–806.

Harwood, C. R. and Cutting, S. M., eds (1990), *Molecular Biological Methods for Bacillus*, Modern Microbiological Methods, John Wiley & Sons, Chichester.

Helaine, S., Cheverton, A. M., Watson, K. G., Faure, L. M., Matthews, S. A. and Holden, D. W. (2014), 'Internalization of Salmonella by Macrophages Induces Formation of Nonreplicating Persisters', *Science* **343**(6167), 204–208.

Hilbert, D. W. and Piggot, P. J. (2004), 'Compartmentalization of Gene Expression during Bacillus subtilis Spore Formation', *Microbiology and Molecular Biology Reviews* **68**(2), 234–262.

Hinterdorfer, P. and van Oijen, A., eds (2009), *Handbook of Single-Molecule Biophysics*, Springer, New York.

Hirota, N. and Imae, Y. (1983), 'Na^+-driven Flagellar Motors of an Alkalophilic Bacillus strain YN-1', *The Journal of biological chemistry* **258**(17), 10577–10581.

Hoa, T. T., Tortosa, P., Albano, M. and Dubnau, D. (2002), 'Rok (YkuW) regulates genetic competence in Bacillus subtilis by directly repressing comK.', *Molecular Microbiology* **43**(1), 15–26.

Hobley, L., Kim, S. H., Maezato, Y., Wyllie, S., Fairlamb, A. H., Stanley-Wall, N. R. and Michael, A. J. (2014), 'Norspermidine Is Nota Self-Produced Trigger for Biofilm Disassembly', *Cell* **156**(4), 844–854.

Hobley, L., Ostrowski, A., Rao, F. V., Bromley, K. M., Porter, M., Prescott, A. R., Macphee, C. E., van Aalten, D. M. F. and Stanley-Wall, N. R. (2013), 'BslA is a self-assembling bacterial hydrophobin that coats the Bacillus subtilis biofilm.', *Proceedings of the National Academy of Sciences* **110**(33), 13600–13605.

Horton, R. M. (1995), 'PCR-mediated recombination and mutagenesis. SOEing together tailor-made genes.', *Molecular biotechnology* **3**(2), 93–99.

Hsueh, Y.-H., Cozy, L. M., Sham, L.-T., Calvo, R. A., Gutu, A. D., Winkler, M. E. and Kearns, D. B. (2011), 'DegU-phosphate activates expression of the anti-sigma factor FlgM in Bacillus subtilis', *Molecular Microbiology* **81**(4), 1092–1108.

Härtl, B., Wehrl, W., Wiegert, T., Homuth, G. and Schumann, W. (2001), 'Development of a new integration site within the Bacillus subtilis chromosome and construction of compatible expression cassettes', *Journal of Bacteriology* **183**(8), 2696–2699.

Ikeda, T., Oosawa, K. and Hotani, H. (1996), 'Self-assembly of the filament capping protein, FliD, of bacterial flagella into an annular structure.', *Journal of molecular biology* **259**(4), 679–686.

Ito, K.-i., Tadaki, T., Lee, S., Takada, K., Muto, A. and Himeno, H. (2002), 'Trans-translation mediated by Bacillus subtilis tmRNA.', *FEBS letters* **516**(1-3), 245–252.

Jiang, M., Shao, W., Perego, M. and Hoch, J. A. (2000), 'Multiple histidine kinases regulate entry into stationary phase and sporulation in Bacillus subtilis.', *Molecular Microbiology* **38**(3), 535–542.

Johnston, C., Martin, B., Fichant, G., Polard, P. and Claverys, J.-P. (2014), 'Bacterial transformation: distribution, shared mechanisms and divergent control', *Nature Reviews Microbiology* **12**(3), 181–196.

Jones, S. E., Paynich, M. L., Kearns, D. B. and Knight, K. L. (2014), 'Protection from Intestinal Inflammation by Bacterial Exopolysaccharides', *The Journal of Immunology* **192**(10), 4813–4820.

Jores, L. and Wagner, R. (2003), 'Essential Steps in the ppGpp-dependent Regulation of Bacterial Ribosomal RNA Promoters Can Be Explained by Substrate Competition', *Journal of Biological Chemistry* **278**(19), 16834–16843.

Kearns, D. B. (2010), 'A field guide to bacterial swarming motility', *Nature Reviews Microbiology* **8**(9), 634–644.

Kearns, D. B., Chu, F., Rudner, R. and Losick, R. (2004), 'Genes governing swarming in Bacillus subtilis and evidence for a phase variation mechanism controlling surface motility', *Molecular Microbiology* **52**(2), 357–369.

Kearns, D. B. and Losick, R. (2005), 'Cell population heterogeneity during growth of Bacillus subtilis', *Genes & development* **19**(24), 3083–3094.

Khasanov, F. K., Zvingila, D. J., Zainullin, A. A., Prozorov, A. A. and Bashkirov, V. I. (1992), 'Homologous recombination between plasmid and chromosomal DNA in Bacillus subtilis requires approximately 70 bp of homology.', *Molecular & general genetics : MGG* **234**(3), 494–497.

Kirstein, J., Molière, N., Dougan, D. A. and Turgay, K. (2009), 'Adapting the machine: adaptor proteins for Hsp100/Clp and AAA+ proteases.', *Nature Reviews Microbiology* **7**(8), 589–599.

Kirstein, J., Strahl, H., Molière, N., Hamoen, L. W. and Turgay, K. (2008), 'Localization of general and regulatory proteolysis in Bacillus subtilis cells', *Molecular Microbiology* **70**(3), 682–694.

Knobloch, J. K. M., Nedelmann, M., Kiel, K., Bartscht, K., Horstkotte, M. A., Dobinsky, S., Rohde, H. and Mack, D. (2003), 'Establishment of an Arbitrary PCR for Rapid Identification of Tn917 Insertion Sites in Staphylococcus epidermidis: Characterization of Biofilm-Negative and Nonmucoid Mutants', *Applied and environmental microbiology* **69**(10), 5812–5818.

Kobayashi, K. and Iwano, M. (2012), 'BslA(YuaB) forms a hydrophobic layer on the surface of Bacillus subtilis biofilms', *Molecular Microbiology* **85**(1), 51–66.

Kolodkin-Gal, I., Cao, S., Chai, L., Böttcher, T., Kolter, R., Clardy, J. and Losick, R. (2012), 'A Self-Produced Trigger for Biofilm Disassembly that Targets Exopolysaccharide', *Cell* **149**(3), 684–692.

Kolodkin-Gal, I., Elsholz, A. K. W., Muth, C., Girguis, P. R., Kolter, R. and Losick, R. (2013), 'Respiration control of multicellularity in Bacillus subtilis by a complex of the cytochrome chain with a membrane-embedded histidine kinase', *Genes & development* **27**(8), 887–899.

Kolodkin-Gal, I., Romero, D., Cao, S., Clardy, J., Kolter, R. and Losick, R. (2010), 'D-Amino Acids Trigger Biofilm Disassembly', *Science* **328**(5978), 627–629.

Krásný, L., Tiserová, H., Jonák, J., Rejman, D. and Sanderová, H. (2008), 'The identity of the transcription +1 position is crucial for changes in gene expression in response to amino acid starvation in Bacillus subtilis.', *Molecular Microbiology* **69**(1), 42–54.

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., Cummings, N. J., Daniel, R. A., Denizot, F., Devine, K. M., Düsterhöft, A., Ehrlich, S. D., Emmerson, P. T., Entian, K. D., Errington, J., Fabret, C., Ferrari, E., Foulger, D., Fritz, C., Fujita, M., Fujita, Y., Fuma, S., Galizzi, A., Galleron, N., Ghim, S. Y., Glaser, P., Goffeau, A., Golightly, E. J., Grandi, G., Guiseppi, G., Guy, B. J., Haga, K., Haiech, J., Harwood, C. R., Hénaut, A., Hilbert, H., Holsappel, S., Hosono, S., Hullo, M. F., Itaya, M., Jones, L., Joris, B., Karamata, D., Kasahara, Y., Klaerr-Blanchard, M., Klein, C., Kobayashi, Y., Koetter, P., Koningstein, G., Krogh, S., Kumano, M., Kurita, K., Lapidus, A., Lardinois, S., Lauber, J., Lazarevic, V., Lee, S. M., Levine, A., Liu, H., Masuda, S., Mauël, C., Médigue, C., Medina, N., Mellado, R. P., Mizuno, M., Moestl, D., Nakai, S., Noback, M., Noone, D., O'Reilly, M., Ogawa, K., Ogiwara, A., Oudega, B., Park, S. H., Parro, V., Pohl, T. M., Portetelle, D., Porwollik, S., Prescott, A. M., Presecan, E., Pujic, P., Purnelle, B., Rapoport, G., Rey, M., Reynolds, S.,

Rieger, M., Rivolta, C., Rocha, E., Roche, B., Rose, M., Sadaie, Y., Sato, T., Scanlan, E., Schleich, S., Schroeter, R., Scoffone, F., Sekiguchi, J., Sekowska, A., Seror, S. J., Serror, P., Shin, B. S., Soldo, B., Sorokin, A., Tacconi, E., Takagi, T., Takahashi, H., Takemaru, K., Takeuchi, M., Tamakoshi, A., Tanaka, T., Terpstra, P., Tognoni, A., Tosato, V., Uchiyama, S., Vandenbol, M., Vannier, F., Vassarotti, A., Viari, A., Wambutt, R., Wedler, E., Wedler, H., Weitzenegger, T., Winters, P., Wipat, A., Yamamoto, H., Yamane, K., Yasumoto, K., Yata, K., Yoshida, K., Yoshikawa, H. F., Zumstein, E., Yoshikawa, H. and Danchin, A. (1997), 'The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.', *Nature* **390**(6657), 249–256.

Kunz, B. A. and Kohalmi, S. E. (1991), 'Modulation of mutagenesis by deoxyribonucleotide levels.', *Annual Review of Genetics* **25**, 339–359.

Kutsukake, K., Iyoda, S., Ohnishi, K. and Iino, T. (1994), 'Genetic and molecular analyses of the interaction between the flagellum-specific sigma and anti-sigma factors in Salmonella typhimurium.', *The EMBO journal* **13**(19), 4568–4576.

Kærn, M., Elston, T. C., Blake, W. J. and Collins, J. J. (2005), 'Stochasticity in gene expression: from theories to phenotypes', *Nature Reviews Genetics* **6**(6), 451–464.

Labhsetwar, P., Cole, J. A., Roberts, E., Price, N. D. and Luthey-Schulten, Z. A. (2013), 'Heterogeneity in protein expression induces metabolic variability in a modeled Escherichia coli population.', *Proceedings of the National Academy of Sciences* **110**(34), 14006–14011.

Le Breton, Y., Mohapatra, N. P. and Haldenwang, W. G. (2006), 'In vivo random mutagenesis of Bacillus subtilis by use of TnYLB-1, a mariner-based transposon.', *Applied and environmental microbiology* **72**(1), 327–333.

Lederberg, E. M. (1950), 'Lysogenicity of Escherichia coli strain K-12', *Microbial Genetics Bulletin* **1**, 5–9.

Leiman, S. A., May, J. M., Lebar, M. D., Kahne, D., Kolter, R. and Losick, R. (2013), 'D-Amino Acids Indirectly Inhibit Biofilm Formation in Bacillus subtilis by Interfering with Protein Synthesis', *Journal of Bacteriology* **195**(23), 5391–5395.

Levdikov, V. M., Blagova, E., Joseph, P., Sonenshein, A. L. and Wilkinson, A. J. (2006), 'The structure of CodY, a GTP- and isoleucine-responsive regulator of stationary phase and virulence in gram-positive bacteria.', *The Journal of biological chemistry* **281**(16), 11366–11373.

Leveau, J. H. J. and Lindow, S. E. (2001), 'Predictive and Interpretive Simulation of Green Fluorescent Protein Expression in Reporter Bacteria', *Journal of Bacteriology* **183**(23), 6752–6762.

Levine, J. H., Fontes, M. E., Dworkin, J. and Elowitz, M. B. (2012), 'Pulsed feedback defers cellular differentiation.', *PLoS Biology* **10**(1), e1001252.

Lewis, K. (2010), 'Persister Cells', *Annual Review of Microbiology* **64**(1), 357–372.

Lewis, P. J. and Marston, A. L. (1999), 'GFP vectors for controlled expression and dual labelling of protein fusions in Bacillus subtilis', *Gene* **227**(1), 101–110.

Li, X., Lindahl, L., Sha, Y. and Zengel, J. M. (1997), 'Analysis of the Bacillus subtilis S10 ribosomal protein gene cluster identifies two promoters that may be responsible for transcription of the entire 15-kilobase S10-spc-α cluster.', *Journal of Bacteriology* **179**(22), 7046–7054.

Liu, J. and Zuber, P. (1998), 'A Molecular Switch Controlling Competence and Motility: Competence Regulatory Factors ComS, MecA, and ComK Control ςD-Dependent Gene Expression in Bacillus subtilis', *Journal of Bacteriology* **180**(16), 4243–4251.

Liu, L., Nakano, M. M., Lee, O. H. and Zuber, P. (1996), 'Plasmid-amplified comS enhances genetic competence and suppresses sinR in Bacillus subtilis.', *Journal of Bacteriology* **178**(17), 5144–5152.

Lo Scrudato, M. and Blokesch, M. (2012), 'The Regulatory Network of Natural Competence and Transformation of Vibrio cholerae', *PLoS Genetics* **8**(6), e1002778.

Locke, J. C. W. and Elowitz, M. B. (2009), 'Using movies to analyse gene circuit dynamics in single cells', *Nature Reviews Microbiology* **7**(5), 383–392.

López, D., Fischbach, M. A., Chu, F., Losick, R. and Kolter, R. (2009a), 'Structurally diverse natural products that cause potassium leakage trigger multicellularity in Bacillus subtilis.', *Proceedings of the National Academy of Sciences of the United States of America* **106**(1), 280–285.

López, D., Vlamakis, H. and Kolter, R. (2009b), 'Generation of multiple cell types in Bacillus subtilis.', *FEMS Microbiology Reviews* **33**(1), 152–163.

López, D., Vlamakis, H., Losick, R. and Kolter, R. (2009c), 'Cannibalism enhances biofilm development in Bacillus subtilis', *Molecular Microbiology* **74**(3), 609–618.

Lopez, J. M., Dromerick, A. and Freese, E. (1981), 'Response of guanosine 5'-triphosphate concentration to nutritional changes and its significance for Bacillus subtilis sporulation.', *Journal of Bacteriology* **146**(2), 605–613.

Losick, R. and Desplan, C. (2008), 'Stochasticity and Cell Fate', *Science* **320**(5872), 65–68.

Luttinger, A., Hahn, J. and Dubnau, D. (1996), 'Polynucleotide phosphorylase is necessary for competence development in Bacillus subtilis.', *Molecular Microbiology* **19**(2), 343–356.

Maamar, H. and Dubnau, D. (2005), 'Bistability in the Bacillus subtilis K-state (competence) system requires a positive feedback loop', *Molecular Microbiology* **56**(3), 615–624.

Magnusson, L. U., Farewell, A. and Nyström, T. (2005), 'ppGpp: a global regulator in Escherichia coli', *Trends in microbiology* **13**(5), 236–242.

Malik, S., Zalenskaya, K. and Goldfarb, A. (1987), 'Competition between sigma factors for core RNA polymerase', *Nucleic acids research* **15**(20), 8521–8530.

Marlow, V. L., Porter, M., Hobley, L., Kiley, T. B., Swedlow, J. R., Davidson, F. A. and Stanley-Wall, N. R. (2013), 'Phosphorylated DegU Manipulates Cell Fate Differentiation in the Bacillus subtilis Biofilm', *Journal of Bacteriology* **196**(1), 16–27.

Márquez, L. M., Helmann, J. D., Ferrari, E., Parker, H. M., Ordal, G. W. and Chamberlin, M. J. (1990), 'Studies of sigma D-dependent functions in Bacillus subtilis', *Journal of Bacteriology* **172**(6), 3435–3443.

Márquez-Magaña, L. M. and Chamberlin, M. J. (1994), 'Characterization of the sigD transcription unit of Bacillus subtilis', *Journal of Bacteriology* **176**(8), 2427–2434.

Marvasi, M., Visscher, P. T. and Casillas Martinez, L. (2010), 'Exopolymeric substances (EPS) from Bacillus subtilis: polymers and genes encoding their synthesis', *FEMS microbiology letters* **313**(1), 1–9.

McLoon, A. L., Kolodkin-Gal, I., Rubinstein, S. M., Kolter, R. and Losick, R. (2011), 'Spatial Regulation of Histidine Kinases Governing Biofilm Formation in Bacillus subtilis', *Journal of Bacteriology* **193**(3), 679–685.

Mehling, M. and Tay, S. (2014), 'Microfluidic cell culture', *Current opinion in biotechnology* **25**, 95–102.

Metropolis, N. (1987), 'The beginning of the Monte Carlo method', *Los Alamos Science* **15**(584), 125–130.

Miller, J. H. (1972), *Experiments in molecular genetics*, 3 edn, Cold Spring Harbor Laboratory, New York.

Mirel, D. B. and Chamberlin, M. J. (1989), 'The Bacillus subtilis flagellin gene (hag) is transcribed by the sigma 28 form of RNA polymerase.', *Journal of Bacteriology* **171**(6), 3095–3101.

Mirkin, E. V. and Mirkin, S. M. (2005), 'Mechanisms of Transcription-Replication Collisions in Bacteria', *Molecular and Cellular Biology* **25**(3), 888–895.

Mirouze, N., Desai, Y., Raj, A. and Dubnau, D. (2012), 'Spo0AâĹijP Imposes a Temporal Gate for the Bimodal Expression of Competence in Bacillus subtilis', *PLoS Genetics* **8**(3), e1002586.

Mitra, A., Kesarwani, A. K., Pal, D. and Nagaraja, V. (2010), 'WebGeSTer DB—a transcription terminator database', *Nucleic acids research* **39**(Database), D129–D135.

Molière, N. and Turgay, K. (2013), General and Regulatory Proteolysis in Bacillus subtilis, *in* 'Regulated Proteolysis in Microorganisms', Springer Netherlands, Dordrecht, pp. 73–103.

Molle, V., Fujita, M., Jensen, S. T., Eichenberger, P., González-Pastor, J. E., Liu, J. S. and Losick, R. (2003a), 'The Spo0A regulon of Bacillus subtilis', *Molecular Microbiology* **50**(5), 1683–1701.

Molle, V., Nakaura, Y., Shivers, R. P., Yamaguchi, H., Losick, R., Fujita, Y. and Sonenshein, A. L. (2003b), 'Additional targets of the Bacillus subtilis global regulator CodY identified by chromatin immunoprecipitation and genome-wide transcript analysis', *Journal of Bacteriology* **185**(6), 1911–1922.

Moran, C. P., Lang, N., LeGrice, S. F., Lee, G., Stephens, M., Sonenshein, A. L., Pero, J. and Losick, R. (1982), 'Nucleotide sequences that signal the initiation of transcription and translation in Bacillus subtilis.', *Molecular & general genetics : MGG* **186**(3), 339–346.

Morimoto, T., Loh, P. C., Hirai, T., Asai, K., Kobayashi, K., Moriya, S. and Ogasawara, N. (2002), 'Six GTP-binding proteins of the Era/Obg family are essential for cell growth in Bacillus subtilis', *Microbiology* **148**(Pt 11), 3539–3552.

Moszer, I., Glaser, P. and Danchin, A. (1995), 'SubtiList: a relational database for the Bacillus subtilis genome.', *Microbiology (Reading, England)* **141 (Pt 2)**, 261–268.

Mukai, K., Kawata, M. and Tanaka, T. (1990), 'Isolation and phosphorylation of the Bacillus subtilis degS and degU gene products.', *The Journal of biological chemistry* **265**(32), 20000–20006.

Mulder, K. C. L. and Schumann, W. (2013), 'Construction and Analysis of a Modified Transposable Element Carrying an Outward Directed Inducible Promoter for Bacillus subtilis', *Current Microbiology* .

Murray, E. J., Kiley, T. B. and Stanley-Wall, N. R. (2009), 'A pivotal role for the response regulator DegU in controlling multicellular behaviour.', *Microbiology (Reading, England)* **155**(Pt 1), 1–8.

Murray, H. D., Schneider, D. A. and Gourse, R. L. (2003), 'Control of rRNA expression by small molecules is dynamic and nonredundant.', *Molecular cell* **12**(1), 125–134.

Mäder, U., Antelmann, H., Buder, T., Dahl, M., Hecker, M. and Homuth, G. (2002), 'Bacillus subtilis functional genomics: genome-wide analysis of the DegS-DegU regulon by transcriptomics and proteomics', *Molecular Genetics and Genomics* **268**(4), 455–467.

Nanamiya, H., Kasai, K., Nozawa, A., Yun, C.-S., Narisawa, T., Murakami, K., Natori, Y., Kawamura, F. and Tozawa, Y. (2008), 'Identification and functional analysis of novel (p)ppGpp synthetase genes in Bacillus subtilis.', *Molecular Microbiology* **67**(2), 291–304.

Nicolas, P., Mäder, 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., Härtig, 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. T., Nannapaneni, P., Noone, D., Pohl, S., Rinn, B., Rügheimer, F., Sappa, P. K., Samson, F., Schaffer, M., Schwikowski, B., Steil, L., Stülke, J., Wiegert, T., Devine, K. M., Wilkinson, A. J., Maarten van Dijl, J., Hecker, M., Völker, U., Bessières, P. and Noirot, P. (2012), 'Condition-Dependent Transcriptome Reveals High-Level Regulatory Architecture in Bacillus subtilis', *Science* **335**(6072), 1103–1106.

Nijland, R., Hall, M. J. and Burgess, J. G. (2010), 'Dispersal of biofilms by secreted, matrix degrading, bacterial DNase', *PLoS ONE* **5**(12), e15668.

Norman, T. M., Lord, N. D., Paulsson, J. and Losick, R. (2013), 'Memory and modularity in cell-fate decision making', *Nature* **503**(7477), 481–486.

Ohlsen, K. L., Grimsley, J. K. and Hoch, J. A. (1994), 'Deactivation of the sporulation transcription factor Spo0A by the Spo0E protein phosphatase.', *Proceedings of the National Academy of Sciences of the United States of America* **91**(5), 1756–1760.

Pandey, R., Ter Beek, A., Vischer, N. O. E., Smelt, J. P. P. M., Brul, S. and Manders, E. M. M. (2013), 'Live Cell Imaging of Germination and Outgrowth of Individual Bacillus subtilis Spores; the Effect of Heat Stress Quantitatively Analyzed with SporeTracker', *PLoS ONE* **8**(3), e58972.

Pédelacq, J.-D., Cabantous, S., Tran, T., Terwilliger, T. C. and Waldo, G. S. (2005), 'Engineering and characterization of a superfolder green fluorescent protein', *Nature Biotechnology* **24**(1), 79–88.

Perego, M. (2001), 'A new family of aspartyl phosphate phosphatases targeting the sporulation transcription factor Spo0A of Bacillus subtilis.', *Molecular Microbiology* **42**(1), 133–143.

Persuh, M., Turgay, K., Mandic-Mulec, I. and Dubnau, D. (1999), 'The N- and C-terminal domains of MecA recognize different partners in the competence molecular switch.', *Molecular Microbiology* **33**(4), 886–894.

Piersma, S., Denham, E. L., Drulhe, S., Tonk, R. H. J., Schwikowski, B. and van Dijl, J. M. (2013), 'TLM-Quant: An Open-Source Pipeline for Visualization and Quantification of Gene Expression Heterogeneity in Growing Microbial Cells', *PLoS ONE* **8**(7), e68696.

Piggot, P. J. and Hilbert, D. W. (2004), 'Sporulation of Bacillus subtilis', *Current opinion in microbiology* **7**(6), 579–586.

Pozsgai, E. R., Blair, K. M. and Kearns, D. B. (2011), 'Modified mariner transposons for random inducible-expression insertions and transcriptional reporter fusion insertions in Bacillus subtilis.', *Applied and environmental microbiology* **78**(3), 778–785.

R Development Core Team (2008), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
URL: *http://www.R-project.org*

Rashid, M. H. and Sekiguchi, J. (1996), 'flaD (sinR) mutations affect SigD-dependent functions at multiple points in Bacillus subtilis.', *Journal of Bacteriology* **178**(22), 6640–6643.

Ratnayake-Lecamwasam, M., Serror, P., Wong, K. W. and Sonenshein, A. L. (2001), 'Bacillus subtilis CodY represses early-stationary-phase genes by sensing GTP levels.', *Genes & development* **15**(9), 1093–1103.

Rico, A. I., García-Ovalle, M., Palacios, P., Casanova, M. and Vicente, M. (2010), 'Role of Escherichia coli FtsN protein in the assembly and stability of the cell division ring', *Molecular Microbiology* **76**(3), 760–771.

Robert, L., Paul, G., Chen, Y. and Taddei, F. (2010), 'Pre-dispositions and epigenetic inheritance in the Escherichia coli lactose operon bistable switch', *Molecular systems biology* **6**, 1.

Rocha, E. P. C. (2004), 'The replication-related organization of bacterial genomes', *Microbiology* **150**(6), 1609–1627.

Romero, D., Vlamakis, H., Losick, R. and Kolter, R. (2011), 'An accessory protein required for anchoring and assembly of amyloid fibres in B. subtilis biofilms', *Molecular Microbiology* **80**(5), 1155–1168.

Rosenberg, A., Sinai, L., Smith, Y. and Ben-Yehuda, S. (2012), 'Dynamic Expression of the Translational Machinery during Bacillus subtilis Life Cycle at a Single Cell Level', *PLoS ONE* **7**(7), e41921.

Rychlik, W. and Rhoads, R. E. (1989), 'A computer program for choosing optimal oligonucleotides for filter hybridization, sequencing and in vitro amplification of DNA.', *Nucleic acids research* **17**(21), 8543–8551.

Schneider, C. A., Rasband, W. S. and Eliceiri, K. W. (2012), 'NIH Image to ImageJ: 25 years of image analysis', *Nature Methods* **9**(7), 671–675.

Scholefield, G., Whiting, R., Errington, J. and Murray, H. (2011), 'Spo0J regulates the oligomeric state of Soj to trigger its switch from an activator to an inhibitor of DNA replication initiation', *Molecular Microbiology* **79**(4), 1089–1100.

Schultz, D., Wolynes, P. G., Ben Jacob, E. and Onuchic, J. N. (2009), 'Deciding fate in adverse times: sporulation and competence in Bacillus subtilis.', *Proceedings of the National Academy of Sciences* **106**(50), 21027–21034.

Sen, M., Maillard, R. A., Nyquist, K., Rodriguez-Aliaga, P., Pressé, S., Martin, A. and Bustamante, C. (2013), 'The ClpXP Protease Unfolds Substrates Using a Constant Rate of Pulling but Different Gears', *Cell* **155**(3), 636–646.

Serizawa, M., Yamamoto, H., Yamaguchi, H., Fujita, Y., Kobayashi, K., Ogasawara, N. and Sekiguchi, J. (2004), 'Systematic analysis of SigD-regulated genes in Bacillus subtilis by DNA microarray and Northern blotting analyses', *Gene* **329**, 125–136.

Serror, P. and Sonenshein, A. L. (1996), 'CodY is required for nutritional repression of Bacillus subtilis genetic competence', *Journal of Bacteriology* **178**(20), 5910–5915.

Shah, D., Zhang, Z., Khodursky, A., Kaldalu, N., Kurg, K. and Lewis, K. (2006), 'Persisters: a distinct physiological state of E. coli.', *BMC microbiology* **6**, 53.

Sharp, P. M. and Li, W.-H. (1987), 'The codon Adaptation Index—a measure of directional synonymous codon usage bias, and its potential applications.', *Nucleic acids research* **15**(3), 1281–1295.

Shivers, R. P. and Sonenshein, A. L. (2004), 'Activation of the Bacillus subtilis global regulator CodY by direct interaction with branched-chain amino acids', *Molecular Microbiology* **53**(2), 599–611.

Sliusarenko, O., Heinritz, J., Emonet, T. and Jacobs-Wagner, C. (2011), 'High-throughput, subpixel precision analysis of bacterial morphogenesis and intracellular spatio-temporal dynamics', *Molecular Microbiology* **80**(3), 612–627.

Smith, K. and Hunter, I. S. (2008), 'Efficacy of common hospital biocides with biofilms of multi-drug resistant clinical isolates', *Journal of medical microbiology* **57**(8), 966–973.

Smith, T. J., Blackman, S. A. and Foster, S. J. (2000), 'Autolysins of Bacillus subtilis: multiple enzymes with multiple functions.', *Microbiology (Reading, England)* **146**, 249–262.

Snel, B., Lehmann, G., Bork, P. and Huynen, M. A. (2000), 'STRING: a web-server to retrieve and display the repeatedly occurring neighbourhood of a gene.', *Nucleic acids research* **28**(18), 3442–3444.

Sonenshein, A. L. (2005), 'CodY, a global regulator of stationary phase and virulence in Gram-positive bacteria', *Current opinion in microbiology* **8**(2), 203–207.

Sonenshein, A. L. (2007), 'Control of key metabolic intersections in Bacillus subtilis.', *Nature Reviews Microbiology* **5**(12), 917–927.

Sonenshein, A. L., Hoch, J. A. and Losick, R. (2002), *Bacillus Subtilis and Its Closest Relatives*, American Society Mic Series, ASM Press, Washington, DC.

Sonenshein, A. L., Hoch, J. A. and Losick, R., eds (1993), *Bacillus subtilis and Other Gram-Positive Bacteria*, American Society for Microbiology.

Srivatsan, A., Han, Y., Peng, J., Tehranchi, A. K., Gibbs, R., Wang, J. D. and Chen, R. (2008), 'High-Precision, Whole-Genome Sequencing of Laboratory Strains Facilitates Genetic Studies', *PLoS Genetics* **4**(8), e1000139.

Stewart, E. J., Madden, R., Paul, G. and Taddei, F. (2005), 'Aging and Death in an Organism That Reproduces by Morphologically Symmetric Division', *PLoS Biology* **3**(2), e45.

Tagami, K., Nanamiya, H., Kazo, Y., Maehashi, M., Suzuki, S., Namba, E., Hoshiya, M., Hanai, R., Tozawa, Y., Morimoto, T., Ogasawara, N., Kageyama, Y., Ara, K., Ozaki, K., Yoshida, M., Kuroiwa, H., Kuroiwa, T., Ohashi, Y. and Kawamura, F. (2012), 'Expression of a small (p)ppGpp synthetase, YwaC, in the (p)ppGpp^0 mutant of Bacillus subtilis triggers YvyD-dependent dimerization of ribosome', *MicrobiologyOpen* **1**(2), 115–134.

Terra, R., Stanley-Wall, N. R., Cao, G. and Lazazzera, B. A. (2012), 'Identification of Bacillus subtilis SipW as a Bifunctional Signal Peptidase That Controls Surface-Adhered Biofilm Formation', *Journal of Bacteriology* **194**(11), 2781–2790.

Tojo, S., Hirooka, K. and Fujita, Y. (2013), 'Expression of kinA and kinB of Bacillus subtilis, necessary for sporulation initiation, is under positive stringent transcription control.', *Journal of Bacteriology* **195**(8), 1656–1665.

Tsien, R. Y. (1998), 'The green fluorescent protein', *Annual Review of Biochemistry* **67**, 509–544.

Turgay, K., Hahn, J., Burghoorn, J. and Dubnau, D. (1998), 'Competence in Bacillus subtilis is controlled by regulated proteolysis of a transcription factor.', *The EMBO Journal* **17**(22), 6730–6738.

Turgay, K., Hamoen, L. W., Venema, G. and Dubnau, D. (1997), 'Biochemical characterization of a molecular switch involving the heat shock protein ClpC, which controls the activity of ComK, the competence transcription factor of Bacillus subtilis.', *Genes & development* **11**(1), 119–128.

Vagner, V., Dervyn, E. and Ehrlich, S. D. (1998), 'A vector for systematic gene inactivation in Bacillus subtilis.', *Microbiology* **144**(11), 3097–3104.

van Sinderen, D., Luttinger, A., Kong, L., Dubnau, D., Venema, G. and Hamoen, L. (1995), 'comK encodes the competence transcription factor, the key regulatory protein for competence development in Bacillus subtilis.', *Molecular Microbiology* **15**(3), 455–462.

van Sinderen, D. and Venema, G. (1994), 'comK acts as an autoregulatory control switch in the signal transduction route to competence in Bacillus subtilis.', *Journal of Bacteriology* **176**(18), 5762–5770.

Veening, J.-W., Igoshin, O. A., Eijlander, R. T., Nijland, R., Hamoen, L. W. and Kuipers, O. P. (2008a), 'Transient heterogeneity in extracellular protease production by Bacillus subtilis', *Molecular systems biology* **4**.

Veening, J.-W., Murray, H. and Errington, J. (2009), 'A mechanism for cell cycle regulation of sporulation initiation in Bacillus subtilis', *Genes & development* **23**(16), 1959–1970.

Veening, J.-W., Smits, W. K. and Kuipers, O. P. (2008b), 'Bistability, Epigenetics, and Bet-Hedging in Bacteria', *Annual Review of Biochemistry* **62**(1), 193–210.

Veening, J.-W., Stewart, E. J., Berngruber, T. W., Taddei, F., kuipers, O. P. and Hamoen, L. W. (2008c), 'Bet-hedging and epigenetic inheritance in bacterial cell development', *Proceedings of the National Academy of Sciences* **105**(11), 4393–4398.

Vischer, N., Huls, P. G. and Woldringh, C. L. (1994), 'Object-Image: an interactive image analysis program using structured point collection', *Binary* .

Vlamakis, H., Aguilar, C., Losick, R. and Kolter, R. (2008), 'Control of cell fate by the formation of an architecturally complex bacterial community', *Genes & development* **22**(7), 945–953.

Walker, K. A., Mallik, P., Pratt, T. S. and Osuna, R. (2004), 'The Escherichia coli Fis promoter is regulated by changes in the levels of its transcription initiation nucleotide CTP.', *The Journal of biological chemistry* **279**(49), 50818–50828.

Wang, J. D., Berkmen, M. B. and Grossman, A. D. (2007a), 'Genome-wide coorientation of replication and transcription reduces adverse effects on replication in Bacillus subtilis', *Proceedings of the National Academy of Sciences* **104**, 5608–5613.

Wang, J. D., Sanders, G. M. and Grossman, A. D. (2007b), 'Nutritional control of elongation of DNA replication by (p)ppGpp', *Cell* **128**(5), 865–875.

Wang, W. and Bechhofer, D. H. (1996), 'Properties of a Bacillus subtilis polynucleotide phosphorylase deletion strain.', *Journal of Bacteriology* **178**(8), 2375–2382.

Wendrich, T. M., Blaha, G., Wilson, D. N., Marahiel, M. A. and Nierhaus, K. H. (2002), 'Dissection of the mechanism for the stringent factor RelA', *Molecular cell* **10**(4), 779–788.

Wendrich, T. M. and Marahiel, M. A. (1997), 'Cloning and characterization of a relA/spoT homologue from Bacillus subtilis', *Molecular Microbiology* **26**(1), 65–79.

Werner, R. (1971), 'Nature of DNA precursors', *Nature New Biology* **233**(38), 99–103.

Wilkinson, D. J. (2009), 'Stochastic modelling for quantitative description of heterogeneous biological systems', *Nature Reviews Genetics* **10**(2), 122–133.

Winkelman, J. T., Bree, A. C., Bate, A. R., Eichenberger, P., Gourse, R. L. and Kearns, D. B. (2013), 'RemA is a DNA-binding protein that activates biofilm matrix gene expression in Bacillus subtilis', *Molecular Microbiology* **88**(5), 984–997.

Withey, J. H. and Friedman, D. I. (2003), 'A salvage pathway for protein synthesis: tmRNA and trans-translation', *Annual Review of Microbiology* **57**(1), 101–123.

Wünsche, A., Hammer, E., Bartholomae, M., Völker, U., Burkovski, A., Seidel, G. and Hillen, W. (2012), 'CcpA forms complexes with CodY and RpoA in Bacillus subtilis', *FEBS Journal* **279**(12), 2201–2214.

Young, J. W., Locke, J. C. W., Altinok, A., Rosenfeld, N., Bacarian, T., Swain, P. S., Mjolsness, E. and Elowitz, M. B. (2011), 'Measuring single-cell gene expression dynamics in bacteria using fluorescence time-lapse microscopy', *Nature Protocols* **7**(1), 80–88.

Youngman, P., Perkins, J. B. and Losick, R. (1984), 'Construction of a cloning site near one end of Tn917 into which foreign DNA may be inserted without affecting transposition in Bacillus subtilis or expression of the transposon-borne erm gene', *Plasmid* **12**(1), 1–9.

Zafra, O., Lamprecht-Grandío, M., de Figueras, C. G. and González-Pastor, J. E. (2012), 'Extracellular DNA Release by Undomesticated Bacillus subtilis Is Regulated by Early Competence', *PLoS ONE* **7**(11), e48716.

Zeigler, D. R., Prágai, Z., Rodriguez, S., Chevreux, B., Muffler, A., Albert, T., Bai, R., Wyss, M. and Perkins, J. B. (2008), 'The origins of 168, W23, and other Bacillus subtilis legacy strains', *Journal of Bacteriology* **190**(21), 6983–6995.

Zenobi, R. (2013), 'Single-Cell Metabolomics: Analytical and Biological Perspectives', *Science* **342**(6163), 1243259–1243259.

Zhou, J., Lloyd, S. A. and Blair, D. F. (1998), 'Electrostatic interactions between rotor and stator in the bacterial flagellar motor.', *Proceedings of the National Academy of Sciences of the United States of America* **95**(11), 6436–6441.

Österberg, S., del Peso-Santos, T. and Shingler, V. (2011), 'Regulation of Alternative Sigma Factor Use', *Annual Review of Microbiology* **65**(1), 37–55.