Using high-throughput approaches to unveil sporulation and germination mechanisms in *Clostridium sporogenes*

Raquel Sofia Mesquita Rodrigues, MSc

Thesis submitted to the University of Nottingham
for the degree of Doctor of Philosophy

September, 2019



Declaration

Unless otherwise acknowledged, the work presented in this thesis is my own. No part has been submitted for another degree in the University of Nottingham or any other institute of learning.

Raquel Sofia Mesquita Rodrigues,
September 2019

Abstract

Intravenously injected clostridial spores infiltrate and selectively germinate in the hypoxic regions of solid tumours. The exquisite selectivity of this natural phenomenon is being exploited to bring about tumour destruction by engineering the delivered clostridia to enzymatically convert a subsequently injected circulatory prodrug into a highly cytotoxic species. This approach is termed 'Clostridial-Directed Enzyme Prodrug Therapy' (CDEPT), and its leading chassis is *Clostridium sporogenes*, a non-pathogenic member of the genus and a prolific producer of spores.

Despite its central role in CDEPT, spore formation and germination in *C. sporogenes* remain relatively poorly understood. A better understanding would allow the implementation of further refinements to CDEPT, in particular the creation of a conditionally sporulating strain. Accordingly, to gain greater insight into these two developmental processes, TraDIS (Transposon Directed Insertion-site Sequencing) and RNA-Seq were undertaken. The former to identify those genes essential to sporulation and germination, the latter to identify the regulons of the main sporulation sigma factors, σ^F, σ^E, σ^G and σ^K.

The implementation of TraDIS initially required improvements to the frequency of DNA transfer into *C. sporogenes* NCIMB 10696 through circumvention of the barriers posed by two native restriction modification systems. Conjugation efficiencies were increased more than 10-fold using *E. coli* sExpress as the conjugative donor. The TraDIS library made contained 118,894 unique insertion sites, equivalent to one transposon insertion every 35 bp. A total of 471 genes were identified as essential for the survival of the organism, corresponding to almost 13% of the annotated genes.

Some 183 and 142 genes were considered essential for sporulation and germination, respectively. Some of these genes identified as essential were further characterised, namely *sigF*, *sigE*, *sigG*, *sigK*, *spoIIAA*, *spoIIE*, *spoIIR*, *gerKA4*, CLSPO_c22010 and CLSPO_c33910.

Overall, this work shed some light on sporulation and germination mechanisms in *C. sporogenes* and highlighted some similarities and differences relative to other spore formers. Furthermore, the genes identified using TraDIS as essential for sporulation can be explored in the future regarding their use in the generation of a conditionally sporulating strain.

Acknowledgements

I would like to thank my main supervisor, Professor Nigel Minton, for the opportunity to work in his lab, as well as for the guidance and support during the course of the PhD. To my second supervisor, Dr Alan Cockayne, thank you for the continuous support, for all the suggestions and for the patience to hear my frequent frustrations. And to my bench supervisor, Dr Aleksandra Kubiak, thanks for welcoming me in the lab and teaching me so many useful techniques.

This work would not have been possible without the funding from the European Union (Marie Skłodowska-Curie Actions) – to all those involved in the Clospore network, thank you.

I also thank Dr Robert Mansfield for the help with the restriction systems work, Dr Craig Woods for the help with the TraDIS work and Dr Andrew Dempster for the help with the RNA-Seq work. The high-throughput approaches in particular were very challenging, and I couldn't have done it without your help. I also appreciate all the help and support from my friends and colleagues in the SBRC. You certainly made this journey easier and fun. And thank you Evangelos for always preparing everything I needed for those massive experiments.

To Inés, Maria and Daphne, a very special thank you. Not only are you incredible scientists, but also amazing people. I learned a lot from all of you, and your advice and support made this thesis possible. These last 4 years wouldn't have been the same without you, and your friendship made it all worth it.

Para os meus pais. Não há palavras para agradecer tudo o que fizeram por mim, tudo o sacrificaram por mim. Devo a pessoa que sou a vocês, e não teria chegado aqui se não tivesse o vosso apoio incondicional. Ao meu irmão Bruno, obrigada por acreditares que mereci chegar aqui. E obrigada pelo meu sobrinho Daniel, que me trouxe bastante alegria numa altura em que só conseguia pensar em trabalho.

E para o meu namorado Tiago... Sei que os últimos anos não foram fáceis, com muito stress e imensas frustrações. Mas saber que te tenho do meu lado, saber que caminhamos e crescemos juntos, torna tudo bem mais fácil. Tu vês o melhor em mim e acreditas em mim, mesmo quando eu não consigo. You trully raise me up.

Para a minha madrinha, o meu primo e um grande amigo... Onde estiverem, espero que tenham orgulho em mim.

Table of contents

Declaration ..... **i**

Abstract ..... **ii**

Acknowledgements ..... **iii**

Table of contents ..... **iv**

List of Figures ..... **xii**

List of Tables ..... **xvi**

List of Abbreviations ..... **xvii**

Chapter 1: General introduction **1**

1.1 Introduction to spore formers 2

1.1.1 Clostridial spores as a threat 2

1.1.2 Clostridial spores as an opportunity 3

1.1.2.1 Clostridial-Directed Enzyme Prodrug Therapy (CDEPT) 3

1.2 Mechanisms of sporulation and germination 5

1.2.1 Stages of sporulation 5

1.2.2 Morphology of spores 7

1.2.3 Regulation of sporulation 9

1.2.3.1 Spo0A 9

1.2.3.2 Sigma factors 11

1.2.3.3 Other regulatory proteins 11

1.2.4 Germination 12

1.2.4.1 Initiation of germination 12

1.2.4.2 Degradation of the cortex 13

1.2.4.3 Resumption of metabolic activity 13

1.3 Hypothesis and aims 15

Chapter 2: Materials and methods **16**

2.1 Microbiology techniques 17
2.1.1 Growth media, buffers and supplements 17
2.1.2 Bacterial strains and vectors 18
2.1.3 General culture and storage conditions 19
2.1.4 Quantification of bacterial growth 19
2.1.5 DNA transfer into *E. coli* strains 19
2.1.5.1 Preparation of electro-competent cells of *E. coli* 19
2.1.5.2 Transformation of electro-competent *E. coli* cells with DNA by electroporation 20
2.1.6 DNA transfer into *C. sporogenes* strains 20
2.1.6.1 Transfer of plasmids from an *E. coli* donor into *C. sporogenes* by conjugation 20
2.1.6.2 Preparation of electro-competent cells of *C. sporogenes* 21
2.1.6.3 Transformation of electro-competent cells of *C. sporogenes* with DNA by electroporation 21
2.1.7 Plasmid loss 21
2.2 Molecular biology techniques 22
2.2.1 Extraction and purification of plasmid DNA 22
2.2.2 Extraction and purification of genomic DNA 22
2.2.3 Extraction and purification of total RNA 22
2.2.4 Reverse transcription of RNA into cDNA 23
2.2.5 Amplification of DNA 24
2.2.5.1 Polymerase Chain Reaction (PCR) 24
2.2.5.2 Splicing by overlap extension (SOE) PCR 24
2.2.5.3 Colony PCR 24
2.2.5.4 Inverse PCR 24
2.2.5.5 RT-PCR 25
2.2.5.6 RT-qPCR 25

2.2.5.7 List of primers 26
2.2.6 Agarose gel electrophoresis 26
2.2.6.1 Extraction of DNA from agarose gels 26
2.2.7 Restriction endonuclease digestion 26
2.2.8 Ligation and membrane dialysis of DNA 26
2.2.9 DNA sequencing 26
2.2.10 Southern blot 27
2.3 Phenotypic characterisation 29
2.3.1 Sporulation assay 29
2.3.2 Growth curves 29
2.3.3 Conjugation efficiency assays 29
2.3.4 Transformation efficiency assays 30
2.3.5 Microscopy 30
2.4 Methods used to assess gene expression levels 31
2.4.1 Enzymatic assay of chloramphenicol acetyltransferase (CAT assay) 31
2.4.1.1 Construction of vectors tested with the CAT assay 32
2.4.2 Red fluorescence protein (RFP) reporter assay 32
2.4.2.1 Construction of vectors tested with the RFP assay 33
2.5 Genome editing 33
2.5.1 ClosTron 34
2.5.2 CRISPR/Cas9 34
2.5.2.1 Construction of CRISPR/Cas9 vectors 35
2.5.2.2 Construction of RiboCas vectors 36
2.5.2.3 Construction of vectors used to revert each target gene to the original locus 37
2.6 Cloning of vectors used in Chapter 3 38
2.7 TraDIS 39

2.7.1 Spore purification 40
2.7.2 Sequencing library preparation 40
2.7.3 Analysis and processing of sequencing data 42
2.8 RNA-Seq 42
2.9 Bioinformatics 43
Chapter 3: Overcoming restriction systems to facilitate gene transfer 44
3.1 Background 45
3.1.1 Importance of gene transfer for this work 45
3.1.2 The role of restriction systems as a barrier to gene transfer 45
3.1.2.1 Restriction systems in *Clostridium sporogenes* 46
3.1.2.2 Type I restriction systems 47
3.1.2.3 Type IV restriction systems 48
3.1.3 Strategies to overcome restriction 49
3.1.3.1 ClosTron 50
3.1.3.2 CRISPR/Cas9 51
3.2 Chapter overview 53
3.3 Results 54
3.3.1 Overcoming the type IV restriction system 54
3.3.1.1 Targeting the type IV system using ClosTron 54
3.3.1.2 Targeting the type IV system using CRISPR/Cas9 60
3.3.1.3 Testing an *E. coli* strain that avoids methylation in cytosine nucleotides: sExpress 63
3.3.1.4 Comparison between the different approaches to overcome the type IV restriction system 65
3.3.2 Overcoming the type I restriction system 67
3.3.2.1 Targeting the type I system using ClosTron 67
3.3.2.2 Targeting the type I system using CRISPR/Cas9 70

3.3.2.3 Changing the methylation pattern of the DNA before transfer to avoid restriction 72

3.3.2.4 Identification of potential type I recognition sites 76

3.3.2.5 A final attempt 76

3.4 Discussion 79

Chapter 4: Identification of genes essential for sporulation or germination in *Clostridium sporogenes* 81

4.1 Background 82

4.1.1 TraDIS: a high-throughput approach to identify essential genes 82

4.1.1.1 Use of tetracycline inducible promoters to regulate the expression of relevant genes 83

4.2 Chapter overview 85

4.3 Results 86

4.3.1 Selection of the best plasmid to mediate transposon mutagenesis 86

4.3.1.1 Optimisation of expression of transposase and induction of plasmid loss 86

4.3.1.2 Enzymatic assay of chloramphenicol acetyltransferase 88

4.3.1.3 Description of the plasmid used to generate the transposon mutant library 90

4.3.2 Comparison between two donor strains and determination of the number of conjugations required to obtain one million mutants 91

4.3.3 Optimisation of a spore purification protocol 93

4.3.4 Generation of the transposon mutant library 96

4.3.5 Analysis of the TraDIS results 97

4.3.5.1 Identification of genes essential for sporulation 101

4.3.5.2 Identification of genes essential for germination 105

4.4 Discussion 109

Chapter 5: Regulation of the sporulation cascade by the sigma factors σ^E, σ^F, σ^G and σ^K 116

5.1 Background 117
5.1.1 Activation of sigma factors and their role in regulation of sporulation 117
5.1.1.1 Co-regulation of genes by more than one sigma factor/regulatory gene 120
5.1.2 RNA-Seq: a high-throughput approach to transcriptomics 120
5.1.3 Other methods to assess gene expression 121
5.1.4 RiboCas: a gene editing tool used to target the genes encoding the sigma factors 123
5.2 Chapter overview 123
5.3 Results 124
5.3.1 Deletion of the genes encoding σ^F, σ^E, σ^G and σ^K 124
5.3.2 Identification of the time points for RNA extraction 125
5.3.2.1 RT-PCR 125
5.3.2.2 RT-qPCR 127
5.3.3 Identification of the regulon of each sigma factor 128
5.3.4 Regulation of expression and activity of the sigma factors 130
5.3.4.1 The σ^F regulon 132
5.3.4.2 The σ^E regulon 134
5.3.4.3 The σ^G regulon 136
5.3.4.4 The σ^K regulon 139
5.3.4.5 Other regulatory proteins 140
5.4 Discussion 142
Chapter 6: Characterisation of genes involved in sporulation and/or germination in *C. sporogenes* 145
6.1 Background 146
6.1.1 Brief overview of the selected genes 146
6.1.1.1 Sigma factors 146
6.1.1.2 CLSPO_c31830: *spoIIAA* 146
6.1.1.3 CLSPO_c37040: *spoIIE* 147

6.1.1.4 CLSPO_c01510: *spoIIR* 149
6.1.1.5 CLSPO_c01480: *gerKA4* 151
6.1.1.6 CLSPO_c22010: M24 family metallopeptidase (APP) 151
6.1.1.7 CLSPO_c33910: Hypothetical protein 152
6.2 Chapter overview 152
6.3 Results 154
6.3.1 Creation of the desired mutants 154
6.3.1.1 Deletion of selected genes using RiboCas 154
6.3.1.2 Insertion of ClosTron in selected genes 156
6.3.2 Characterisation of the sigma factor mutants 156
6.3.3 Characterisation of the *spoIIAA* mutant 162
6.3.4 Characterisation of the APP mutant 165
6.3.5 Characterisation of the CLSPO_c33910 mutant 167
6.3.6 Sporulation ability of other selected genes 171
6.3.6.1 Characterisation of the *spoIIE* mutant 171
6.3.6.2 Characterisation of the *spoIIR* mutant 172
6.3.6.3 Characterisation of the *gerKA4* mutant 173
6.4 Discussion 175
Chapter 7: General conclusions 177
Bibliography 188
Appendix 200
Chapter 2: Supporting information 201
List of bacterial strains 201
List of vectors 203
List of primers 206
Vectors used in the RFP reporter assay 217

Vector used to express the *hsdS* and *hsdM* genes from the type I restriction system 218

Vectors containing sequences potentially recognised by the type I restriction system 219

CRISPR/Cas9 vectors 220

RiboCas vectors 221

RiboCas vectors used to revert genes to their original locus 222

Chapter 4: Supporting information 223

Insertion index distributions 223

List of genes essential for survival 226

List of genes essential for vegetative growth 243

List of genes essential for sporulation 251

List of genes essential for germination 261

Chapter 5: Supporting information 269

RNA quality control 269

qPCR efficiencies 271

RNA-Seq: Sequencing data 272

RNA-Seq: PCA analysis 273

RNA-Seq: Heatmaps 277

RNA-Seq: List of genes from the σ^F regulon 278

RNA-Seq: List of genes from the σ^E regulon 280

RNA-Seq: List of genes from the σ^G regulon 289

RNA-Seq: List of genes from the σ^K regulon 295

List of Figures

Figure 1.1: Representation of the CDEPT therapy. 4
Figure 1.2: Stages of sporulation. 6
Figure 1.3: Spore structure. 8
Figure 1.4: Representation of the phosphorelay system in *B. subtilis*. 10
Figure 2.1: Apparatus used to transfer the *Hin*dIII-digested DNA from the agarose gel onto a Hybond H^+ nitrocellulose membrane. 28
Figure 2.2: Schematic representation of the primer binding sites for colony PCR to screen for the insertion of the ClosTron in the target gene. 34
Figure 2.3: Schematic representation of the SOEing PCR method used to join the homology arms. 37
Figure 2.4: Schematic representation of the SOEing PCR method used to introduce a SNP in the stop codon. 38
Figure 3.1: Restriction systems in *C. sporogenes* NCIMB 10696. 46
Figure 3.2: Functional domains identified in the DNA repair helicase Rad25, a putative type IV restriction system from *C. sporogenes*. 48
Figure 3.3: Summary of the strategies that can be used to overcome the restriction barriers found in *C. sporogenes*. 50
Figure 3.4: Representation of the ClosTron system. 51
Figure 3.5: Representation of the CRISPR/Cas9 system. 52
Figure 3.6: Representation of the CLSPO_c06750 gene with and without the intron insertion. 55
Figure 3.7: Sporulation profile of the *C. sporogenes*-c06750-1240a::CT strain. 56
Figure 3.8: Comparison between conjugation efficiencies in *C. sporogenes* WT and *C. sporogenes*-c06750-1240a::CT. 58
Figure 3.9: Comparison between transformation efficiencies in *C. sporogenes* WT and *C. sporogenes*-c06750-1240a::CT. 59
Figure 3.10: Gel electrophoresis of colony PCR samples demonstrating the insertion of the ClosTron into CLSPO_c06750 (CLSPO_c06750::CT) and the in-frame deletion of the same gene using CRISPR/Cas9 (ΔCLSPO_c06750). 61
Figure 3.11: Sporulation profile of the *C. sporogenes* Δc06750 strain. 61

Figure 3.12: Comparison between conjugation efficiencies in *C. sporogenes* WT and *C. sporogenes* Δc06750. 62
Figure 3.13: Comparison between conjugation efficiencies using *E. coli* CA434 or *E. coli* sExpress as donor strains. 64
Figure 3.14: Comparison between conjugation efficiencies using *E. coli* CA434 or *E. coli* sExpress as donor strains, and *C. sporogenes* WT, *C. sporogenes*-c06750-1240a::CT or *C. sporogenes* Δc06750 as recipient strains. 66
Figure 3.15: A: Gel electrophoresis of colony PCR samples demonstrating the insertion of the ClosTron into *hsdR*. 68
Figure 3.16: Sporulation profile of the *C. sporogenes-hsdR*1537a::CT strain. 68
Figure 3.17: Comparison between conjugation efficiencies in *C. sporogenes* WT and *C. sporogenes-hsdR*1537a::CT. 69
Figure 3.18: Gel electrophoresis of colony PCR samples screening the in-frame deletion of the *hsdR* gene. 71
Figure 3.19: Fluorescence levels of RFP driven by different promoters. 73
Figure 3.20: Comparison between conjugation efficiencies using *E. coli* sExpress or *E. coli* sExpress carrying pMTL80221-P_{BAD}-*hsdSM* as donor strains. 75
Figure 3.21: Comparison between conjugation efficiencies using *E. coli* sExpress or *E. coli* sExpress carrying pMTL80221-PBAD-*hsdSM* as donor strains. 77
Figure 4.1: Schematic representation of the steps required for TraDIS. 83
Figure 4.2: Schematic representation of the tetracycline inducible system. 85
Figure 4.3: Schematic representation of the inducible system used by Dembek *et al.*[114]and in the present study. 87
Figure 4.4: Schematic representation of the Tet inducible system from pMTL-tet3nO. 87
Figure 4.5: Schematic representation of the Tet inducible system from pMTL-tet3nO_RFtetR. 88
Figure 4.6: CAT assay comparing enzymatic activity resulting from expression of *catP*. 89
Figure 4.7: pMTL-CW21 vector used for the generation of the transposon mutant library. 91
Figure 4.8: Comparison between conjugation efficiencies using *E. coli* CA434 or *E. coli* sExpress as donor strains to transfer pMTL-CW21 into *C. sporogenes* WT. 92

Figure 4.9: Microscopic visualisation of *C. sporogenes* spores following spore purification using Sucrose or HistoDenz gradients. 95
Figure 4.10: Schematic representation of the strategy used to identify genes essential for sporulation and germination in *C. sporogenes*. 96
Figure 4.11: Plot demonstrating the identification of genes essential for survival of *C. sporogenes*. 99
Figure 4.12: Plot showing the difference in mutant frequencies in the spores relative to vegetative cells. 102
Figure 4.13: Plot showing the difference in mutant frequencies in germinated cells relative to spores. 106
Figure 5.1: Schematic representation of the criss-cross regulation of sigma factors. 118
Figure 5.2: Flowchart representing the different steps usually required to perform RNA-Seq experiments in bacteria. 121
Figure 5.3: Representation of how fluorescence increases with cycle number in qPCR experiments. 122
Figure 5.4: Representation of the colony PCR method used to confirm the deletion of the *sigF* gene, as an example. 124
Figure 5.5: Gel electrophoresis showing changes in expression levels over time (RT-PCR). 126
Figure 5.6: RT-qPCR results. 128
Figure 5.7: Overview of the differential gene expression between sigma factor mutants and wild type. 129
Figure 6.1: Schematic representation of the σ^F activation mechanism. 147
Figure 6.2: Schematic representation of the σ^E activation mechanism. 150
Figure 6.3: Schematic representation of the strategy used for reversion of the gene to the original locus. 154
Figure 6.4: Gel electrophoresis picture of colony PCR samples demonstrating the deletion and reversion of the target genes. 155
Figure 6.5: Evaluation of the ClosTron mutants. 156
Figure 6.6: Sporulation profile of the *C. sporogenes* Δ*sigF*, *C. sporogenes* Δ*sigF* + *sigF*, *C. sporogenes* Δ*sigE* and *C. sporogenes* Δ*sigE* + *sigE* strains. 158
Figure 6.7: *C. sporogenes* strains stained using the Schaeffer-Fulton method. 159

Figure 6.8: Sporulation profile of the *C. sporogenes* Δ*sigG*, *C. sporogenes* Δ*sigG* + *sigG*, *C. sporogenes* Δ*sigK* and *C. sporogenes* Δ*sigK* + *sigK* strains. 160
Figure 6.9: *C. sporogenes* strains stained using the Schaeffer-Fulton method. 161
Figure 6.10: Growth profiles of *C. sporogenes* WT, *C. sporogenes* Δ*sigF*, *C. sporogenes* Δ*sigF* + *sigF*, *C. sporogenes* Δ*sigE*, *C. sporogenes* Δ*sigE* + *sigE*, *C. sporogenes* Δ*sigG*, *C. sporogenes* Δ*sigG* + *sigG*, *C. sporogenes* Δ*sigK* and *C. sporogenes* Δ*sigK* + *sigK*. 162
Figure 6.11: Characterisation of *C. sporogenes* Δ*spoIIAA* and *C. sporogenes* Δ*spoIIAA* + *spoIIAA* strains. 163
Figure 6.12: *C. sporogenes* strains stained using the Schaeffer-Fulton method. 164
Figure 6.13: Characterisation of *C. sporogenes* ΔAPP and *C. sporogenes* ΔAPP + APP strains. 166
Figure 6.14: *C. sporogenes* strains stained using the Schaeffer-Fulton method. 167
Figure 6.15: Characterisation of *C. sporogenes* Δc33910 and *C. sporogenes* Δc33910 + c33910 strains. 169
Figure 6.16: *C. sporogenes* strains stained using the Schaeffer-Fulton method. 170
Figure 6.17: Sporulation profile of the *C. sporogenes-spoIIE*2368s::CT strain. 171
Figure 6.18: Sporulation profile of the *C. sporogenes-spoIIR*335s::CT strain. 172
Figure 6.19: Sporulation profile of the *C. sporogenes-gerKA4*775s::CT strain. 174
Figure 7.1: Schematic representation of the regulation of sigma factors in *C. sporogenes*. 181

List of Tables

Table 2.1: Compositions of media and buffers used in this study. 17
Table 2.2: List of supplements used in this study. 18
Table 2.3: Components used in the reaction for reverse transcription of RNA into cDNA, and respective volumes (Omniscript Kit, Qiagen). 23
Table 2.4: Inverse PCR components and conditions. 25
Table 2.5: sgRNA seed sequences designed using Benchling. 35
Table 2.6: sgRNA seed sequences designed using Benchling. 36
Table 2.7: List of primers used in the preparation of the sequencing library for TraDIS. 41
Table 2.8: List of bioinformatic tools used in this study. 43
Table 4.1: Transposition ratio and percentage of plasmid loss obtained in the preliminary TraDIS experiment. 93
Table 4.2: Summary of the data obtained from the MiSeq Illumina sequencing. 99
Table 4.3: KEGG pathway analysis of essential genes. 100
Table 4.4: Partial list of candidate genes essential for sporulation in *C. sporogenes*. 104
Table 4.5: Partial list of candidate genes essential for germination in *C. sporogenes*. 107
Table 4.6: List of genes that are conserved and relevant for sporulation in *B. subtilis*, *C. sporogenes* and *C. difficile*. 111
Table 5.1: Part of the σ^F regulon from *C. sporogenes*.. 133
Table 5.2: Part of the σ^E regulon from *C. sporogenes*. 135
Table 5.3: Part of the σ^G regulon from *C. sporogenes*. 137
Table 5.4: Part of the σ^K regulon from *C. sporogenes*. 140
Table 7.1: List of genes essential for sporulation/germination identified with TraDIS and RNA-Seq. 183

List of Abbreviations

AdoMet	S-Adenosylmethionine
AGFK	Mixture of asparagine, glucose, fructose and K^+
APP	Aminopeptidase P
ATCC	American type culture collection
ATP	Adenosine triphosphate
BLAST	Basic local alignment search tool
bp	Base pair
Cas	CRISPR-associated
CAT	Chloramphenicol acetyltransferase
CDEPT	Clostridial-directed enzyme prodrug therapy
cDNA	Complementary DNA
CFU	Colony forming unit
CRISPR	Clustered regularly interspaced palindromic repeat
crRNA	CRISPR RNA
Dcm	DNA cytosine methyltransferase
dH_2O	Distilled water
DMSO	Dimethyl sulphoxide
DNA	Deoxyribonucleic acid
DPA	Dipicolinic acid
GMO	Genetically modified organism
Hsd	Host specificity for DNA
ITR	Inverted terminal repeat sequences
Kb	Kilobase
KEGG	Kyoto encyclopedia of genes and genomes
LB	Luria Bertani
LHA	Left homology arm
mRNA	Messenger RNA
NCIMB	National collection of industrial food and marine bacteria
NEB	New England Biolabs
NRT	No-reverse transcription
NTC	No-template control

OD_{600}	Optical density at 600 nm
PAM	Protospacer adjacent motif
PBS	Phosphate-buffered saline
PCA	Principal component analysis
PCE	Prodrug converting enzyme
PCR	Polymerase chain reaction
RAM	Retrotransposition-activated marker
RBS	Ribosome binding site
RFP	Red fluorescent protein
RHA	Right homology arm
RIN	RNA integrity number
RNA	Ribonucleic acid
RNA-Seq	RNA sequencing
rRNA	Ribosomal RNA
RT-PCR	Reverse transcription PCR
RT-qPCR	Quantitative reverse transcription PCR
SASPs	Small acid-soluble proteins
sgRNA	Single-guide RNA
skin	Sigma K intervening
SNP	Single nucleotide polymorphism
SNVs	Single nucleotide variants
SOE	Splicing by overlap extension
TEM	Transmission electron microscopy
Tet	Tetracycline-inducible
TraDIS	Transposon directed insertion-site sequencing
trCas9	Truncated Cas9
tRNA	Transfer RNA
UIS	Unique insertion sites
UV	Ultraviolet radiation
WT	Wild type

Chapter 1: General introduction

1

1.1 Introduction to spore formers

The bacterial endospore is one of the most highly resistant life-forms on earth and allows the bacterium to survive exposure to extremes of temperature, desiccation, radiation, disinfectants and, in the case of *Clostridium*, oxygen. In these remarkable structures the DNA is protected by several proteins within the core, which in turn is protected by layers of membrane, coat, cortex and, in some cases, exosporium[1] (Figure 1.3). Spores are not only resistant to extreme conditions on earth but also to the harsh environment of space[2]. Their protective layers are so effective that they allow them to survive for many years, possibly millions[3, 4]. The downside of sporulation is that the spores are metabolically dormant. However, when the environment is no longer a threat to cell survival, the spores are able to germinate and resume metabolic activity[5]. Several spore formers have been identified over the years, and most of them are Firmicutes. Accordingly, several bacteria belonging to the taxonomic classes Bacilli and Clostridia are Gram-positive and able to form spores[6]. The main difference between the two classes is that Clostridia are generally anaerobic bacteria, while Bacilli can be obligate or facultative aerobes[7, 8]. In addition, some differences regarding the formation and development of spores have been observed[6, 9, 10]. The process of sporulation has been extensively studied in *Bacillus subtilis*[11-13], so this species usually serves as model for sporulation mechanisms and spore structure[14, 15]. Conversely, the mechanisms of spore formation are less clear in Clostridia.

1.1.1 Clostridial spores as a threat

Clostridia are mostly known for the pathogenicity of some of its species. *Clostridium botulinum*, for instance, can germinate in food, wounds or the intestine and produce a lethal toxin that causes botulism[16, 17]. *Clostridioides difficile* (formerly *Clostridium difficile*), on the other hand, is usually associated with antibiotic treatment. This results in an altered intestinal flora, allowing *C. difficile* endospores to germinate and cause diarrhoea and sometimes pseudomembranous colitis[17, 18]. Other examples of diseases caused by *Clostridium* are food poisoning (*Clostridium perfringens*) and tetanus (*Clostridium tetani*)[17]. Due to their ability to form spores, these pathogens are very hard to eliminate and pose a serious threat to human health[15].

1.1.2 Clostridial spores as an opportunity

Despite the threats described above, most *Clostridium* species are non-pathogenic and can actually have some interesting applications. Some of these species are known to produce commercially relevant solvents (*Clostridium acetobutylicum*, *Clostridium stercorarium* and *Clostridium thermocellum*)[19-21]. Furthermore, *Clostridium cellulolyticum* has been investigated for its ability to degrade cellulose[22].

Recently, non-pathogenic clostridia have been explored for their use in cancer therapy, as described below. One of the species of interest for this therapy is *Clostridium sporogenes*, which is the focus of this study. Like other *Clostridium* species, *C. sporogenes* is rod-shaped, anaerobic, produces endospores and can be found in a variety of environments, including soil and human/animal intestines. This species is considered to be safe, being classified as a harmless hazard group I organism by the UK Advisory Committee on Dangerous Pathogens and as a harmless biosafety level 1 organism by the American Type Culture Collection (ATCC 3584)[23, 24]. In fact, *C. sporogenes* is often used as a surrogate for *C. botulinum* in food industry quality assessments (the two species show many similarities, but *C. sporogenes* does not produce botulinum toxin)[25, 26]. The genome of *C. sporogenes* NCIMB 10696, the strain used in this study, has been sequenced and is available in the GenBank database under the accession number CP009225[24].

1.1.2.1 Clostridial-Directed Enzyme Prodrug Therapy (CDEPT)

Clostridial-Directed Enzyme Prodrug Therapy (CDEPT) involves the use of non-pathogenic strains of *Clostridium* as a vector for delivering anti-cancer drugs directly to solid tumour cells[27]. CDEPT was developed because traditional cancer therapies usually cause unwanted side effects. In fact, the drugs that are commonly used in chemotherapy frequently target essential cellular processes, such as DNA replication, mitosis or cell multiplication. However, healthy cells can also be affected by these drugs, leading to side effects[28]. To overcome these problems, the concept of a prodrug was developed: a compound that is delivered to the tumour in an inactive form, and only there is converted to a highly toxic drug[29]. CDEPT involves the administration of a prodrug alongside a prodrug converting enzyme (PCE)[27]. Ideally, prodrug conversion occurs only in the tumour, leaving the healthy tissue unharmed. This mechanism is represented in Figure 1.1.

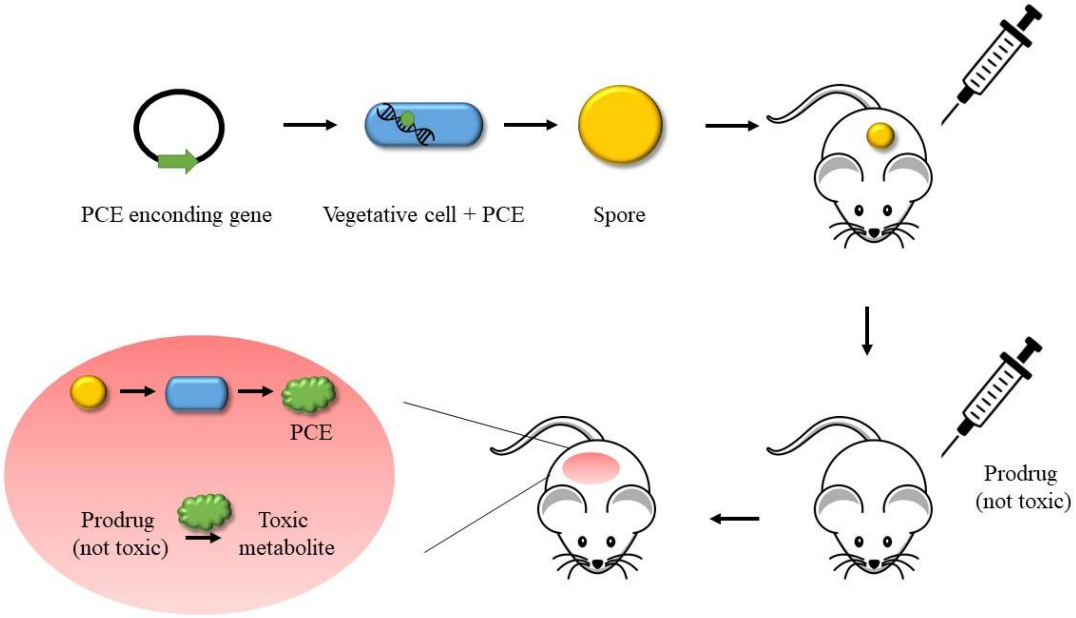


Figure 1.1: Representation of the CDEPT therapy. Spores of an engineered strain of *C. sporogenes* are administered intravenously to a tumour-bearing animal. The engineered strain contains a gene encoding a prodrug converting enzyme (PCE). Then, a prodrug is also delivered to the animal. In the tumour, the spores are able to germinate and express the PCE, which converts the prodrug into a toxic derivative. Adapted from Kubiak *et al.*[27].

Nitroreductases can be used as prodrug converting enzymes. HsoNTR and NmeNTR, for instance, are able to reduce the 4-nitro group of the anti-tumour prodrug CB1954, leading to the formation of the 4-hydroxylamine derivative. This reduced and active form of the drug shows much higher toxicity than the prodrug[30].

An advantage of the CDEPT approach is the fact that it uses obligate anaerobes as vectors to deliver the PCE to hypoxic tumour tissues. Hypoxia is common in tumours and is a result of the greater distance between some tumour tissues and blood vessels, making it difficult to treat using conventional methods (*e.g.*, radio- and chemotherapy)[31]. The administration of *Clostridium* spores to cancer patients has been reported to cause tumour regression (due their oncolytic properties) and target hypoxic regions, as spores germinate only in poorly oxygenated areas[32, 33]. Spores from *Clostridium* species (including *C. sporogenes*) have hence been studied as vectors to deliver PCE to tumour hypoxic tissues. In fact, the administration of engineered *C. sporogenes* spores along with the pro-drug CB1954 resulted in reduced tumour volume in mice[30].

1.2 Mechanisms of sporulation and germination

As previously mentioned, *B. subtilis* is the model species for sporulation and germination studies. Consequently, most of the structural features presented in the following section were studied in this organism, although at least some of them have also been identified in clostridia. Unless otherwise stated, all the spore structures and mechanisms described in this thesis correspond to *B. subtilis*.

1.2.1 Stages of sporulation

Spore formation is a highly regulated process and requires several morphological changes. First, the cell undergoes asymmetric division, leading to the formation of a forespore (which will later become the spore itself) and a mother cell. The mother cell and the forespore are initially separated by a septum, but eventually the forespore is engulfed by the mother cell, which will aid in the development of the spore. When the morphogenesis is complete, the mother cell lyses and releases the fully developed spore[1, 10, 12]. In order to facilitate its study, this process has been divided into different stages[34] (Figure 1.2):

- Stage 0 – vegetative cell: this stage precedes cell division and involves the initiation of DNA replication;
- Stage I – axial filamentation: the replicating chromosomes stretch and form a filament across the long axis of the cell;
- Stage II – asymmetric division: a septum is formed near one of the poles, separating the forespore from the mother cell;
- Stage III – engulfment: the mother cell envelope surrounds the forespore;
- Stage IV – cortex synthesis: in addition to the cell wall, another peptidoglycan layer (cortex) forms around the forespore;
- Stage V – coat synthesis: specialised proteins form an additional layer surrounding the forespore, the coat;
- Stage VI – maturation: the spore becomes resistant to several conditions, such as UV radiation and heat;
- Stage VII – mother cell lysis: the mature spore is released from the mother cell.

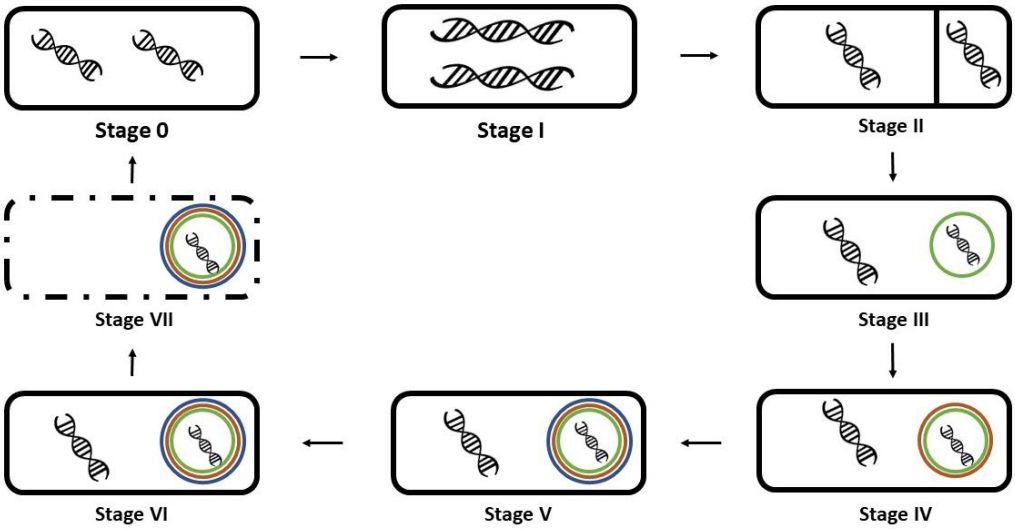


Figure 1.2: Stages of sporulation. The vegetative cell and the septum are represented in black. The chromosome is represented as a DNA molecule. The prespore is represented as a circle and the different layers are coloured: the membrane is represented in green, the cortex in orange and the coat in blue. Stage 0: Vegetative cell after DNA replication. Stage I: Axial filamentation. Stage II: Asymmetric division. Stage III: Engulfment. Stage IV: Cortex synthesis. Stage V: Coat synthesis. Stage VI: Maturation. Stage VII: Mother cell lysis. Adapted from Hilbert *et al.*[34].

The nomenclature of genes involved in sporulation is based on the stage at which sporulation is stopped when that gene is mutated. For instance, the deletion or mutation of a certain gene might block engulfment after asymmetric division (stage II) is complete. Then, that gene will be named *spoII* followed by letters to distinguish between different genes[34].

During axial filamentation (stage I), the chromosomes are extended across the cell and the origins of replication are located at the poles. This ensures that one of the origins will be inside the forespore compartment when the septum forms and separates the two cells[34, 35].

The formation of the septum (stage II) is essential for asymmetric division. Some of the proteins involved in this process also have a role in vegetative cell division (*e.g.*, DivIVA, FtsA, FtsZ). However, the sporulation septum is thinner, forms at the pole rather than at the middle of the cell, and associates with some proteins that are not part of vegetative division[34]. At the beginning of the division process FtsZ localises at both poles of the cell, but the septum only forms at one[36]. It has been suggested that spores form preferentially at the older pole of the cell, and that FtsA aids in the selection of the pole at which the spore will develop[34]. The formation of the second septum is prevented by the products of *spoIID*, *spoIIM* and *spoIIP*, which are regulated by the mother cell specific σ^{E} [34, 35].

When septum formation is complete, only one third of the chromosome (the origin-proximal region) is present in the forespore. After a short period of time, the DNA translocase SpoIIIE transports the rest of the chromosome to the forespore[35]. This delay in the transfer of the entire chromosome might be relevant to the compartmentalisation of gene expression and the forespore-specific activation of σ^F [37, 38].

Engulfment (stage III) is mediated by SpoIID, SpoIIM and SpoIIP, which are located in the septum and might be responsible for the degradation of its wall. This degradation starts in the middle of the septum rather than at the periphery[34, 37]. Following the degradation of the septum wall, the mother cell membrane starts migrating from the septum to the pole of the cell until both ends merge and the forespore is completely inside the mother cell[34, 35]. The final step, the merging of the mother cell membranes, appears to be mediated by SpoIIQ and SpoIIIE[34, 35].

The proteins SpoIIQ and SpoIIIAA-AH are involved in the formation of a channel ("feeding tube") between the mother cell and the forespore, which is essential for the coordination of regulation mechanisms between the two compartments. The channel has been hypothesised to allow the transport of small molecules that are involved in the activation of σ^G [39].

Stages IV and V of sporulation are characterised by the formation of the cortex and coat layers, respectively. The location and composition of these layers is described below. Finally, following spore maturation (stage VI), the mother cell lyses and the mature spore is released[34].

1.2.2 Morphology of spores

The development of the spore involves the formation of a multi-layered cell wall that will provide protection against the adverse environmental conditions[14]. A representation of the several layers that surround the spore is shown in Figure 1.3.

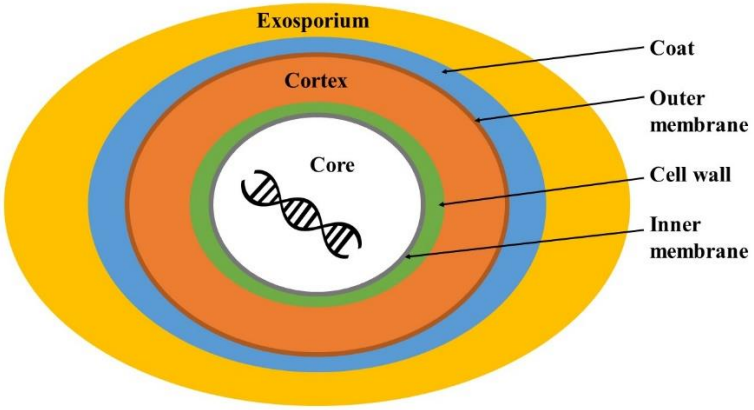


Figure 1.3: Spore structure. The spore core, containing the DNA, is protected by several layers: inner membrane, cell wall, cortex, outer membrane, coat and, in some, cases, exosporium. The representation and relative sizes are not drawn to scale. Adapted from Setlow[1].

The inner membrane is a low-permeability layer that surrounds the spore core, protecting it against chemicals that might damage the spore DNA[1]. The germ cell wall, which surrounds the inner membrane, is composed of peptidoglycan and is the precursor of the vegetative cell wall when the spore germinates[40, 41]. The spore cortex forms around the cell wall and is composed of peptidoglycan with two modifications: muramic acid δ-lactam and muramic acid linked only to alanine[14, 41]. This layer helps the spore core to maintain its dehydrated state, which is highly relevant for its resistance to heat treatments[42]. In later stages of sporulation, the mother cell synthesises proteins which will form the coat[1, 43]. The coat can be divided in sub-layers and protects the spore against chemicals such as hydrogen peroxide[2]. The exosporium is the most external layer of the spore, but it is only developed in some species. This layer is mainly composed of proteins and carbohydrates and has been hypothesised to be responsible for the interactions between the spore and the environment[1, 44].

When the spore is formed, the DNA is stored in the core and needs to be protected from adverse environmental conditions such as UV exposure and heat treatment. If the protective mechanisms fail and the DNA is damaged, then it needs to be properly repaired. However, repair mechanisms are usually more efficient during spore outgrowth[1]. During spore maturation, small acid-soluble proteins (SASPs) are synthesised in the forespore. These proteins have an important role in the spore core, as they bind to DNA and protect it. In fact, most of the proteins synthesised during spore maturation are SASPs (including α/β-type)[1]. SASPs have been identified in both *Bacillus* and *Clostridium* species[1, 45]. Furthermore, the mother cell synthesises pyridine-2,6-dicarboxylix acid (dipicolinic acid – DPA), which is transferred to the

forespore core and is often associated with calcium, giving Ca-DPA. Ca-DPA is also involved in DNA protection and forespore dehydration. Both α/β-type SASPs and Ca-DPA are essential to protect spore DNA[1].

While the spore is dormant, metabolic activity is minimal due to the dehydration and decreased pH of the spore core. In fact, the enzymes necessary for the catabolism of high-energy compounds are present in the spore core, but are not active[1].

1.2.3 Regulation of sporulation

The sporulation cascade is complex and highly regulated. The main regulatory proteins are Spo0A and four sigma factors (σ^F, σ^E, σ^G and σ^K), all of which are highly conserved across spore formers[6, 9, 37]. The roles of each of these regulators is briefly described below.

1.2.3.1 Spo0A

Regulation of sporulation in *B. subtilis* starts with the activation of Spo0A. The master regulator for sporulation will then regulate the expression of several genes involved in this process, including several other regulatory proteins, and will remain active until late stages of sporulation, particularly in the mother cell[13, 37].

In *B. subtilis*, Spo0A is activated in response to a phosphorelay system (Figure 1.4). In this mechanism, five sensory kinases (KinA-KinE) phosphorylate Spo0F (sporulation initiation phosphotransferase), triggering the sporulation cascade. These kinases can respond to a variety of signals, both intracellular and extracellular. The phosphate group is then transferred to Spo0B (single-domain response regulator) and finally to Spo0A (master regulator of sporulation)[37, 46]. The transcription of *spo0A*, *spo0F* and some genes encoding kinases appears to be regulated by σ^H, which also plays an important part in sporulation[11].

It should be noted that Spo0A~P can be dephosphorylated (and consequently inactivated) by some other phosphatases. This mechanism acts as an extra regulatory system to control the initiation of sporulation. Additional regulation of the phosphorelay system ensures that sporulation is not initiated if the DNA is damaged or its replication is blocked[34].

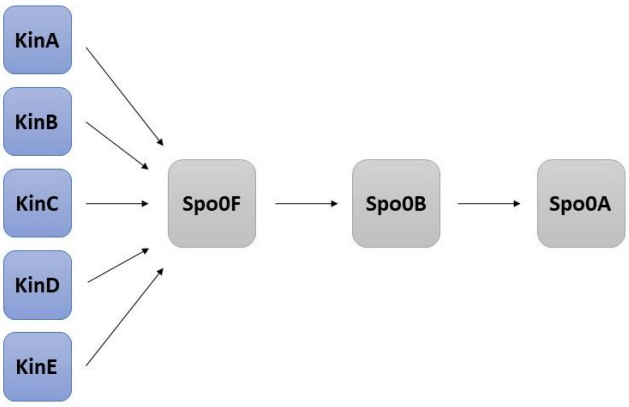


Figure 1.4: Representation of the phosphorelay system in *B. subtilis*. Extracellular signals are detected by kinases (A to E), which are then able to phosphorylate Spo0F. Spo0F~F phosphorylates Spo0B and Spo0B~F phosphorylates Spo0A. Adapted from Piggot *et al.*[37].

In clostridia, Spo0A is also responsible for the initiation of the sporulation cascade. Indeed, several studies showed that Spo0A regulates sporulation in various *Clostridium* species, such as *C. difficile*[47], *C. acetobutylicum*[48], *C. perfringens*[49] and *Clostridium beijerinckii*[50]. However, the phosphorelay system described above appears not to be responsible for the phosphorylation of Spo0A in clostridia[51, 52].

Some hypotheses have been proposed for the mechanism of Spo0A phosphorylation in clostridia. The first hypothesis is that clostridia might have specific kinases that can directly phosphorylate Spo0A. Based on the fact that the five phosphorelay kinases in *B. subtilis* lack an adjacent response regulator and that three of them contain transmembrane domains, Paredes et al.[15] tried to identify clostridial kinases with the same features. The authors identified five kinases in *C. acetobutylicum*, seven in *C. perfringens* and none in *C. tetani* that showed the desired characteristics. Furthermore, based on a microarray study in *C. acetobutylicum*[53], the authors confirmed that four of the five kinases identified in this organism are highly expressed just before the onset of sporulation. A second potential mechanism of Spo0A phosphorylation in clostridia is the presence of a novel phosphorelay system. In bacilli, some of the features of Spo0F, Spo0B and Spo0A are conserved and potentially relevant for the phosphorelay system. However, the phosphorelay components could not be identified in *Clostridium* species, indicating the absence of a phosphorelay system[15]. The third hypothesis is that a phosphodonor (*e.g.*, butyryl phosphate or acetyl phosphate) might be able to phosphorylate Spo0A[54]. This hypothesis is supported by a study that shows a positive correlation between increased intracellular levels of butyryl phosphate and solventogenesis/sporulation in *C. acetobutylicum*[55]. Finally, it

appears that in *C. difficile* Spo0A is phosphorylated by a sporulation-associated kinase, forming a two-component signal transduction system[56].

Once activated, Spo0A~P is able to bind to certain promoters and regulate transcription of several genes. Spo0A is responsible for the transcription of genes involved in stages I and II, including the early stage sigma factors and their regulators[13, 34]. Following asymmetric division, Spo0A continues to regulate the transcription of some genes. However, this regulation appears to be confined mainly to the mother cell compartment[57], possibly due to the inhibition of Spo0A activity by dephosphorylated SpoIIAA in the forespore[34]. In addition to sporulation, Spo0A is involved in the regulation of other processes, such as toxin production and solventogenesis[47, 49, 50, 56]. It has been hypothesised that the level of phosphorylation of Spo0A determines its regulatory activity[37].

1.2.3.2 Sigma factors

In addition to Spo0A, regulation of sporulation in *B. subtilis* involves sigma factors that control the expression of genes involved in sporulation and germination, as well as other regulators[10, 12]. Gene expression during sporulation has been shown to be dependent on time (early/late-sporulation) and location (forespore/mother cell)[58, 59]. It is worth noting that, despite the fact that the mother cell and the forespore undergo different regulation mechanisms, their regulation is coordinated[15, 60].

The regulation of the sporulation cascade by the sigma factors σ^F, σ^E, σ^G and σ^K will be described in Chapter 5. Briefly, σ^F and σ^E regulate early sporulation in the forespore and mother cell, respectively. These early-stage sigma factors are activated after asymmetric division and are involved in the transcription and/or activation of the other sigma factors. The late-stage σ^G and σ^K regulate gene expression in the forespore and mother cell, respectively, and are activated following engulfment[34].

1.2.3.3 Other regulatory proteins

In the mother cell, the sporulation cascade is regulated by σ^E, SpoIIID, GerR, σ^K and GerE (in hierarchical order). As a result, genes involved in sporulation are transcribed in pulses, *i.e.*, they are induced by a certain sigma factor and repressed by the next regulatory protein. This mechanism, which is based on a series of feed-forward loops, allows proper morphogenesis. It has also been observed that genes that are expressed in the same pulse often have related functions[12]. Moreover, some regulators other than

σ^F and σ^G also appear to control gene expression in the forespore. These include SpoIIQ, RsfA, SplA and SpoVT[34].

1.2.4 Germination

When the environment is favourable to the cell, spores can germinate and produce metabolically active vegetative cells. Several factors can induce germination, such as the presence of germinants[1]. In order for germination to occur, the spore cortex must be hydrolysed. Once that happens, the core is rehydrated, SASPs are degraded and metabolic activity is resumed. After these processes take place, the spore is no longer dormant and resistant to heat shock. The germination process also involves the release of ions (Zn^{2+}, K^+, Na^+ and Ca-DPA) from the spore, increase of internal pH and reabsorption of potassium ions[61, 62]. Spore germination can be divided into 3 main steps, as described below.

1.2.4.1 Initiation of germination

Germination can be triggered by the recognition of nutrients by receptors present in the inner spore membrane[63]. Those nutrients, referred to as germinants, have been hypothesised to be transported through the layers of the spore with the help of GerP[64]. Although homologues of the *gerP* gene are widely found in Bacilli, they could not be detected in Clostridia[65].

In *B. subtilis*, receptors are usually composed of three proteins, which are typically encoded by tricistronic operons: *gerA*, *gerB* and *gerK*. GerA is involved in L-alanine response[63, 66], while *gerB* and *gerK* are involved in response to AGFK (combination of L-asparagine, D-glucose, and K^+)[66, 67]. A Na^+/H^+-K^+ antiporter is also involved in germination and is encoded by *gerN* and *gerO*[68, 69]. Although there is homology between germinant receptors in *Bacillus* and *Clostridium* species, there are also some differences. The main difference is in the number of genes encoding these receptors, which is lower in Clostridia. Furthermore, the operons found in several *Clostridium* species have a variable number of genes, which are organised/oriented differently between species[65]. In fact, several incomplete operons (or operons with unusual structure) can be found in different clostridial species[62, 67]. Unlike in *Bacillus*, where the three subunits appear to be essential for functionality, these incomplete operons are fully functional in clostridia[65]. The receptors GerAA and GerAB have been

identified in *C. botulinum* and *C. sporogenes*, and their role in germination appears to be similar to *B. subtilis*[67].

The cell wall of Gram-positive bacteria contains peptidoglycan. The degradation of this peptidoglycan layer leads to the production of muropeptides, which can bind to germination receptors and initiate germination. A gene belonging to the serine/threonine protein kinase family, *prkC*, is thought to be involved in this process, as it is highly conserved among Bacilli and clostridia, and the protein encoded by this gene binds species-specific peptidoglycan fragments[70].

1.2.4.2 Degradation of the cortex

The rehydration of the core is only possible after the degradation of the spore cortex. In *Bacillus*, some germination-specific cortex-lytic enzymes have been identified: CwlJ is activated by Ca-DPA and SleB and YpeB are required for cortex degradation in *B. subtilis*[71]. The process appears to be similar in clostridia, as suggested by the presence of homologues of CwlJ, SleB and YpeB[72, 73]. Of note, is the apparent presence of enzymes with similar function but low similarity to SleB in *Clostridium* species[65].

Some other spore cortex-lytic enzymes and germination-specific proteases have been identified in *Clostridium* species. This suggests the presence of two different cortex-degradation mechanisms: one similar to *Bacillus* (involving CwlJ, SleB and YpeB), and one involving a different set of enzymes: SleC (spore cortex-lytic enzyme), SleM (cortical fragment-lytic enzyme), CspA, CspB and CspC (germination-specific proteases)[74-76]. These enzymes have been described in *C. perfringens*. SleC is inactive after production, and requires proteolytic activation by a germination-specific protease (GSP) in the cortex-core boundary in order to degrade the cortex[76]. In some *C. perfringens* strains (S40, ATCC13124 and 13), GSP is encoded by *cspA*, *cspB* and *cspC*, whereas in others (SM101) only one *csp* gene was detected[65, 76]. This alternative mechanism also appears to occur in other *Clostridium* species, as homologues of all or some of the enzymes have been identified[74]. However, homologues could not be found in *Bacillus* species[65].

1.2.4.3 Resumption of metabolic activity

In the later stages of spore outgrowth, DNA transcription needs to be resumed. However, this is only possible if the SASPs are eliminated. Hence, after re-hydration of the spore core, α/β-type SASPs are dissociated from the DNA and degraded by

germination proteases (GPR). Then, the cleavage products of SASPs (in particular SASP-γ) are further degraded to amino acids that will later be used to synthesise proteins[1, 65].

Previous studies reveal some variability in the amino acid sequence of SASPs in *Clostridium* species. On the other hand, genes encoding germination proteases are highly conserved in both *Bacillus* and *Clostridium*, and amino acid residues which are essential for functionality have been identified[65].

1.3 Hypothesis and aims

The main aim of this thesis is to elucidate the processes of sporulation and germination in *C. sporogenes*. This organism is relevant due to its potential to be used in cancer therapy, and its ability to form spores is a key part of the treatment. However, as described above, *C. sporogenes* needs to be genetically engineered in order to increase its anti-cancer properties. As the CDEPT therapy involves the administration of spores of the genetically engineered strain to the patient, the subsequent release of the organism to the environment is a concern. The obligate anaerobe *C. sporogenes* should not survive outside of the tumour unless it forms spores. Thus, a better understanding and control of sporulation should allow the containment of the strain by preventing the formation of spores outside of the laboratory environment. This can be achieved by generating a conditionally sporulating strain.

The concept of conditional sporulation has been applied in *C. difficile* by placing the gene encoding the master regulator of sporulation, *spo0A*, under the control of a tetracycline inducible promoter[77]. Moreover, an inducible system based on theophylline responsive riboswitches has already been developed in *C. sporogenes* for this purpose[78]. The use of theophylline is preferred in cancer therapy because tetracycline is an antibiotic that could be administered to the patient and trigger sporulation. However, the transcriptional regulator Spo0A controls many cellular functions in addition to sporulation, so other target genes are desirable to generate the conditionally sporulating strain.

In this thesis, the Transposon Directed Insertion-site Sequencing (TraDIS) approach was used to identify genes that are required for sporulation and germination in *C. sporogenes* (Chapter 4). These genes can potentially be used to generate the conditionally sporulating strain. The use of TraDIS required several optimisation steps, including the study of restriction systems to increase DNA transfer frequencies (Chapter 3). A better understanding of the regulation of sporulation involved the study of sigma factors and the sequencing of transcriptomes by RNA-Seq (Chapter 5). Finally, some genes hypothesised to be essential for sporulation/germination were characterised (Chapter 6). An overview of the main results is discussed in Chapter 7.

Chapter 2: Materials and methods

2

2.1 Microbiology techniques

2.1.1 Growth media, buffers and supplements

The composition of all media and buffers is listed in Table 2.1.

Supplements were prepared as described in Table 2.2.

Table 2.1: Compositions of media and buffers used in this study.

Medium	Composition	Amount (g/L)
LB (Luria-Bertani)	Tryptone	10
	Yeast Extract	5
	NaCl	10
TYG	Tryptone	30
	Yeast Extract	20
	Sodium thioglycolate	1
SOC	Tryptone	20
	Yeast Extract	5
	NaCl	0.584
	KCl	0.186
	$MgCl_2$	0.952
	$MgSO_4$	1.204
	Glucose	3.603
PBS (Phosphate-Buffered Saline)	NaCl	8
	KCl	0.2
	Na_2HPO_4	1.44
	KH_2PO_4	0.24
Electroporation buffer (pH 7.4)	Sucrose	92.4
	Na_2HPO_4	0.71
	NaH_2PO_4	0.6

Table 2.2: List of supplements used in this study.

Supplement	Solvent	Stock concentration	Working concentration: *E. coli*	Working concentration: *Clostridium*
D-Cycloserine	dH_2O	50 mg/mL	-	250 µg/mL
Chloramphenicol	100% ethanol	25 mg/mL	12.5 µg/mL (broth) 25 µg/mL (agar)	-
Thiamphenicol	50% ethanol	15 mg/mL	-	15 µg/mL
Kanamycin	dH_2O	50 mg/mL	50 µg/mL	-
Erythromycin	100% ethanol	50 mg/mL	300 µg/mL (broth) 500 µg/mL (agar)	30 µg/mL (broth) 50 µg/mL (agar)
Glucose	dH_2O	50% (w/v)	-	0.05% (w/v)
Anhydrotetracycline	100% ethanol	1 mg/mL	-	100 ng/mL
Lactose	dH_2O	100 mM	-	10 mM
Theophylline	DMSO*	266 mM	-	2 mM (broth) 5 mM (agar)

*DMSO: dimethyl sulphoxide

All media and buffers were prepared in distilled water (dH_2O) and autoclaved at 121°C for 20 minutes or sterilised by filtration through a 0.2 µm pore size filter (Ministart®). Agar plates were prepared by adding 1.5% (w/v) No.1 Bacteriological Agar (Oxoid) to the media prior to autoclaving. All media and supplements were supplied by Sigma-Aldrich or Fisher Scientific.

2.1.2 Bacterial strains and vectors

The bacterial strains and vectors used and generated in the present study are listed in the Appendix, Table S1 and Table S2 (respectively).

2.1.3 General culture and storage conditions

C. sporogenes NCIMB 10696 and derivatives were grown in an anaerobic cabinet (model MG1000 Mark II, Don Whitley Scientific Limited), under anaerobic conditions (80% N_2, 10% CO_2, 10% H_2), at 37°C. All media and reagents were pre-reduced before use by incubation in the anaerobic cabinet for at least 4 hours. *C. sporogenes* strains were grown in TYG media and stored at -80°C in TYG broth supplemented with 10% (v/v) glycerol.

E. coli strains were grown in Luria Bertani (LB) media, at 37°C under aerobic conditions. Liquid cultures were incubated with agitation at 200 rpm. *E. coli* strains were stored at -80°C using the MicrobankTM Long Term Bacterial Storage vials (Pro-lab Diagnostics).

Growth media was supplemented whenever required (specific information given where necessary).

2.1.4 Quantification of bacterial growth

The growth of the bacterial cultures was assessed by measuring the optical density of the culture at 600 nm (OD_{600}). The instrument used to measure OD_{600} was the Genesys™ 10S UV-Vis spectrophotometer (Thermo Scientific). Alternatively, when several cultures were being measured simultaneously, the CLARIOstar® microplate reader (BMG LABTECH) was used with flat bottom 96-well plates (Costar, Corning®). All cultures were diluted as necessary and measured against a blank of sterile medium.

2.1.5 DNA transfer into *E. coli* strains

2.1.5.1 Preparation of electro-competent cells of *E. coli*

In order to prepare competent cells, overnight cultures were generated by using a frozen stock of the selected *E. coli* strain (TOP10, CA434 or sExpress) to inoculate 5 mL of LB broth supplemented with appropriate antibiotics. After overnight growth, 1 mL of the starter culture was used to inoculate 200 mL of sterile, pre-warmed LB broth and the cells were harvested when an OD_{600} of 0.5-0.7 was achieved. The culture was then transferred to pre-chilled Falcon tubes and centrifuged at 3500 ×*g* for 10 minutes at 4°C. The pellet was gently washed three times with sterile ice cold dH_2O (twice with 20 mL and once with 10 mL). After each wash step, the cells were

centrifuged at 3500 ×*g* for 10 minutes (4°C) and the pellets were combined. The final pellet was re-suspended in 1 mL of 1 mM MOPS with 10% (v/v) glycerol. The cells were aliquoted into 50 µL samples and stored at -80°C.

2.1.5.2 Transformation of electro-competent *E. coli* cells with DNA by electroporation

For each electroporation, an aliquot of electro-competent cells was mixed with 200-500 ng of purified plasmid DNA or 10 µL of dialysed (0.025 µM Merck Millipore dialysis discs over dH_2O for at least 20 minutes) ligation product. The cell-plasmid mixture was transferred to a pre-chilled electroporation cuvette (0.2 cm, Bio-Rad), which was put in a Bio-Rad MicroPulser™. A pulse of 22.5 kV/cm, 200 Ω, 25 µF was applied to the cuvette, and 400 µL of sterile SOC medium were used to re-suspend the cells immediately after. The re-suspended cells were transferred to Eppendorf tubes and incubated for 1 hour at 37°C (or 2 hours at 30°C) with agitation. Finally, transformed cells were plated in suitable dilutions onto selective LB agar plates.

2.1.6 DNA transfer into *C. sporogenes* strains

2.1.6.1 Transfer of plasmids from an *E. coli* donor into *C. sporogenes* by conjugation

The protocol used to introduce shuttle plasmids into *C. sporogenes* NCIMB 10696 was based on the one described by Purdy *et al.*[79]. Selected donor strains (*E. coli* CA434 or *E. coli* sExpress carrying shuttle plasmids) and recipient strains (*C. sporogenes* WT or mutants) were grown overnight in broth with appropriate selection. On the following day, cells from 1 mL of *E. coli* culture were centrifuged for 1 minute at 5000 ×*g*, carefully washed with 1 mL of phosphate buffer saline (PBS) and centrifuged again. The pellets were transferred to the anaerobic cabinet, where they were re-suspended in 200 µL of *C. sporogenes* overnight culture. The donor-recipient mixture was spotted onto a TYG plate supplemented with glucose and incubated in the anaerobic cabinet at 37°C. After at least 7 hours, the cells were harvested with a sterile loop and re-suspended in 500 µL of sterile, pre-reduced PBS. The cells were plated onto media supplemented with D-cycloserine to counter select *E. coli* donor cells, and appropriate antibiotics to select the conjugated plasmid.

2.1.6.2 Preparation of electro-competent cells of *C. sporogenes*

In order to prepare competent cells, 5 mL TYG were inoculated with the appropriate *C. sporogenes* strain and grown overnight. On the following day, 230 mL of sterile, pre-reduced TYG broth was inoculated with 1 mL of overnight culture and grown at 37°C to an OD_{600} of 0.2-0.25. When the correct OD was achieved, 46 mL of 2 M sucrose were added to the culture, which was then grown to an OD_{600} of 0.4-0.5. The culture was transferred to pre-chilled Falcon tubes and centrifuged at 6000 ×*g* for 10 minutes (4°C). Then, the pellets were washed two times with sterile ice cold electroporation buffer (first 40 mL and then 15 mL). After each wash step, the cells were centrifuged at 6000 ×*g* for 10 minutes (4°C). Finally, the pellets were combined into one tube and re-suspended in 1.6 mL of electroporation buffer. The cells were aliquoted into 590 µL samples and transformed immediately. The whole procedure was done in the anaerobic cabinet, except the centrifugation steps.

2.1.6.3 Transformation of electro-competent cells of *C. sporogenes* with DNA by electroporation

Electro-competent cells were mixed with 2-5 µg of purified plasmid. The cell-plasmid mixture was transferred to a chilled electroporation cuvette (0.2 cm, Bio-Rad), which was put in a Bio-Rad Gene Pulser. A pulse (2.0 kV, ∞ Ω, 25 µF) was applied to the cuvette, and 1 mL of sterile TYG broth was used to resuspend the cells immediately after. The resuspended cells were transferred to Eppendorf tubes and incubated for 3–4 hours at 37°C. Finally, transformed cells were plated onto appropriate selective media. The whole procedure was done in the anaerobic cabinet.

2.1.7 Plasmid loss

Plasmid loss was achieved by growing the cells carrying the plasmid in non-selective liquid media for 48 hours. Then, appropriate dilutions of the cultures were plated onto agar plates without selection. After 24 hours of growth, isolated colonies were randomly selected and patch-plated onto agar plates with and without antibiotic selection. The following day, colonies that could not grow in the presence of the antibiotic were screened using colony PCR (section 2.2.5.3) to confirm plasmid loss.

2.2 Molecular biology techniques

2.2.1 Extraction and purification of plasmid DNA

Plasmid DNA extraction and purification was performed using the Monarch® Plasmid MiniPrep Kit (NEB - New England BioLabs), according to manufacturer's instructions. The plasmid DNA was eluted in dH_2O and stored at -20°C.

2.2.2 Extraction and purification of genomic DNA

Genomic DNA was extracted and purified using the GenElute™ Bacterial Genomic DNA Kit (Sigma-Aldrich), according to manufacturer's instructions. The genomic DNA was eluted in dH_2O and stored at -20°C.

2.2.3 Extraction and purification of total RNA

Prior to any work involving RNA, all surfaces and pipettes were cleaned with RNase Zap (Sigma-Aldrich). In addition, only filter tips and RNase-free tubes were used to handle RNA samples.

In order to extract RNA, selected strains of *C. sporogenes* were grown overnight in TYG broth. The following day, fresh TYG broth was inoculated with the overnight cultures to an OD_{600} of 0.01. At pre-determined time-points, 1.5 mL of culture were mixed with 3 mL of RNA protect and incubated in the anaerobic cabinet for 5 minutes. Then, the cells were taken outside of the cabinet and centrifuged at 15000 ×*g* for 1 minute (4°C). The supernatants were discarded and the pellets were stored at -80°C for less than 24 hours. The RNA was then extracted using the FastRNA™ Pro Blue Kit (MP Biomedicals™). The pellets were thawed on ice, resuspended in 1 mL RNA pro solution and transferred to a RNA pro matrix tube. The cells were homogenised at 6400 rpm for 45 seconds in a Bertin Corp Precellys® 24 bead beating tissue homogeniser (Fisher Scientific). The homogenised sample was centrifuged (15000 ×*g*, 15 minutes, 4°C), and the supernatant was carefully transferred to a fresh RNase-free tube and incubated at room temperature for 5 minutes. In the fume hood, 300 µL of ice-cold chloroform were added to each sample, which was then mixed for 6 seconds in a vortex mixer. After a 5-minute incubation at room temperature, the samples were centrifuged (15000 ×*g*, 15 minutes, 4°C) and the upper phase was transferred to fresh tubes containing 500 µL of absolute ethanol. The samples were mixed by inversion and incubated at -80°C for 1-2 hours to precipitate the RNA. The samples were

centrifuged again at (15000 ×*g*, 15 minutes, 4°C), the supernatant was discarded and the RNA pellet was washed with 500 µL of 70% ethanol. After another centrifugation (15000 ×g, 15 minutes, 4°C), the supernatant was discarded and the RNA pellet was air dried for 30 minutes. The pellet was resuspended in 50 µL DEPC-treated water and stored at -80°C. The following day, the RNA samples were treated with DNase (Ambion™ TURBO DNA-free kit, Fisher Scientific) following the manufacturer's instructions. The RNA was then purified using the RNeasy Mini Kit (Qiagen). First, 3 µL of β-mercaptoethanol were added to 400 µL of RLT buffer. Then, 350 µL of RLT+β-mercaptoethanol and 200µL of ethanol were added to each DNase-treated sample and mixed by inversion. The mixture was transferred to an RNeasy column and centrifuged at 10000 ×*g* for 15 seconds. The column was washed twice with 500 µL RPE buffer and centrifuged at 10000 ×*g* for 2 minutes to remove ethanol residues. The RNA was eluted in 50 µL of DEPC-treated water and stored at -80°C. The quality and concentration of the RNA were assessed using the NanoDrop Lite Spectrophotometer (Thermo Scientific) and the Agilent 2100 Bioanalyzer (Agilent Technologies).

2.2.4 Reverse transcription of RNA into cDNA

The reverse transcription reaction was performed using the Omniscript Kit (Qiagen) according to manufacturer's instructions. Briefly, 1.5 µg of RNA was used as template for cDNA synthesis by the Omniscript reverse transcriptase using random primers (Random primer 6, NEB). The reaction mix (Table 2.3) was incubated at 37°C for 1 hour, and then stored at -20°C.

Table 2.3: Components used in the reaction for reverse transcription of RNA into cDNA, and respective volumes (Omniscript Kit, Qiagen).

Component	Volume (µL)
10X buffer RT	2
dNTP mix (5mM each dNTP)	2
Random primer 6 (NEB; 10 µM)	2
RNase inhibitor (10 U/µL)	1
Omniscript reverse transcriptase	1
RNA (1.5 µg)	x
RNase free water	Up to 20 µL

2.2.5 Amplification of DNA

2.2.5.1 Polymerase Chain Reaction (PCR)

All genes of interest were amplified using KOD Hot Start Master Mix (Novagen). The PCR program conditions were selected according to manufacturer's instructions. Annealing temperatures (Ta) were calculated based on the melting temperature (Tm) of the primers used, and extension times were dependent on the expected size of the amplicon. All PCR reactions were carried out in a thermal cycler block (Applied Biosystems Gene Amp PCR system 2400).

2.2.5.2 Splicing by overlap extension (SOE) PCR

The gene editing cassettes used in the CRISPR/Cas9 systems were obtained using SOEing PCR. This approach relies on the design of specific oligonucleotides with overlapping sections, as previously described[80] and shown in Figure 2.3.

2.2.5.3 Colony PCR

Single colonies were picked from agar plates and re-suspended in 10 µL of sterile dH_2O. Then, 1 µL of the colony suspension was added to the PCR mix (DreamTaq™ Green PCR Master Mix, primers and dH_2O) prepared according to manufacturer's instructions. The PCR program used allowed 10 minutes for the initial denaturation step at 98˚C to facilitate complete cell lysis.

2.2.5.4 Inverse PCR

Inverse PCR was one of the methods used to assess if the ClosTron was inserted only in the target gene, and the procedure has been described by Cartman *et al.*[81]. Genomic DNA was extracted from overnight cultures as described in section 2.2.2. Then, the DNA was digested with *Hin*dIII at 37°C for 2 hours and incubated at 80°C for 20 minutes to inactivate the restriction enzyme. The digestion mix was self-ligated with T4 DNA ligase (Promega) as described in section 2.2.8. The DNA was amplified using KOD Hot Start Master Mix (Novagen), and the reaction components and program are presented in Table 2.4. The PCR products were purified from the agarose gel (section 2.2.6.1) and sequenced with primer erm_RAM_seq to confirm the site of insertion.

Table 2.4: Inverse PCR components and conditions.

PCR reaction components	
Component	**Volume (µL)**
Ligation product	6
Primer (erm_inv_F)	1
Primer (erm_inv_R)	1
KOD Hot Start	12.5
dH_2O	4.5
Total	25
PCR program	
Conditions	**Cycles**
95°C, 2 minutes	1
95°C, 20 seconds 57°C, 10 seconds 70°C, 3 minutes and 45 seconds	40
70°C, 10 minutes	1

2.2.5.5 RT-PCR

The RT-PCR reaction was carried out as described in section 2.2.5.1, but with cDNA as template instead of DNA.

2.2.5.6 RT-qPCR

To assess gene expression levels, RT-qPCR was performed in a Light Cycler 480 II (Roche), using Power SYBR Green Master Mix (Thermo Fisher), and according to the manufacturer's instructions. The thermal cycle profile was 10 minutes at 95°C, followed by 40 cycles of 95°C for 15 seconds and 60°C for 1 minute. Each reaction was followed by the analysis of a melting curve with a temperature range from 60°C to 95°C, with 5 acquisitions per °C to allow assessment of reaction specificity. Reaction efficiency was determined by analysing standard curves from serially diluted cDNA. Two negatives controls were included in each analysis, one with RNA as template (non-reverse transcription; NRT) and another with water as template (non-template control; NTC). The gene expression level was determined by relative quantification and calculated with the E-method[82] using the LightCycler® 480 software, version 1.5.

2.2.5.7 List of primers

The list of primers used in this study is presented in the Appendix, Table S3.

Primers to be used in RT-qPCR were carefully evaluated: i) specificity of primer binding was assessed using BLAST; ii) secondary structure of the cDNA, namely at the primer binding site, was checked using mfold; iii) melting temperatures, primer dimers and self-annealing were also taken into account; iv) primers were designed to amplify fragments of 100-300 base pairs.

2.2.6 Agarose gel electrophoresis

The size of DNA fragments (either PCR products or restriction fragments) was evaluated by agarose gel electrophoresis. The samples were loaded into previously prepared 1% (w/v) agarose gels stained with SYBR® Safe DNA Gel Stain (Invitrogen). The gels were run in TAE buffer at 90-100 V for 30-90 minutes. The fragments were then visualised under ultraviolet light using a Gel DocTM XR+ system (Bio Rad).

2.2.6.1 Extraction of DNA from agarose gels

Whenever necessary, DNA fragments were extracted from excised agarose gel pieces using the Monarch® DNA Gel Extraction Kit (NEB), following manufacturer's instructions.

2.2.7 Restriction endonuclease digestion

DNA digestions were performed using restriction endonucleases supplied by NEB according to manufacturer's instructions.

2.2.8 Ligation and membrane dialysis of DNA

In order to ligate DNA fragments, T4 DNA polymerase (Promega) and other essential ligation components were used according to manufacturer's instructions. Prior to electroporation, ligation products were dialysed using 0.025 μM Merck Millipore dialysis discs over dH_2O for at least 20 minutes.

2.2.9 DNA sequencing

The sequences of the plasmid DNA and purified PCR products were verified by nucleotide sequencing performed by Eurofins Genomics.

Whole genome sequencing was provided by DeepSeq facility (University of Nottingham) using the MiSeq Illumina platform. The sequencing data was mapped to the *C. sporogenes* NCIMB 10696 reference genome using the CLC Genomics Workbench version 12 (Qiagen), and single nucleotide variants (SNVs) were identified using the Basic Variant Detection tool (default parameters).

2.2.10 Southern blot

Southern blot was used to confirm single insertions in ClosTron mutants. A probe specific for the intron was obtained by amplifying a ClosTron template with primers EBS2 and SalI-R1. DNA denaturation was achieved by heating the amplicon at 100°C for 10 minutes followed by incubation on ice. Then, 4 µL of the DIG High Prime enzyme (Roche) was added to the amplicon to label the DNA with Digoxigenin-11-dUTP, and the mixture was incubated overnight at 37°C. The reaction was stopped by incubating the mixture at 65° for 10 minutes. The ladder (λ DNA/*Hin*dIII marker, Promega) was also labelled as described above.

Genomic DNA was extracted from the ClosTron mutants as described in section 2.2.2 and then digested with *Hin*dIII endonuclease. The digested DNA and the labelled ladder were run on an agarose gel as described in section 2.2.6. Then, an apparatus assembled as shown in Figure 2.1. The apparatus was kept for at least 5 hours to allow the capillary transfer of the DNA from the agarose gel onto the nitrocellulose membrane.

After the capillary transfer, the DNA was fixed onto the membrane by incubating for 2 minutes under UV light. The membrane was then washed in 100 mL 2× SSC (3 M NaCl, 300 mM Tri-Na-citrate), and incubated with 10 mL of pre-hybridisation buffer (DIG Easy Hyb, Roche) for 1 hour at 42°C. A probe mix was prepared, containing 24 µL of the labelled ClosTron probe, 19 µL of the labelled ladder probe and 10 mL of DIG Easy Hyb. The probe mix was boiled for 10 minutes to allow denaturation of the DNA and placed on ice. The membrane was incubated with the probe mix overnight at 42°C.

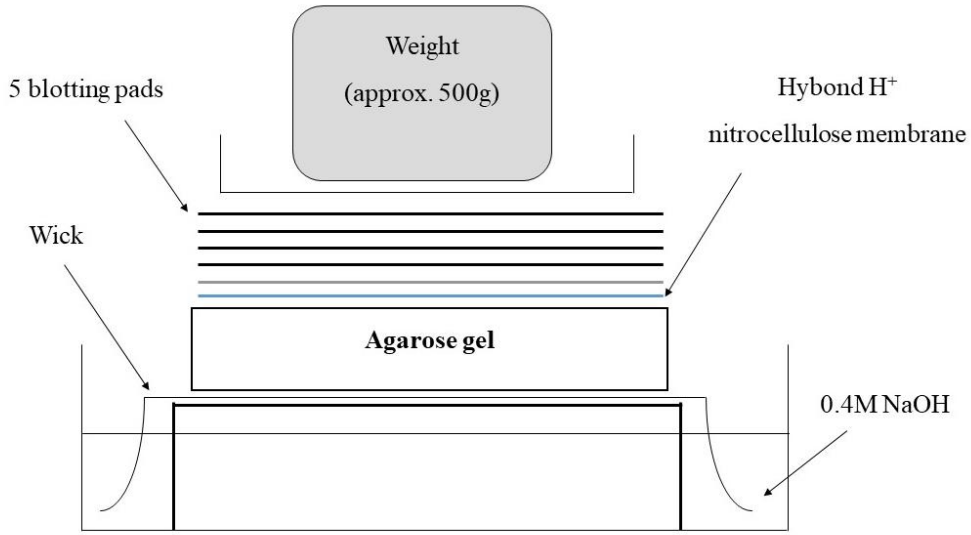


Figure 2.1: Apparatus used to transfer the *Hin*dIII-digested DNA from the agarose gel onto a Hybond H^+ nitrocellulose membrane. The gel was put on a support inside a container filled with 0.4 M NaOH. A blotting paper was pre-wet on 0.4 M NaOH and acted as a wick for the transfer solution. The agarose gel was placed on top of the wick, and the Hybond H^+ nitrocellulose membrane (pre-wet) was placed on top of the gel. Then, 5 blotting pads were placed on top of the membrane (the first one was pre-wet) and a weight was positioned on top of the layers.

On the following day, the membrane was thoroughly washed:

i) Membrane was washed twice with 100 mL of washing solution 1 (10 mL of 20× SSC, 1 mL of 10% SDS, 89 mL of dH_2O), incubating at room temperature for 5 minutes.

ii) Membrane was washed twice with 100 mL of washing solution 2 (2.5 mL of 20× SSC, 1 mL of 10% SDS, 96.5 mL of dH_2O), incubating at 68°C for 15 minutes.

After the washes, the membrane was equilibrated by incubating at room temperature for 1 minute with 100 mL of maleic acid buffer T (20 mL of 5× maleic acid buffer, 0.3 mL of Tween 20, 79.7 mL of dH_2O). Then, the membrane was blocked by incubating at room temperature for 30 minutes with 30 mL of blocking buffer (3 mL of 10× blocking reagent (Roche), 6 mL of 5× maleic acid buffer, 21 mL of dH_2O). The Anti-Digoxigenin-AP antibody (Roche) was diluted in blocking buffer (3 µL in 30 mL), and the mixture was added to the membrane and incubated at room temperature for 30 minutes. Finally, the membrane was washed three times with 100 mL of maleic acid buffer T (at room temperature for 15 minutes) and incubated for 2 minutes at room temperature with detection buffer (2.5 mL of 1 M Tris-HCl, 2.5 mL of 1 M NaCl, 20 mL of dH_2O). A chemiluminescent substrate (1.5 mL of CSPD, Roche) was added to the membrane, which was incubated at 37°C for 10 minutes. Then, the membrane

was developed for 15 minutes using developer and fixer supplied by Calumet Photography.

2.3 Phenotypic characterisation

2.3.1 Sporulation assay

Frozen stocks of *C. sporogenes* strains were used to inoculate 1 mL of TYG broth and the cultures were grown overnight. On the following day, the cells were sub-cultured (1:100) and grown for 3 hours. Then the cells were sub-cultured again into 10 mL TYG broth (1:100) and the first time point (t = 0 hours) was collected immediately. The cultures were kept for 5 days under anaerobic conditions to allow sporulation to occur. At several time points (0 hours, 24 hours, 48 hours, 72 hours, 96 hours and 120 hours after inoculation), 180 µL of each culture were collected and heat-treated at 80°C for 20 minutes. Then, heat treated and non-heat treated samples were serially diluted in PBS and plated onto TYG agar (each biological replicate was plated in triplicates) using the drop plate technique[83]. Plates were incubated for 24 hours in order to allow the spores to germinate. Formed colonies were enumerated and expressed as colony forming units per mL of culture (CFU/mL).

2.3.2 Growth curves

In order to assess the growth profile of the different strains, overnight cultures of selected *C. sporogenes* strains were sub-cultured into 20 mL of fresh TYG broth to an OD_{600} of 0.01. Samples were collected at the moment of inoculation, and then every 2 hours during the first 12 hours of growth. Another sample was collected 24 hours after inoculation. The OD_{600} was measured using the CLARIOstar® microplate reader as described in section 2.1.4.

2.3.3 Conjugation efficiency assays

In order to assess conjugation efficiency, donor strains were grown (in triplicates) overnight, in LB broth supplemented with kanamycin (to maintain the R702 factor) and appropriate selection to maintain the vector to be conjugated. In the case of the strains carrying the pMTL80221-P_{BAD}-*hsdMS* vector, erythromycin and L-arabinose were also added to the cultures. The recipient strains were also grown overnight with appropriate antibiotics. Conjugation was carried out as described in section 2.1.6.1.

For each donor strain replicate, an additional 1 mL of culture was pelleted and washed as described in section 2.1.6.1. The pellet was resuspended in 1 mL of PBS, serially diluted and plated onto LB agar plates with appropriate selection using the drop plate technique[83]. This allowed the enumeration of the donor strains' CFUs.

After approximately 7 hours of conjugation, the cells were harvested, resuspended in 500 µL of PBS and serially diluted. The diluted cultures were plated using the drop plate technique[83] onto i) TYG plates with no selection to enumerate total CFUs; ii) TYG plates with appropriate selection to enumerate transconjugants. After 24 to 48 hours of growth, formed colonies were enumerated and expressed as colony forming units per mL of culture (CFU/mL).

2.3.4 Transformation efficiency assays

Electro-competent cells of selected *C. sporogenes* strains were prepared as described in section 2.1.6.2. Immediately after, 5 µg of the vector pMTL83151 were electroporated into the competent cells as described in section 2.1.6.3. Following recovery (incubation at 37°C for 4 hours), the cells were centrifuged (4000 ×*g*, 10 minutes) and the pellets were resuspended in 500 µL PBS. Then, the cultures were serially diluted and plated using the drop plate technique[83] onto i) TYG plates with no selection to enumerate total CFUs; ii) TYG plates with appropriate selection to enumerate transformants. After 24 to 48 hours of growth, formed colonies were enumerated and expressed as colony forming units per mL of culture (CFU/mL). All transformations were done in triplicates.

2.3.5 Microscopy

Bacterial cultures were observed using a Nikon Eclipse Ci-E microscope (Nikon Instruments Europe B.V., UK) using standard microscopy procedures.

Spores were stained using the Schaeffer and Fulton Spore Stain Kit (Sigma-Aldrich) following manufacturer's instructions.

2.4 Methods used to assess gene expression levels

2.4.1 Enzymatic assay of chloramphenicol acetyltransferase (CAT assay)

With the purpose of evaluating the strength and tightness of the Tet inducible system, 20 mL of TYG broth supplemented with appropriate antibiotics were inoculated (1:100) with overnight cultures of *C. sporogenes* strains carrying the constructs to be tested. Four hours after inoculation, the cultures were induced with 100 ng/mL or 200 ng/mL of anhydrotetracycline (aTc). Strains carrying the pMTL82254_Lac constructs were induced with 10 mM of lactose instead of aTc. Uninduced cultures were used as controls. After 8 hours of inoculation, 2 mL were taken from each culture and centrifuged at 6,000 ×*g* for 10 minutes (4°C). The supernatants were discarded and the pellets were stored at -20°C.

In order to obtain the cell lysates, the frozen pellets were thawed on ice for 15 minutes. A working solution was prepared by dissolving a tablet of cOmplete™ Protease Inhibitor Cocktail (EDTA free; Roche) in BugBuster (Novagen). Then, 300 µL of working solution was added to each pellet and vortexed gently. Cell suspensions were incubated at room temperature for 20 minutes, with gentle agitation. Finally, the suspensions were centrifuged at 6,000 ×*g* for 10 minutes (4°C) to remove cell debris and the soluble fraction was transferred to a new tube.

In each well of a 96-well plate, 10 µL of cell lysates were mixed with 40 µL Tris Buffer (100 mM; pH 7.8). Two of the wells contained only Tris Buffer and no sample – this was the blank. Each 96-well plate was prepared in triplicates (technical replicates). Note that for each sample there are also three biological replicates for each condition being tested.

A CAT assay master-mix was prepared based on the protocol from Sigma-Aldrich:

- Chloramphenicol: 500 µL of 0.3% (w/v) stock prepared in dH_2O (final concentration: 0.005%);
- Acetyl CoA sodium salt: 1 mL of 5 mM stock prepared in dH_2O (final concentration: 0.167 mM; filter sterilized);
- DNTB (5,5'-Dithio-bis (2-Nitrobenzoic Acid)): 1 mL of 2.5 mM stock prepared in 100 mM Tris Buffer - pH 7.8 (final concentration: 0.0833mM; filter sterilized);

- 20 mL of 100 mM Tris Buffer - pH 7.8 (final concentration: 94 mM; autoclaved).

The 96-well plates were analysed in a CLARIOstar® plate reader set at 25°C. The volume of master-mix to be added to each sample was set to 150 µL.

CAT activity was determined according to the following equations:

$$Units\ CAT/mg\ protein\ = \frac{Units/mL\ CAT}{mg\ \ protein/mL\ enzyme}$$

$$Units/mL\ CAT\ = \frac{(\Delta A_{412}/\ \mathrm{min.\,test} - \Delta A_{412}/\ \mathrm{min.\,blank}) \times 0.2\ \times \mathrm{df}}{0.0136\ \times 0.01}$$

0.2 = total volume of the assay (mL)
0.01 = volume of the cell lysate (mL)
0.0136 = micromolar extinction coefficient for DTNB at 412 nm.
Df = dilution factor

The protein concentration (mg protein/mL enzyme) was determined using the BCA Protein Assay (Thermo Fisher Scientific) according to manufacturer's instructions. The absorbance (ΔA_{412}/min test - ΔA_{412}/min blank) was determined using the CLARIOstar's software, based on the maximum of slope (blank corrected) for the time range tested.

2.4.1.1 Construction of vectors tested with the CAT assay

Several vectors were tested using the CAT assay, and most of them were available in the SBRC culture collection. Only the pMTL-tet3nO_RBS vector was generated as part of this study. A DNA fragment containing the P_{fdx} promoter without the second RBS was obtained by amplifying pMTL-tet3nO with primers tet3nO_SalI_F and tet3nO_1RBS_R. The obtained fragment contained the restriction sites *SalI* and *NdeI*, which were used to clone the fragment into the linearized pMTL-tet3nO vector originating pMTL-tet3nO_RBS.

2.4.2 Red fluorescence protein (RFP) reporter assay

The RFP reporter assay was used to assess gene expression levels in *E. coli* cultures. Accordingly, *E. coli* TOP10 cells carrying the selected constructs were grown overnight in LB broth supplemented with erythromycin. In the case of strains carrying

constructs with the P_{BAD} promoter, each technical replicate was done in duplicates: one was induced with 100 mM L-arabinose and the other was left uninduced. After 16 hours of growth, 200 µL of each culture were transferred to a flat bottomed Greiner 96-well plate. The fluorescence levels of the samples were measured in the CLARIOstar® microplate reader with 20 flashes per well of excitation at 585 nm wavelength and an emission detection at 620 nm. The OD_{600} was also measured in each well using the CLARIOstar® microplate reader as described in section 2.1.4. All samples were measured against a blank of sterile LB broth and, for each well, the fluorescence intensity was normalised to the OD_{600}.

2.4.2.1 Construction of vectors tested with the RFP assay

To generate the promoterless vector pMTL80221-P_{neg}-RFP, the gene encoding RFP was amplified with primers RFP_NotI_F and RFP_NheI_R and inserted into pMTL80221 through the *Not*I and *Nhe*I restriction sites.

To generate pMTL80221-P_{hsd}-RFP, the P_{hsd} promoter was amplified from genomic DNA extracted from *C. sporogenes* WT using primers Phsd_NotI_F and Phsd_BsaI_R, which introduced the restriction sites *Not*I and *Bsa*I. The gene encoding RFP was amplified with RFP_NotI_F and RFP_NheI_R, introducing the restriction sites *Bsa*I and *Nhe*I. The promoter and the gene were ligated using the *Bsa*I restriction site and introduced into pMTL80221 through the *Not*I and *Nhe*I restriction sites.

To generate pMTL80221-P_{BAD}-RFP, a fragment containing the P_{BAD} promoter and the *araC* gene was amplified using primers araC_NotI_F and PBAD_BsaI_R, introducing the restriction sites *Not*I and *Bsa*I. The gene encoding RFP was amplified with RFP_NotI_F and RFP_NheI_R, introducing the restriction sites *Bsa*I and *Nhe*I. The promoter and the gene were ligated using the *Bsa*I restriction site and introduced into pMTL80221 through the *Not*I and *Nhe*I restriction sites.

A generic representation of the vectors used in the RFP assay can be found in the Appendix, Figure S1.

2.5 Genome editing

For each genome editing tool, the same procedure was followed for all target genes. Thus, primer names and vectors are described using the generic term “gene”.

2.5.1 ClosTron

The ClosTron technique is a powerful tool for directed mutagenesis and has been previously described[84, 85]. Briefly, the group II introns were re-targeted using Perutka algorithm available in www.clostron.com. Plasmids (pMTL007C-E2) containing the ClosTrons targeting the genes of interest were ordered from DNA2.0. These plasmids were delivered in filters and recovered in 50 µL DEPC-treated H_2O. Then, 2 µL of plasmid solution were electroporated into *E. coli* CA434 and conjugated into *C. sporogenes* NCIMB 10696. *C. sporogenes* cells were plated first onto TYG plates supplemented with thiamphenicol to select cells that acquired the plasmid (pMTL007C-E2 contains the *catP* gene, which confers resistance to chloramphenicol and thiamphenicol). Then, the cells were plated onto TYG supplemented with erythromycin to select cells where the group II intron was inserted into the target gene (contains the RAM, which is based on the *ermB* gene). Finally, the presence of the group II intron in the target gene was confirmed by colony PCR, using different sets of primers: i) RAM_F and RAM_R to confirm the presence of the RAM in the genome; ii) EBS Universal and gene_CTscr_F/R to confirm the presence of the RAM in the target gene; iii) gene_CTscr_F and gene_CTscr_R, also to confirm the presence of the RAM in the target gene (Figure 2.2). Plasmid loss was induced as described in section 2.1.7. A new colony PCR with the same primers confirmed that the cells were still mutants after plasmid loss. The insertion of the intron in the target gene was also confirmed by Sanger sequencing using primers gene_fl_F and/or gene_fl_R. Single insertions of the intron (exclusively at the target gene) were confirmed by either Southern blot (section 2.2.10) or inverse PCR (section 2.2.5.4).

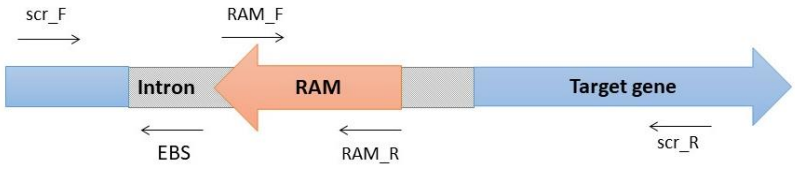


Figure 2.2: Schematic representation of the primer binding sites for colony PCR to screen for the insertion of the ClosTron in the target gene.

2.5.2 CRISPR/Cas9

The CRISPR/Cas9 system used in Chapter 3 has been previously described[86]. Briefly, vectors pMTL43151-CLSPO-gene were electroporated into *E. coli* CA434 and then conjugated into an appropriate *C. sporogenes* strain. Following conjugation, the cells

were plated onto TYG agar plates supplemented with thiamphenicol to select cells that incorporated the plasmid. After 24-48 hours of growth, random colonies were selected and screened using primers gene_scr_F and gene_scr_R. In addition, the in-frame deletion of the gene was confirmed by Sanger sequencing using primers gene_fl_F and/or gene_fl_R.

The in-frame deletion of genes using RiboCas (Chapters 5 and 6) was performed as described in Cañadas *et al.*[87]. Briefly, vectors pRECas1-gene_BM were electroporated into *E. coli* CA434 and conjugated into *C. sporogenes* WT. The conjugated cells were plated into TYG agar supplemented with thiamphenicol, as described above. After 24-48 hours of growth, cells were patch-plated onto agar plates with and without theophylline. Cells that were able to grow on theophylline were screened using primers gene_scr_F and gene_scr_R. In addition, the in-frame deletion of the gene was confirmed by Sanger sequencing using primers gene_fl_F and/or gene_fl_R.

In order to revert each gene to its original locus, vectors pRECas1-gene_rev were conjugated into the respective mutant following the RiboCas procedure described above.

2.5.2.1 Construction of CRISPR/Cas9 vectors

The seed sequence targeting each gene was designed using the "Design CRISPR Guides" tool provided by Benchling (Table 2.5). Primers containing each seed sequence were designed and named gene_sgRNA_F. The primer dimers gene_sgRNA_F and sgRNA_AsiSI_R were annealed by incubation at 60°C for 1 hour and originated a fragment containing the seed region, the sgRNA scaffold and a terminator; the fragment was flanked by *Sal*I and *Asi*SI recognition sites.

Table 2.5: sgRNA seed sequences designed using Benchling.

Target gene	sgRNA seed sequence
ClosTron (RAM)	TGGATATTCACCGAACACTA
hsdR	ACTTCCTACATCCCAAAGGG

The left homology arm (LHA) was amplified from *C. sporogenes* WT genomic DNA using the primers gene_LHA_F and gene_SOE_R and the right homology arm (RHA) using the primers gene_SOE_F and gene_RHA_R. The PCR products were joined using SOEing PCR (with primers gene_LHA_F and gene_RHA_R), originating a fragment containing the LHA, the two first codons of the target gene, the three last

codons of the target gene and the RHA. The fragment containing the homology arms was flanked by *Asi*SI and *Asc*I recognition sites.

The sgRNA and the homology arms fragments were ligated using *Asi*SI restriction site and introduced into pMTL43151 through the *Sal*I and *Asc*I restriction sites, originating pMTL43151-CLSPO-gene. A generic representation of the CRISPR/Cas9 vectors can be found in the Appendix, Figure S4.

2.5.2.2 Construction of RiboCas vectors

The seed sequence targeting each gene was designed using the "Design CRISPR Guides" tool provided by Benchling (Table 2.6). Primers containing each seed sequence were designed and named gene_sgRNA_F. The primer dimers gene_sgRNA_F and sgRNA_AatII_R were annealed by incubation at 60°C for 1 hour and originated a fragment containing the seed region, the sgRNA scaffold and a terminator; the fragment was flanked by *Sal*I and *Aat*II recognition sites.

Table 2.6: sgRNA seed sequences designed using Benchling.

Target gene	sgRNA seed sequence
sigF	TAAAATAAGTGAAAATACAG
sigE	AAGGAAGTAGCAGATATGCT
sigG	TTACTGGTAAGGATTAGGAA
sigK	AGAACTCAAAGAGAAATAGC
spoIIAA	AAGAAATTATCATTAAGGGA
APP	AATGAAATTCTTGTATGCAA
CLSPO_c33910	AGATGTTAGATTTTTCAAAG

The left homology arm (LHA) was amplified from *C. sporogenes* WT genomic DNA using the primers gene_LHA_F and gene_SOE_R and the right homology arm (RHA) using the primers gene_SOE_F and gene_RHA_R. The PCR products were joined using SOEing PCR, originating a fragment containing the LHA, the first two codons of the target gene, a bookmark, the three last codons of the target gene and the RHA (Figure 2.3). The bookmark was introduced to facilitate reversion of the target gene to the original locus. The homology arms fragment was flanked by *Aat*II and *Asc*I recognition sites.

Finally, the sgRNA and the homology arms fragments were ligated using *Aat*II restriction site and introduced into pRECas1 through the *Sal*I and *Asc*I restriction sites,

originating pRECas1-gene_BM. A generic representation of the RiboCas vectors can be found in the Appendix, Figure S5.

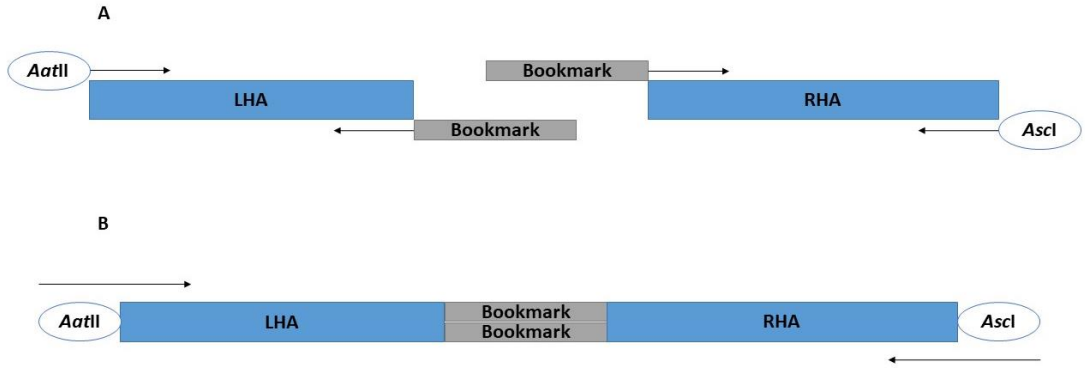


Figure 2.3: Schematic representation of the SOEing PCR method used to join the homology arms. A: The left homology arm (LHA) was amplified using a forward primer containing the *Aat*II restriction site and a reverse primer containing the bookmark sequence. The right homology arm was amplified using a forward primer containing the bookmark sequence and a reverse primer containing the *Asc*I restriction site. **B:** In a second round of PCR, both homology arms were joined due to the complementary bookmark sequences and amplified using the external primers (containing the restriction sites).

2.5.2.3 Construction of vectors used to revert each target gene to the original locus

The construction of the vectors used to revert each gene to its original locus was very similar to the construction of vectors used to knock-out those genes. One difference was the fact that the same sgRNA seed was used for all constructs, and the fragment was obtained using primer-dimer BM_sgRNA_F and sgRNA_AatII_R.

In these constructs, the target gene was inserted between the LHA and the RHA instead of the bookmark, and a silent mutation was introduced in the stop codon of that gene. The homology arms fragments were amplified from *C. sporogenes* WT gDNA using primers gene_LHA_F and gene_SNP_R (LHA); and primers gene_SNP_F and gene_RHA_R (RHA) (Figure 2.4). A generic representation of the RiboCas vectors used to revert the genes can be found in the Appendix, Figure S6.

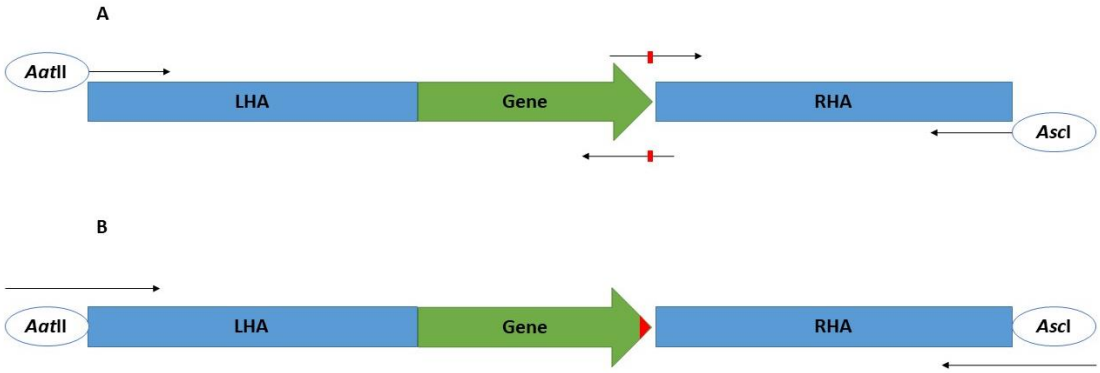


Figure 2.4: Schematic representation of the SOEing PCR method used to introduce a SNP in the stop codon. A: The left homology arm (LHA) was amplified using a forward primer containing the *Aat*II restriction site and a reverse primer containing the SNP (red). The right homology arm was amplified using a forward primer containing the SNP and a reverse primer containing the *Asc*I restriction site. **B:** In a second round of PCR, the both fragments were joined by amplification using the external primers (containing the restriction sites).

2.6 Cloning of vectors used in Chapter 3

To generate the vector pMTL80221-P_{BAD}-*hsdMS*, a fragment containing the P_{BAD} promoter and the *araC* gene was amplified using primers araC_NotI_F and PBAD_BsaI_R, introducing the restriction sites *Not*I and *Bsa*I. A fragment containing the genes *hsdM* and *hsdS* was amplified from genomic DNA extracted from *C. sporogenes* WT using primers hsdM_BsaI_F and hsdS_NheI_R, introducing the restriction sites *Bsa*I and *Nhe*I. The promoter and the genes were ligated using *Bsa*I restriction site and introduced into pMTL80221 through the *Not*I and *Nhe*I restriction sites. A representation of the vector pMTL80221-P_{BAD}-*hsdMS* can be found in the Appendix, Figure S2.

To generate the vector pMTL83151_RS1/2, a fragment was amplified from the *cas9* gene using primers T1_RS1/2_NdeI_F and T1_RS1/2_NotI_R. The fragment was inserted into pMTL83151 through the *Not*I and *Nde*I restriction sites.

To generate the vector pMTL83151_RS3, the primer-dimers T1_RS3_F and T1_RS3_R were annealed to each other by incubation at 50°C for 1 hour. The fragment was inserted into pMTL83151 through the *Not*I and *Nde*I restriction sites.

A generic representation of the vectors containing sequences potentially recognised by the type I restriction system can be found in the Appendix, Figure S3.

2.7 TraDIS

In order to generate the transposon mutant library, *E. coli* sExpress carrying the vector pMTL-CW21 was grown overnight in LB broth supplemented with erythromycin and thiamphenicol. Likewise, *C. sporogenes* WT was grown overnight in TYG broth. The following day, the vector pMTL-CW21 was transferred from *E. coli* sExpress into *C. sporogenes* WT in 20 independent conjugations, as described in section 2.1.6.1. After 7 hours, cells from each conjugation were harvested, resuspended in 500 µL PBS, and spread onto TYG agar plates supplemented with erythromycin to select cells that incorporated the vector. The cells were then grown for 24 hours, harvested, resuspended in 500 µL PBS and serially diluted. The erythromycin-resistant cells were plated onto 3 different selections:

i) TYG supplemented with erythromycin, using the drop plate technique[83];

ii) TYG supplemented with erythromycin and anhydrotetracycline. Diluted cells were plated using the drop plate technique[83], and undiluted cells were spread onto the plates.

iii) TYG supplemented with thiamphenicol and anhydrotetracycline. Diluted cells were plated using the drop plate technique[83], and undiluted cells were spread onto the plates.

Cells were grown again for 24 hours and enumerated in order to calculate plasmid loss percentage and transposition ratio. The undiluted cells growing on thiamphenicol were pooled and resuspended in 40 mL of TYG broth supplemented with anhydrotetracycline and thiamphenicol, originating the transposon mutant library. Genomic DNA was extracted from 1 mL of the mutant library as described in section 2.2.2. The library was then sub-cultured to an OD_{600} of 0.1 in three independent flasks containing 50 mL of TYG broth supplemented with anhydrotetracycline and thiamphenicol. After overnight growth, 2 mL were collected from each culture for genomic DNA extraction. The remaining cultures were incubated in the anaerobic cabinet for 10 days to induce sporulation. After that period, the spores were purified in a sucrose gradient as described in section 2.7.1. Then, the purified spores from each replicate were divided in two aliquots and genomic DNA was extracted from one of the aliquots. For DNA extraction, the spores were homogenised in lysing matrix B (MP Biomedicals), using the Precellys 24 homogeniser (Bertin thechnologies) – 3 cycles (6000 rpm, 45 seconds per cycle). After lysis, the samples were centrifuged

(12000 ×*g*, 5 minutes, 4°C) to remove beads and spore debris. The rest of the procedure was carried out as described in section 2.2.2. The second aliquot of purified spores was used to inoculate 5 mL of TYG broth supplemented with thiamphenicol. After overnight growth, cells were sub-cultured to an OD_{600} of 0.1. After 6 hours of growth, 2 mL were collected from each culture for genomic DNA extraction.

2.7.1 Spore purification

As described above, the cultures were incubated for 10 days to induce sporulation. After that period, the cells were heat-treated at 80°C for 20 minutes and centrifuged (3000 ×*g*, 10 minutes). The pellets were washed twice, first with 10 mL of dH_2O and 10 mL of 95% ethanol, then with 20 mL dH_2O; samples were centrifuged at 3000 ×*g* for 10 minutes after each washing step. The final pellet was resuspended in 1 mL of dH_2O and added to a gradient containing 10 mL of 50% (w/v) sucrose (or 50% (w/v) of HistoDenz, Sigma-Aldrich). The samples were centrifuged at 4000 ×*g* for 20 minutes (4°C) to separate the spores from vegetative cells. The supernatant was carefully removed and the spore pellet was washed in 1 mL of dH_2O. After centrifugation at 5000 ×*g* for 3 minutes, the pellet was resuspended in 44 µL of dH_2O and 5 µL of 10X DNase buffer were added to the spores, as well as 1 µL Turbo DNase (Fisher Scientific). The samples were incubated at 37°C for 30 minutes and then at 75°C for 10 minutes to inactivate the enzyme. The spores were washed again with 1mL of dH_2O and centrifuged at 5000 ×*g* for 3 minutes. The final spore pellet was resuspended in 1 mL dH_2O.

2.7.2 Sequencing library preparation

Genomic DNA samples were fragmented via sonication using a Covaris S-series sonolab. DNA fragments with an average size of around 400 bp were produced using the following settings: 20 W, 200 cycles/burst, intensity 5, 10% duty cycle. A 1 µL sample of sonicated product was run on a 2.5% agarose gel to verify fragmentation as expected. A 1× bead purification was performed using NEB sample purification beads before fragments were end repaired and dA-tailed using NEBNext Ultra II DNA Library Prep reagents. A custom splinkerette adapter[88] was ligated to the ends of fragments using the ligation reagents from the NEBNext Ultra II kit. Adapter-ligated samples were purified again with a 1× bead purification step before an I-*Sce*I digest was performed to selectively cut between the library primer binding site and the

splinkerette adapter on plasmid DNA only. Digested products were again purified using a 1× bead cleanup before a PCR was performed amplifying the transposon-genomic DNA junctions. The PCR employed the Kapa Hifi polymerase which amplified the region between the library primer 5'Tn1 and the adapter primer SplAP5.x where x represents the multiplex number (Table 2.7). SplAP5 primers bind to the adapter and contain a unique region allowing multiplexing of samples. The library primer contains the P5 flow-cell binding region, while the adapter primer contains the P7 flow-cell binding region. PCRs were size-selected on a 2% low-melt agarose gel and then gel extracted using a NEB monarch gel extraction kit. Gel extracted samples were quantified using an agilent 2100 bioanalyzer and a DNA 1000 chip to see the size distribution of the library. The concentration of DNA going onto the sequencing cartridge was assessed via Qubit and qPCR using the flow-cell binding regions to amplify.

Table 2.7: List of primers used in the preparation of the sequencing library for TraDIS.

Primer name	Primer sequence (5'→3')
5'Tn1	AATGATACGGCGACCACCGAGATCTACACTTCGACTGC ACTCCCGACTTAATAACTTC
SplAP5.1	C*AAGCAGAAGACGGCATACGAGATAACGTGATGAGA TCGGTCTCGGCATTC*C
SplAP5.2	C*AAGCAGAAGACGGCATACGAGATAAACATCGGAGA TCGGTCTCGGCATTC*C
SplAP5.3	C*AAGCAGAAGACGGCATACGAGATATGCCTAAGAGA TCGGTCTCGGCATTC*C
SplAP5.4	C*AAGCAGAAGACGGCATACGAGATAGTGGTCAGAGA TCGGTCTCGGCATTC*C
SplAP5.5	C*AAGCAGAAGACGGCATACGAGATACCACTGTGAGAT CGGTCTCGGCATTC*C
SplAP5.6	C*AAGCAGAAGACGGCATACGAGATACATTGGCGAGA TCGGTCTCGGCATTC*C

2.7.3 Analysis and processing of sequencing data

Samples were run on a MiSeq using a custom sequencing primer 5'SeqLong (5'-CTTGAACACTTGCCGAAAAACAATTGTGTCAGACCGGGGACTTATC-3'), which was designed to read out of the transposon giving initially 12 bp of the transposon 'tag' and thereafter genomic DNA. A custom MiSeq recipe was also used which employed dark cycles over the first 12 bp of reads hence preventing run failure from an inability to accurately detect clusters. The dark cycles results in accurate calling of cluster location since after 12 bp the bases are not homogenous throughout the whole sequencing lane.

Reads were filtered for quality and trimmed using TrimGalore (https://github.com/FelixKrueger/TrimGalore). The MiSeq custom recipe already filters for reads that contain the specified transposon tag. Bioinformatic analysis was then undertaken using methods based on those outlied in Barquist *et al.*[88].

2.8 RNA-Seq

In order to sequence the transcriptomes of *C. sporogenes* WT and selected mutants, RNA was extracted from the different strains as described in section 2.2.3. The total RNA samples were then provided to the University of Nottingham DeepSeq facility, where the RNA samples were processed and sequenced. The depletion of rRNA was achieved by treating 1 µg of total RNA with the Ribo-Zero rRNA removal kit (Illumina). Then, Illumina stranded whole transcriptome sequencing libraries were prepared using NEBNext Ultra Directional RNA library prep kit for Illumina (NEB) and barcoded multiplex libraries were generated using Dual Index Primers Set 1 (NEB). The quality of the libraries was assessed using the bioanalyzer HS kit (Agilent Biotechnologies), and the libraries were then quantified using qPCR (Kapa Biosystems). Equimolar library pool was loaded for sequencing according to manufacturer's instructions. Sequencing was performed on the Illumina NextSeq sequencing platform to generate about 10 million 75 paired-end reads per sample.

After sequencing, reads with low sequencing scores and reads aligned to adapter sequences were removed using a filtering pipeline. Raw reads were trimmed using the Skewer tool[89]. After processing, the reads were mapped to the *C. sporogenes* NCIMB 10696 reference genome using the HISAT2 mapping tool[90]. Mapped reads were recorded in BAM formatted alignment files and companion BAM index files. The

accuracy of read mapping was determined according to the map quality of the reads (MAPQ), taking into account the repeat structure of the reference genome, the base quality of the read and the sensitivity of the alignment algorithm; reads with poor mapping quality were filtered.

In order to assess gene expression, the htseq-count program was used to determine the number of uniquely aligned reads per gene. Reads not uniquely aligned and below a MAPQ score of 20 were removed, as well as rRNA. Finally, the results were normalised by calculating the fragments per kilobase of exon per million reads mapped (FPKM). Differential gene expression was calculated using the DESeq package[91], implementing a single analytical approach. Genes were filtered out with an average FPKM of < 1. The similarity between biological replicates was assessed using Principal component analysis (PCA). Finally, heatmaps and hierarchical clustering dendrograms were generated using the 'hclust' R package with a Euclidian distance metric.

2.9 Bioinformatics

Table 2.8: List of bioinformatic tools used in this study.

Bioinformatic tool	Source	Use/description
BLAST	NCBI	Nucleotide or protein similarity searches
CLC Genomics Workbench, version 12	Qiagen	Whole genome assembling and detection of SNVs
GraphPad Prism, version 7	GraphPad Prism	Design of all the graphics presented and statistical analysis
LightCycler® 480, version 1.5	Roche	Analysis of qPCR results
mfold	The RNA Institute	Simulation of DNA folding for primer design
SnapGene, version 3.2.1	SnapGene®	Visualisation of DNA sequences, simulation of cloning experiments, primer design

Chapter 3: Overcoming restriction systems to facilitate gene transfer

3

3.1 Background

3.1.1 Importance of gene transfer for this work

The main aim of the work presented in this thesis is to elucidate sporulation and germination mechanisms in *C. sporogenes*, and Transposon Directed Insertion-site Sequencing (TraDIS) was one of the methods selected to achieve this goal. This approach allows the identification of essential genes by generating a high-density library, of at least one million mutants, through the random insertion of a transposon into the genome. Given the size of the genome and the number of insertions, all genes should be disrupted by the transposon. If mutants in particular genes cannot be isolated, then the mutants did not survive and the genes are probably essential for growth under those conditions[92, 93]. This method will be better described in Chapter 4.
The creation of a library containing one million mutants requires high efficiencies of gene transfer into *C. sporogenes*. However, the efficiencies currently obtained are not high enough to enable approaches such as TraDIS based on suicide plasmid delivery. Thus, the first step towards the study of sporulation and germination in *C. sporogenes* was to increase conjugation and/or transformation efficiencies.

3.1.2 The role of restriction systems as a barrier to gene transfer

Restriction-modification systems are multifunctional enzymes responsible for the modification (methylation) and/or restriction of DNA. By degrading foreign DNA, these complexes constitute a barrier that restricts DNA transfer and inhibits phage infection[94, 95].
There are different types of restriction systems, and the most common are types I, II and III. In these systems, cellular DNA is protected by methylation and unprotected foreign DNA is degraded. In contrast, a fourth type of restriction system only cleaves sequences that have been methylated[94, 95].
It has been previously reported that mutations in certain amino acids of restriction systems, particularly on the restriction subunit, can inhibit the restriction activity[95-97]. Reduced endonuclease activity in the recipient cell can potentially lead to higher transfer frequencies of DNA from donor cells. In fact, it has been previously

demonstrated that mutations in genes encoding restriction systems can lead to increased DNA transformation frequency[98-100].

These findings suggest that restriction systems might be responsible for the relatively low DNA transfer frequencies observed in *C. sporogenes*. Moreover, this barrier to gene transfer can be overcome by mutating the genes encoding those restriction systems.

3.1.2.1 Restriction systems in *Clostridium sporogenes*

REBASE® Genomes[101] was used to identify restriction subunits in *C. sporogenes* NCIMB 10696. Only one type I restriction system could be identified (M. Csp10696ORF17010P). Nevertheless, a DNA repair helicase Rad25 (putative type IV methyl-directed restriction enzyme) was also identified as a restriction subunit (Csp10696ORF6750P). The remaining restriction systems appear to have only methyltransferase activity (Figure 3.1).

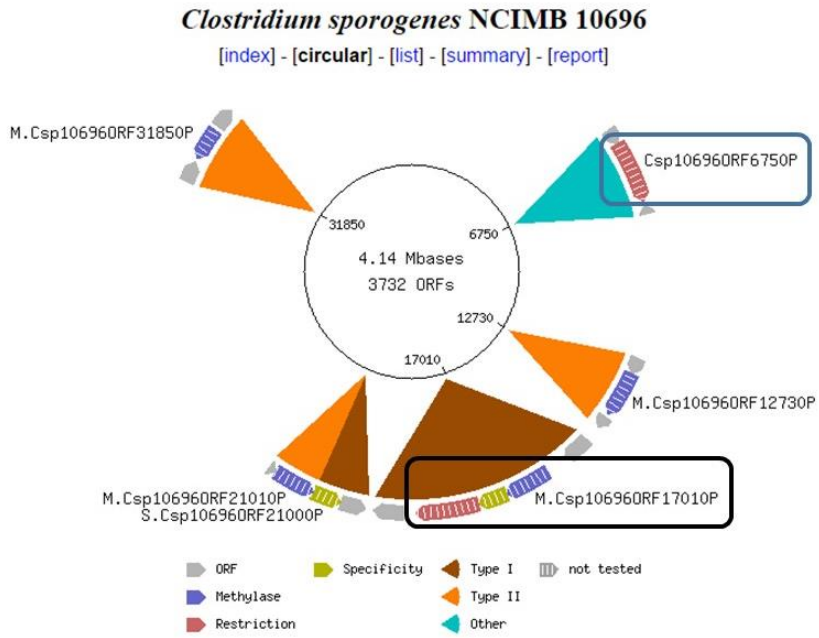


Figure 3.1: Restriction systems in *C. sporogenes* NCIMB 10696. A type I restriction system is highlighted in black and a putative type IV restriction system is highlighted in blue. Source: REBASE® Genomes[101].

As described above, these restriction systems might pose a barrier to DNA transfer in *C. sporogenes*. A brief description of type I and type IV restriction systems is provided below to elucidate their role in this process.

3.1.2.2 Type I restriction systems

Type I systems are multifunctional enzymes often composed of three subunits. One of the subunits, HsdS, is involved in target recognition and specificity. The HsdM subunit, on the other hand, is responsible for the methylation of specific adenine residues in the target sequence (recognised by HsdS) and requires S-Adenosylmethionine (AdoMet) as a methyl donor. Finally, HsdR is the endonuclease that cleaves the DNA, and it requires ATP, AdoMet and Mg^{2+} [94, 95]. Hsd stands for "host specificity for DNA" [94]. It has been suggested that the enzymatic complex contains two subunits of HsdM and HsdR, but only one of HsdS: $R_2M_2S_1$ [95].

The specificity subunit (HsdS) recognises a specific asymmetrical target sequence. That sequence usually contains three or four specific nucleotides, a spacer containing six to eight base pairs, and another specific sequence of four or five nucleotides: 5'XXXXNNNNNNNNXXXXX3' (X represents specific nucleotides and N represents any nucleotide)[95]. When this sequence is recognised, the restriction-modification system binds to it and assesses the methylation status of the sequence:

i) If the specific sequence is hemimethylated, the HsdM subunit is responsible for the methylation of specific adenine residues[95].

ii) If the specific sequence is not methylated, the DNA is recognised as foreign and targeted for restriction. During this process, the enzyme complex translocates across the DNA strand in both directions and cleaves the DNA when it finds an obstacle (this could be another translocase or the topology of the DNA)[95].

The restriction subunit of type I systems (HsdR) is the one responsible for the cleavage of the DNA, and mutations in this subunit have been shown to eliminate restriction activity[96, 97].

The type I system found in *C. sporogenes* appears to be in an operon, with the subunits HsdM, HsdS and HsdR being encoded by the genes CLSPO_c17010, CLSPO_c17020 and CLSPO_c17030, respectively. The structure of the operon diverges from the typical structure of type I restriction systems, where *hsdS* and *hsdM* are in the same operon but *hsdR* is transcribed from an independent promoter[94, 95]. The modification subunit is annotated as a putative type I restriction enzyme MjaXP M protein, while the specificity subunit is annotated as type-I restriction enzyme *Ecoprr*I specificity

protein PrrB. The enzyme *Ecoprr*I belongs to the type IC family[95]. However, its recognition sequence has not been identified in the REBASE database[101].

3.1.2.3 Type IV restriction systems

Type IV restriction systems are characterised for the cleavage of DNA that contains modified cytosine or adenosine nucleotides. These systems may contain several subunits and genes but, not surprisingly, none of them correspond to modification enzymes[94].

The putative type IV system found in *C. sporogenes* is annotated as a DNA repair helicase RAD25. The analysis of the conserved domains of the DNA repair helicase Rad25 on NCBI[102] (Figure 3.2) revealed that the protein is very similar to SauUSI, which was classified as a type IV restriction enzyme[103]. This enzyme cleaves methylated cytosine nucleotides (5mC) and requires ATP hydrolysis and divalent cations in order to be fully active[103].

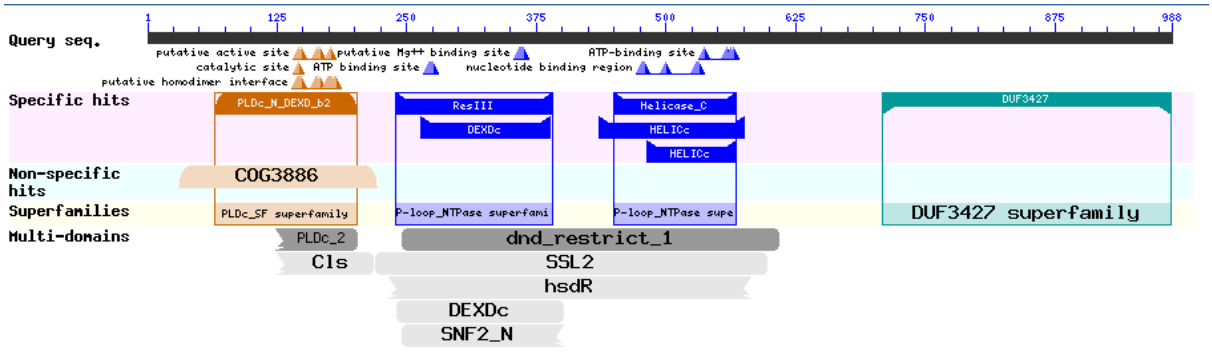


Figure 3.2: Functional domains identified in the DNA repair helicase Rad25, a putative type IV restriction system from *C. sporogenes*. Several domains associated with restriction activity are present, such as the phospholipase D family, hsdR and DEXD motifs. Source: CD-search[102].

Both SauUSI[103] and the DNA repair helicase Rad25 (Figure 3.2) contain a phospholipase D (PLD) family in the N-terminal region. Previous studies suggest that this region may be responsible for DNA cleavage, as mutations in conserved amino acids led to loss of endonuclease activity[96]. A conserved domain family DUF3427 in the C-terminal can be found in both proteins and it might be responsible for the recognition of the restriction site[103]. Also common to SauUSI and the DNA repair helicase Rad25 are helicase domains containing DEAD boxes, an Mg^{2+} binding site and a conserved motif of type III restriction enzymes. It has also been found that DEAD-box motifs appear to be essential for the cleavage of DNA by the HsdR subunit of type I restriction systems[95].

Overall, the analysis of the conserved domains indicates that although the protein is annotated as a DNA repair helicase, it probably functions as a type IV restriction enzyme that targets methylated cytosine residues.

3.1.3 Strategies to overcome restriction

As mentioned above, DNA methylation patterns play a relevant role in the restriction of foreign DNA in bacteria. One of the ways to overcome this barrier is to mutate the genes encoding restriction enzymes, as these are the ones responsible for the cleavage of unmethylated (type I) or methylated (type IV) DNA. This approach has been successfully used before[99], including in clostridia[100, 104]. Targeted mutagenesis approaches such as ClosTron and CRISPR/Cas9 can be used to mutate the restriction subunits, so these techniques are described below.

A different approach is to alter the methylation pattern of the DNA prior to its transfer to the recipient cell. If the recipient cell recognises the DNA as its own, then that DNA will not be targeted by the restriction systems. This approach has also been previously used[100, 105]. It has been shown above that *C. sporogenes* has two restriction systems, a type I and a type IV. The HsdM subunit of the type I system is responsible for the methylation of the DNA, and HsdS is responsible for its recognition – both subunits are essential for the methyltransferase activity[95]. Hence, expressing these proteins in a donor strain would ensure that the DNA would have a specific methylation pattern that would not be restricted by HsdR in the recipient cell. The type IV system, on the other hand, recognises and targets DNA that is methylated at cytosine nucleotides[103]. The methylation of cytosine nucleotides in *E. coli* is mediated by the DNA cytosine methyltransferase (Dcm), so the use of *E. coli* strains lacking this protein as donors should avoid restriction by the type IV system. This should lead to increased DNA conjugation efficiencies. A summary of the strategies to overcome the restriction barrier in *C. sporogenes* is presented in Figure 3.3.

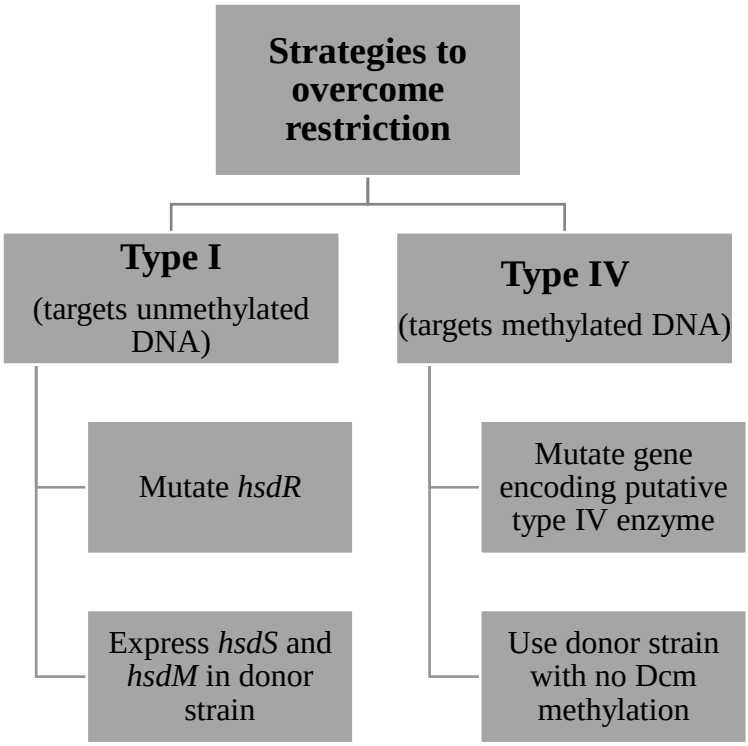


Figure 3.3: Summary of the strategies that can be used to overcome the restriction barriers found in *C. sporogenes*. The barrier posed by the type I and type IV systems can be averted by mutating the genes responsible for restriction. In alternative, the methylation pattern of the DNA can be changed before transfer to avoid restriction.

3.1.3.1 ClosTron

The ClosTron technique allows directed mutagenesis and can be used to target restriction systems. This technique is based on the properties of the group II intron, which can excise itself from RNA transcripts and be inserted somewhere else in the genome. The intron can be manipulated to insert itself into specific target genes, disrupting them and potentially affecting their function. A big advantage of this technique is the fact that the insertion of the intron into the target gene can be selected by using a retrotransposition-activated marker (RAM). The RAM, which is based on the *ermB* gene (confers resistance to erythromycin), is only activated during retrotransposition (*i.e.*, is not expressed when the group II intron is in the plasmid, only when it is inserted into the target gene) (Figure 3.4)[84, 85]. This method is easy and reproducible, and the resulting constructs are predicted to be stable[84].

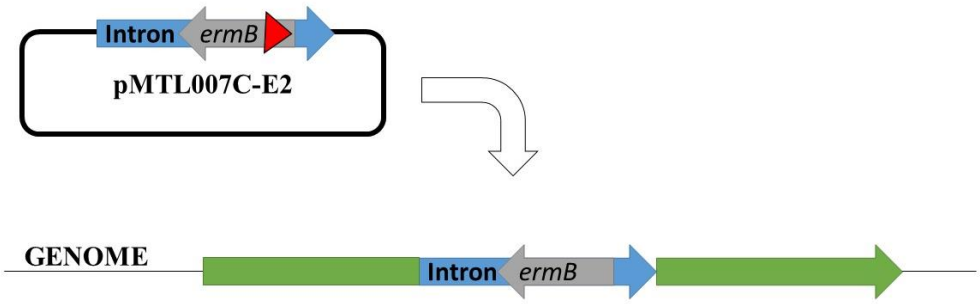


Figure 3.4: Representation of the ClosTron system. The group II intron (blue) is transferred from the plasmid (pMTL007C-E2) into the genome, and it is designed to target a specific gene (green). When the group II intron is in the plasmid, the *ermB* marker (grey) is disrupted by a group I intron (red) and does not confer resistance to erythromycin. The group I intron is spliced out during retrotransposition, so the insertion of the group II intron (and the marker) in the genome results in erythromycin resistance. Adapted from Kuehne *et al.*[106].

3.1.3.2 CRISPR/Cas9

The ClosTron approach mutates genes by inserting a DNA sequence that disrupts them. However, other approaches allow the in-frame deletion of target genes, such as the CRISPR/Cas9 system. The CRISPR/Cas9 gene editing technology is based on a natural immunity system that bacteria developed to protect themselves from foreign DNA[107]. CRISPR stands for "Clustered Regularly Interspaced Palindromic Repeat", which are regions of the genome that contain repeat sequences separated by short DNA spacers[108]. The spacers have been shown to have originated from viral or plasmid DNA[107]. By keeping a "memory record" from previous DNA invasions, the cells are able to respond faster and more effectively when the same source of DNA tries to re-enter the cell. If this happens, the crRNA (transcribed from CRISPR arrays) will recognise the spacer sequence in the foreign DNA. Then, CRISPR-associated (Cas) proteins are guided by the crRNA and introduce double-strand breaks in the invader DNA, degrading it[109].

Some CRISPR systems require several Cas proteins, but type II systems are functional with only one protein – Cas9. Type II systems also require trans-activated crRNA (tracrRNA) for the maturation of crRNA and to stimulate the binding of Cas9 to the target sequence[109, 110]. CRISPR-mediated gene editing is based on the type II CRISPR/Cas9 system from *Streptococcus pyogenes*[110]. In this system, a single guide RNA (sgRNA) has the dual function of crRNA and tracrRNA. Furthermore, the sgRNA only recognises the target sequence if there is a protospacer adjacent motif (PAM) next to it. Consequently, any 20 bp sequence can be targeted using CRISPR/Cas9, provided that there is an adjacent PAM sequence[109, 110].

In bacteria, this technology is used in combination with plasmid-borne editing templates[111]. The CRISPR/Cas9 sgRNA is designed to target the gene of interest, and homology arms to create in-frame deletions of the same gene are introduced into the cell as part of a vector. Cells that undergo homologous recombination no longer have the sequence that is targeted by the CRISPR/Cas9 system, so they survive. Cells that maintain the target gene, however, will die due to the introduction of a lethal double-strand break by Cas9[111].

The system used in this study to target genes encoding restriction systems has been previously described by Ingle *et al.*[86]. This system uses a truncated version of Cas9 (trCas9), which is hypothesised to function as a nickase to introduce single-strand rather than double-strand breaks[86]. A representation of the system used in the present study is shown in Figure 3.5. An editing template containing the left homology arm, part of the target gene (first two and last three codons) and the right homology arm were included in the gene-editing vector[86, 111].

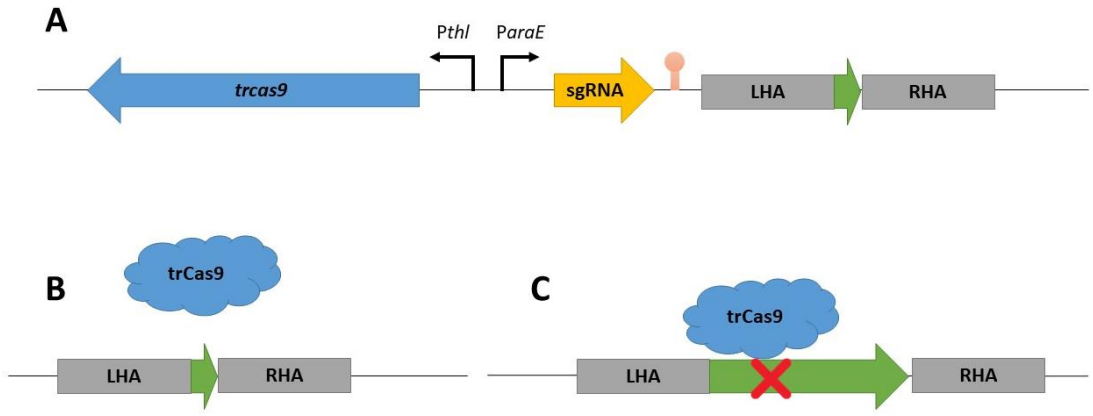


Figure 3.5: Representation of the CRISPR/Cas9 system. A: Part of the vector used for gene editing, containing the CRISPR/Cas9 components. The gene encoding the nuclease is a truncated version of *cas9*, named *trcas9*, and it is represented in blue. The constitutive promoter P_{thl} controls the expression of this gene. The sgRNA is under the control of the P_{araE} promoter and has a terminator placed downstream. The gene editing template contains part of the target gene (first two and last three codons), represented in green, and the two homology arms that allow homologous recombination (LHA and RHA), represented in grey. **B**: Part of the genome after homologous recombination. The sequence recognised by the sgRNA has been removed along with the target gene, so trCas9 does not introduce nicks in the genome. **C**: Part of the genome in cells where homologous recombination did not occur. The target gene is still present, so the trCas9 is able to introduce lethal nicks in the genome.

3.2 Chapter overview

With the aim of increasing DNA transfer frequency in *C. sporogenes*, genes encoding the R subunit of the type I restriction system and the putative type IV restriction system were targeted for directed mutagenesis using ClosTron and CRISPR/Cas9. Moreover, the methylation pattern of DNA to be transferred was changed to prevent restriction by either system. Sporulation assays were performed to ensure that the mutants did not affect the ability of cells to sporulate. Finally, assays to determine conjugation and transformation efficiencies allowed the evaluation of the different strategies in regards to gene transfer. Strains with increased conjugation efficiencies were selected as potential candidates to obtain a transposon mutant library and study sporulation and germination in *C. sporogenes*.

3.3 Results

3.3.1 Overcoming the type IV restriction system

The type IV system was the first one to be targeted because it only contains one gene and it only requires that the DNA is not methylated at cytosine nucleotides. It is, in theory, the easier restriction system to overcome. The first step was to target the gene encoding the putative type IV restriction system with ClosTron, which was the gene editing tool available at the time. Then, when CRISPR/Cas9 became available, an in-frame deletion of the gene was created. Finally, different donors were tested to evaluate the effect of different methylation patterns on conjugation efficiency. Some of the following results have been published[112].

3.3.1.1 Targeting the type IV system using ClosTron

The gene CLSPO_c06750 encodes a DNA repair helicase Rad25, also annotated as a putative type IV restriction enzyme. The plasmid pMTL007::*Csp*-c06750-1240a was obtained from DNA2.0 and used to generate the strain *C. sporogenes*-c06750-1240a::CT. Three independent biological replicates were created using this strategy. The insertion of the ClosTron into the target gene was confirmed by colony PCR (Figure 3.6 - C) and by Sanger sequencing.

When using ClosTron, one of the concerns is that the intron might be inserted into more than one gene in the genome, creating an undesired mutation. The single insertion of the intron into the target gene was evaluated using inverse PCR. The resulting gel picture (Figure 3.6 - D) shows only one band for biological replicate 1, but two bands for biological replicates 2 and 3. All the bands were excised and the corresponding DNA was sequenced, revealing that the bigger band (present in all replicates; approx. 7 Kb) corresponds to the insertion of the intron into the target gene. The DNA from the smaller band (approx. 500 bp) could not be aligned to the genome of *C. sporogenes* and appears to correspond to the intron. It is unclear why the intron was amplified during the inverse PCR. A possible explanation is the unspecific binding of the primers used to amplify the DNA, but in that scenario the smaller band would be expected to be present in all the biological replicates. Nevertheless, the band does not correspond to an additional insertion of the ClosTron into the genome of *C. sporogenes*, indicating that the intron is only present in the target gene.

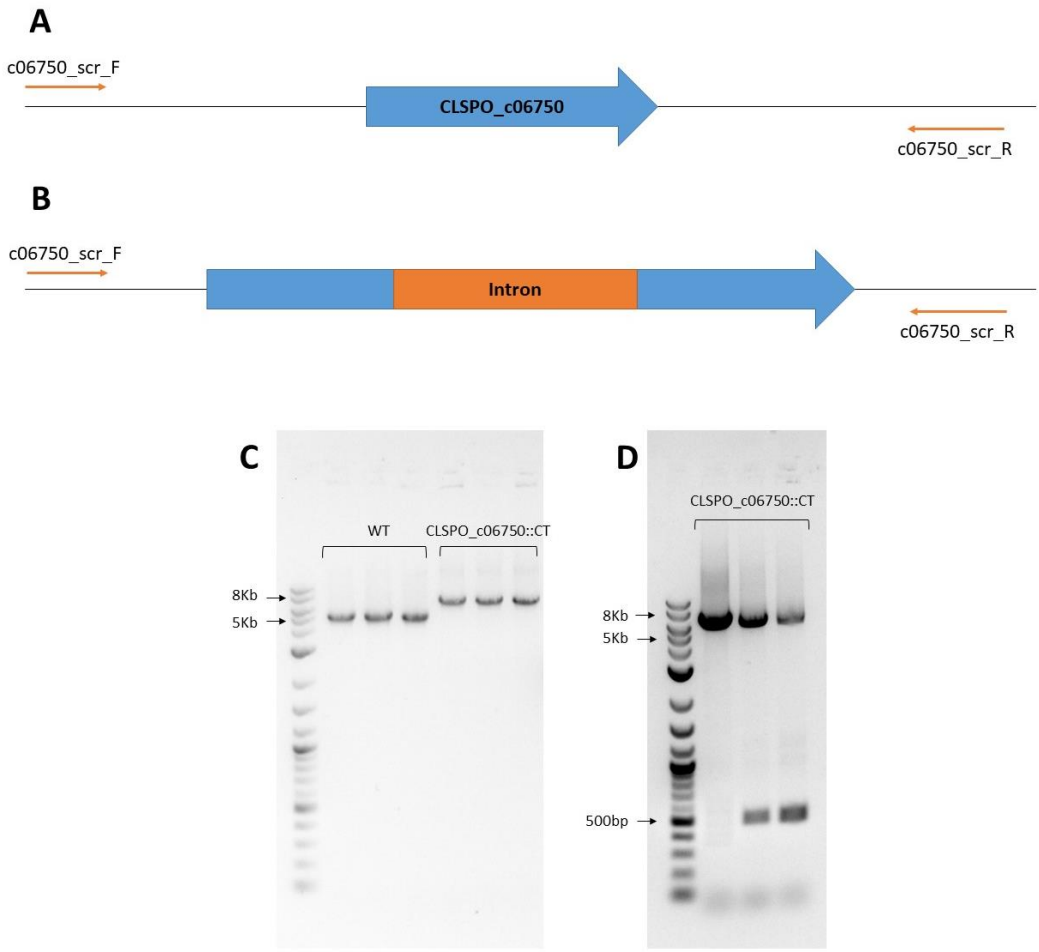


Figure 3.6: Representation of the CLSPO_c06750 gene with (B) and without (A) the intron insertion, including the primers used to screen the mutants. **C**: Gel electrophoresis of colony PCR samples demonstrating the insertion of the ClosTron into CLSPO_c06750 (CLSPO_c06750::CT). The strains were screened using primers c06750_scr_F and c06750_scr_R. Wild type was also included (WT). **D**: Gel picture of inverse PCR showing the number of ClosTron insertions in each biological replicate of the strain *C. sporogenes*-c06750-1240a::CT. Primers erm_inv_F and erm_inv_R were used to amplify the DNA.

3.3.1.1.1 Sporulation assay

The purpose of creating mutants with increased DNA transfer frequency was to use them to study sporulation and germination in *C. sporogenes*. In order to evaluate if mutations in the CLSPO_c06750 gene affect sporulation and/or germination, a sporulation assay was conducted. The results (Figure 3.7) indicate that the mutation has a slight effect on sporulation/germination, as the mutant produces significantly fewer spores at early time points (24 hours and 48 hours; unpaired t-tests performed on GraphPad 7).

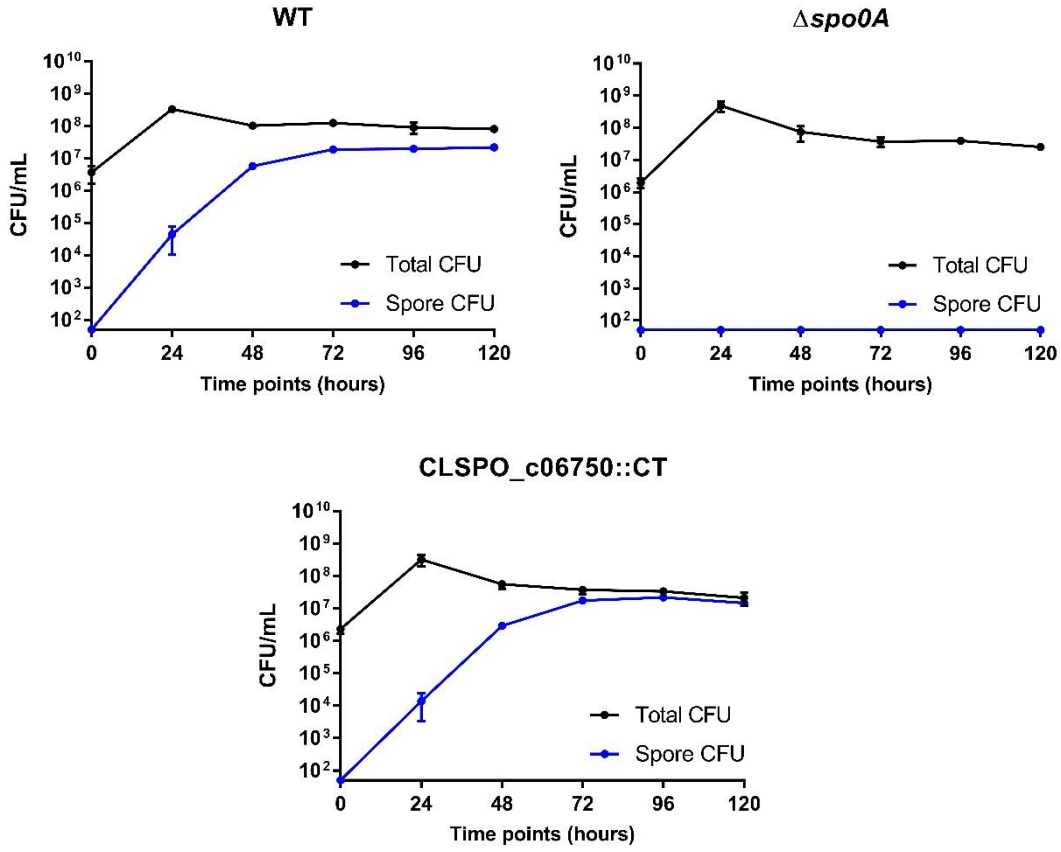


Figure 3.7: Sporulation profile of the *C. sporogenes*-c06750-1240a::CT strain. *C. sporogenes* WT and *C. sporogenes* Δ*spo0A* were included as positive and negative controls, respectively. All the cultures were synchronised at the beginning of the assay, and samples were collected at 24 hour intervals during a total period of 120 hours. For the total CFU count, the cultures were diluted and spotted into TYG agar. For spore CFU count, the cultures were heat-treated at 80°C for 20 minutes prior to the dilutions. Data presented shows the mean and error bars represent the standard deviation of three biological and three technical replicates. The Y axis starts at the limit of detection (50 spores/mL).

3.3.1.1.2 Determination of conjugation efficiency

The effect of the ClosTron insertion into the type IV system was assessed by determining the conjugation efficiency of the strain and comparing it to wild type. This was achieved by conjugating the modular vector pMTL83151 from *E. coli* CA434 into *C. sporogenes* WT and *C. sporogenes*-c06750-1240a::CT. This vector was selected because it contains the pCB102 replicon, known to be stable in *C. sporogenes*. Furthermore, the vector confers resistance to thiamphenicol and chloramphenicol, allowing the distinction between transconjugants and cells that did not receive the plasmid. The assay tested three biological replicates per strain, and three technical replicates per biological replicate.

Following conjugation, all the cultures were serially diluted and 20 µL of each dilution were spotted onto plates containing different selective media, as described in

Chapter 2. The number of transconjugants was determined by counting the number of *C. sporogenes* colony forming units (CFUs) that were able to grow in the presence of thiamphenicol. Different amounts of donor or recipient cells during conjugation might affect efficiency. Hence, in order to be sure that the difference in the number of transconjugants was a consequence of the mutation and not variations in the number of donors/recipients, the number of donor CFUs (*E. coli* CA434 plated onto LB media containing chloramphenicol) and the number of total CFUs were also determined. The number of total CFUs corresponded to all the *C. sporogenes* cells that grew after conjugation with no antibiotic selection. Conjugation efficiencies per recipient and per donor were calculated according to the following equations:

$$Conjugation\ efficiency\ per\ recipient = \frac{CFU/mL\ Transconjugants}{CFU/mL\ Total\ CFUs}$$

$$Conjugation\ efficiency\ per\ donor = \frac{CFU/mL\ Transconjugants}{CFU/mL\ Donor\ CFUs}$$

The results presented in Figure 3.8 show no significant difference between the number of total CFUs in wild type and in the ClosTron mutant (unpaired t-test performed using GraphPad Prism 7). Likewise, a one-way ANOVA indicated that the differences between the biological replicates of the donor strain (*E. coli* CA434) were not significant (p-value = 0.5233). Hence, any difference in the number of transconjugants is likely due to the mutation of the CLSPO_c06750 gene.

The results also indicate that the ClosTron mutant originates an increased number of transconjugants relative to wild type (176-fold). Not surprisingly, the conjugation efficiencies per donor and per recipient were also higher in the *C. sporogenes*-c06750-1240a::CT strain (841 and 55-fold, respectively). These differences between mutant and wild type were statistically significant, with a p-value < 0.0001 (unpaired t-test) (Figure 3.8).

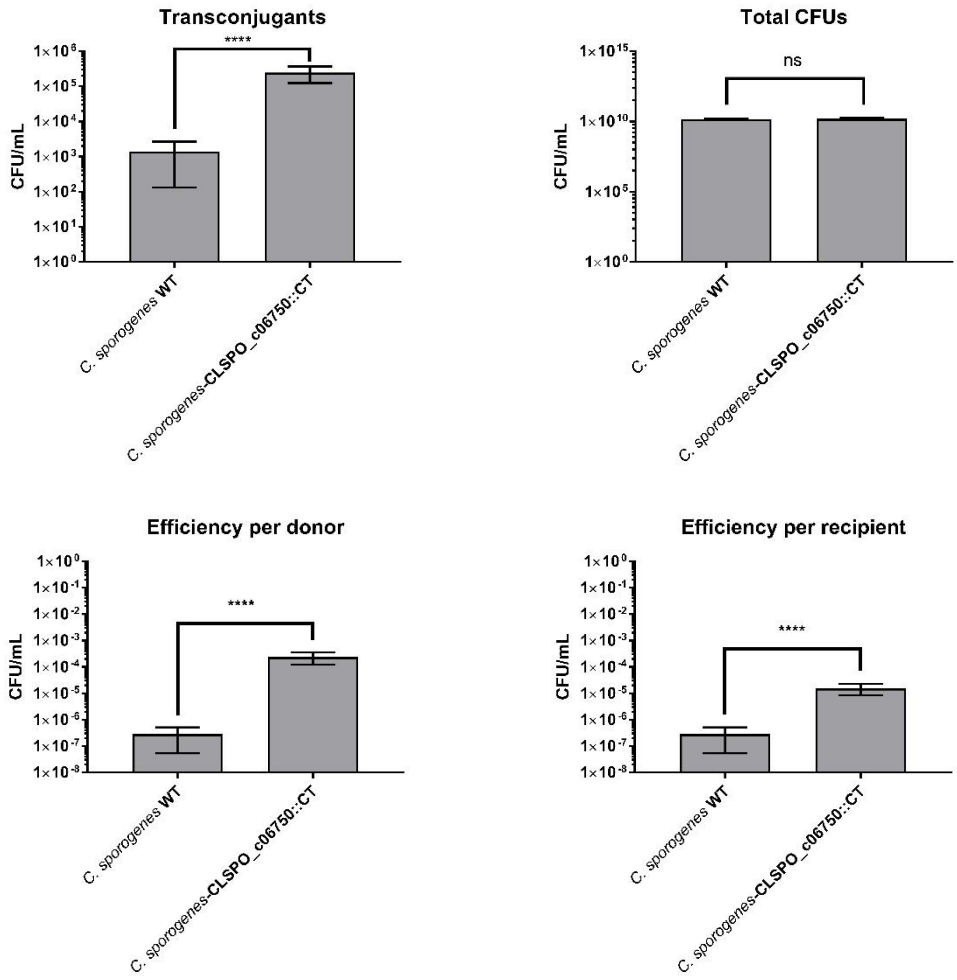


Figure 3.8: Comparison between conjugation efficiencies in *C. sporogenes* WT and *C. sporogenes*-c06750-1240a::CT. The number of transconjugants is presented, as well as the number of total CFUs. Data presented shows the mean and error bars represent the standard deviation of three biological and three technical replicates. Unpaired t-tests were performed using GraphPad 7 to compare the mutant to wild type: ns represent non-significant differences (p-value > 0.05), while **** represent significant differences (p-value < 0.0001).

3.3.1.1.3 Determination of transformation efficiency

The ability of the *C. sporogenes*-c06750-1240a::CT strain to receive DNA via transformation was also assessed. With this aim, the vector pMTL83151 was electroporated into *C. sporogenes* WT and *C. sporogenes*-c06750-1240a::CT. The resulting cells were plated onto TYG (in order to count total CFUs) and onto TYG supplemented with thiamphenicol (in order to count the transformants). For each construct, three biological and three technical replicates were evaluated.

Transformation efficiencies were calculated according to the equation:

$$Transformation\ efficiency = \frac{CFU/mL\ Transformants}{CFU/mL\ Total\ CFUs}$$

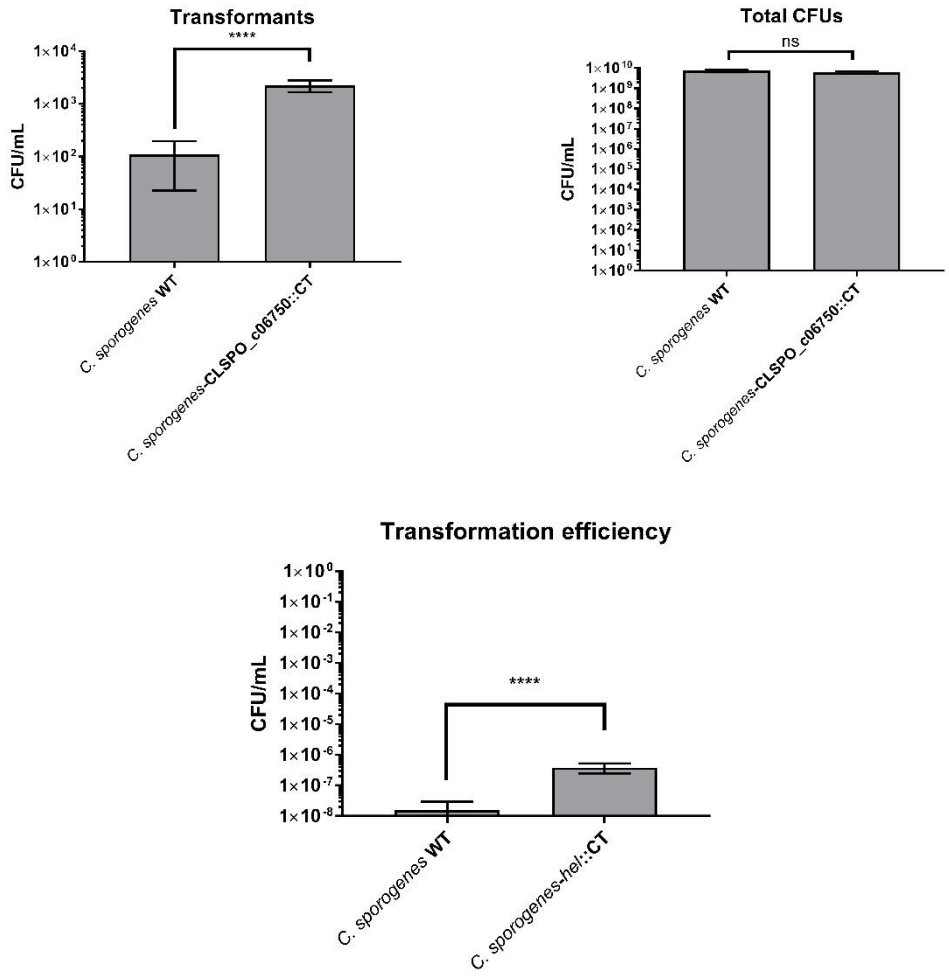


Figure 3.9: Comparison between transformation efficiencies in *C. sporogenes* WT and *C. sporogenes*-c06750-1240a::CT. The number of transformants is presented, as well as the number of total CFUs. Data presented shows the mean and error bars represent the standard deviation of three biological and three technical replicates. Unpaired t-tests were performed using GraphPad 7 to compare the mutant to wild type: ns represent non-significant differences (p-value > 0.05), while **** represent significant differences (p-value < 0.0001).

The results show that the number of total CFUs is the same for both strains (Figure 3.9), which indicates that the amount of competent cells was the same for all transformations. Hence, any variation in the number of CFUs plated onto selective medium results only from variations in plasmid uptake frequency. The number of transformants obtained with *C. sporogenes*-c06750-1240a::CT was 20-fold higher than wild type, and the same was observed for the transformation efficiency (24-fold increase) (Figure 3.9).

Overall, the results indicate that the product of the CLSPO_c06750 gene provides a barrier to gene transfer, as the ClosTron mutants present increased conjugation and transformation efficiencies relative to wild type. Moreover, higher efficiencies were obtained with conjugation (conjugation efficiency per recipient in the CLSPO_c06750

mutant: $1.55\times10^{-5} \pm 7.09\times10^{-6}$; Figure 3.8) than with transformation (transformation efficiency in the CLSPO_c06750 mutant: $3.86\times10^{-7} \pm 1.38\times10^{-7}$; Figure 3.9). Given its higher relative transfer frequency and the fact that the transformation protocol is more laborious, conjugation was selected as the preferred method for gene transfer into *C. sporogenes*.

3.3.1.2 Targeting the type IV system using CRISPR/Cas9

After the generation of the ClosTron mutant targeting the putative type IV restriction system, a CRISPR/Cas9 gene editing tool was developed in the SBRC group[86]. The system is described above and it allows the in-frame deletion of the gene, avoiding the polar effects potentially caused by the insertion of the intron into target genes.

As previously described, Cas9 is able to introduce double-strand breaks in the genome[109] or, in the case of trCas9, single-strand nicks[86]. The consequent toxicity associated with the use of Cas9 or trCas9 is expected to decrease the number of transconjugants/transformants obtained when using the CRISPR/Cas9 system. This issue can be overcome by using hosts with higher conjugation or transformation efficiencies, such as the CLSPO_c06750 ClosTron mutant. Accordingly, the sgRNA seed was designed to target the intron inserted during ClosTron mutagenesis. The vector pMTL43151-CLSPO-c06750CT was generated: it targets the ClosTron intron and contains a gene editing template designed to knock-out the CLSPO-c06750 gene. This vector was conjugated into *C. sporogenes*-c06750-1240a::CT, and three independent replicates of *C. sporogenes* Δc06750 were obtained. The deletion of the target gene was confirmed by colony PCR (Figure 3.10) and by Sanger sequencing.

3.3.1.2.1 Sporulation assay

Similar to what was described for the ClosTron mutant, the impact of the in-frame deletion of CLSPO_c06750 on sporulation was assessed. The sporulation assay results (Figure 3.11) indicate that the deletion of the gene has an impact on the ability of the cells to form heat-resistant spores (statistically significant difference at time points 48 hours, 72 hours, 96 hours and 120 hours; unpaired t-tests performed on GraphPad 7).

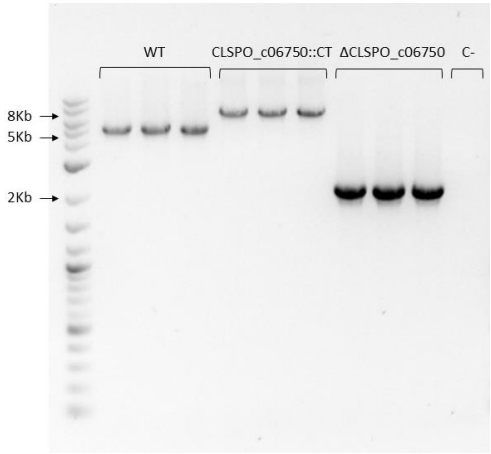


Figure 3.10: Gel electrophoresis of colony PCR samples demonstrating the insertion of the ClosTron into CLSPO_c06750 (CLSPO_c06750::CT) and the in-frame deletion of the same gene using CRISPR/Cas9 (ΔCLSPO_c06750). The strains were screened using primers c06750_scr_F and c06750_scr_R. A positive control was included (WT), using genomic DNA from *C. sporogenes* WT. A negative control (water as template; C-) was also included.

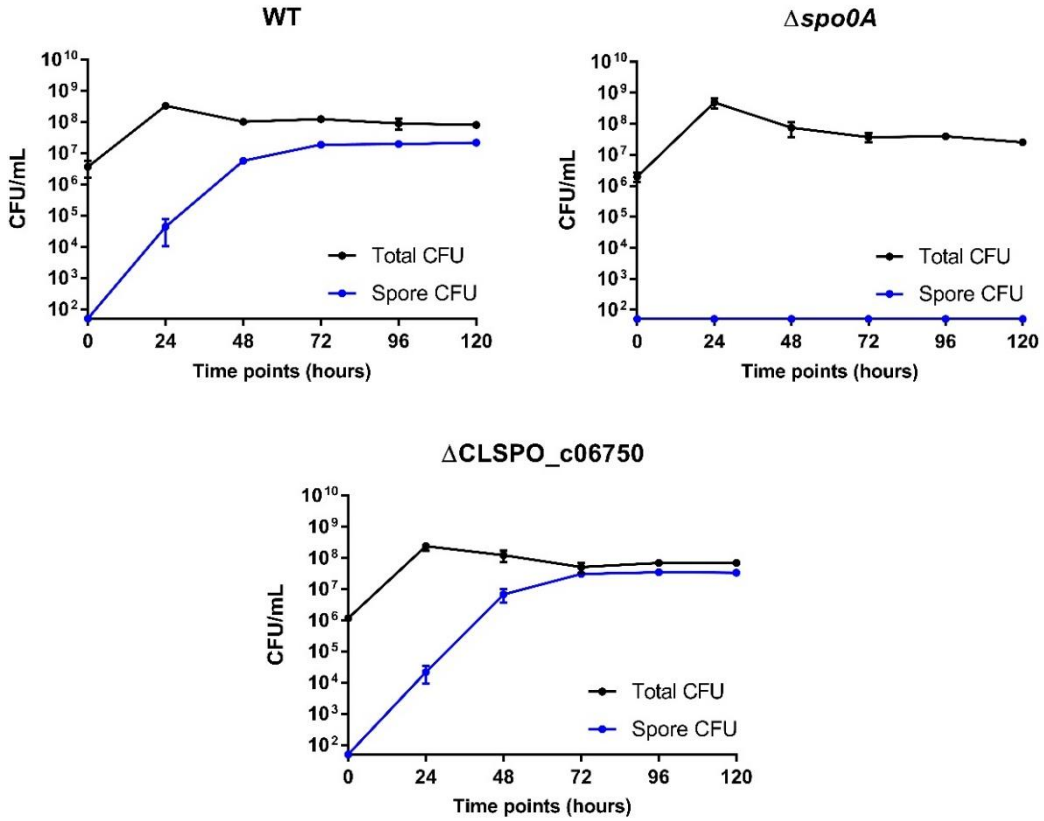


Figure 3.11: Sporulation profile of the *C. sporogenes* Δc06750 strain. *C. sporogenes* WT and *C. sporogenes* Δ*spo0A* were included as positive and negative controls, respectively. All the cultures were synchronised at the beginning of the assay, and samples were collected at 24 hour intervals during a total period of 120 hours. For the total CFU count, the cultures were diluted and spotted into TYG agar. For spore CFU count, the cultures were heat-treated at 80°C for 20 minutes prior to the dilutions. Data presented shows the mean and error bars represent the standard deviation of three biological and three technical replicates. The Y axis starts at the limit of detection (50 spores/mL).

3.3.1.2.2 Determination of conjugation efficiency

The conjugation efficiency of the *C. sporogenes* Δc06750 strain was determined in the same way as the conjugation efficiency of *C. sporogenes*-c06750-1240a::CT.

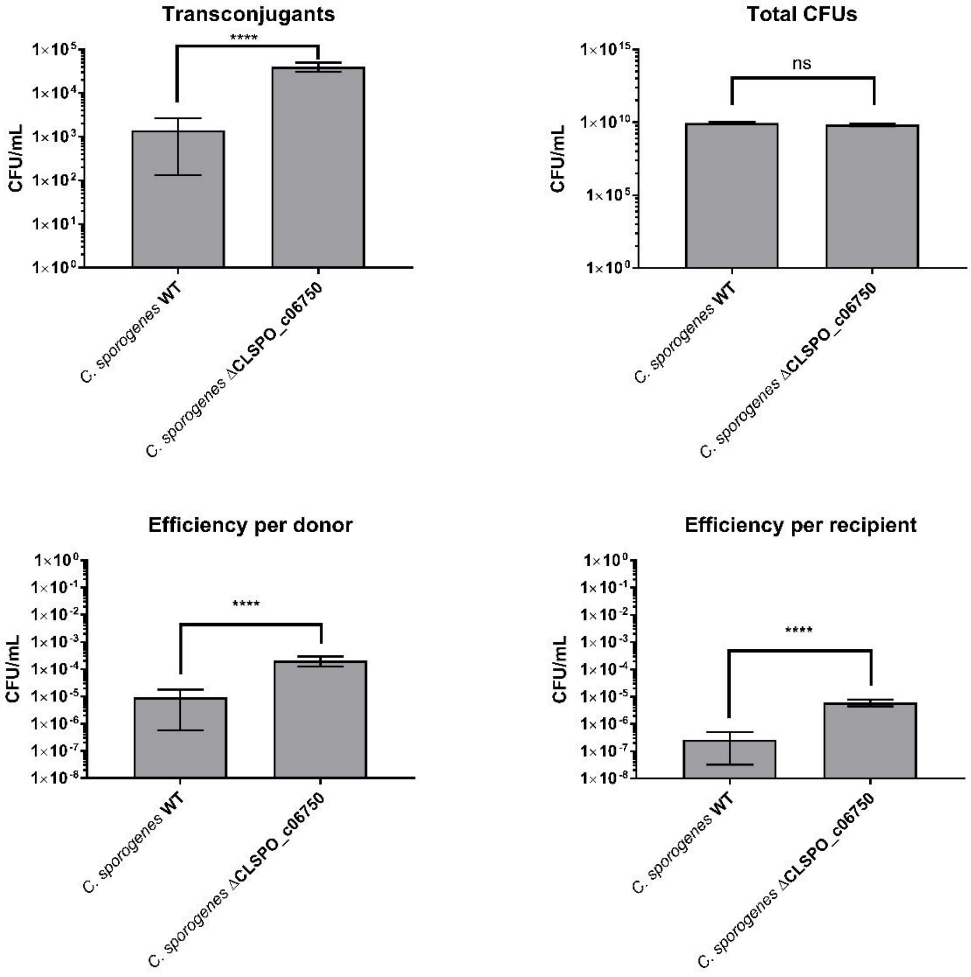


Figure 3.12: Comparison between conjugation efficiencies in *C. sporogenes* WT and *C. sporogenes* Δc06750. The number of transconjugants is presented, as well as the number of total CFUs. Data presented shows the mean and error bars represent the standard deviation of three biological and three technical replicates. Unpaired t-tests were performed using GraphPad 7 to compare the mutant to wild type: ns represent non-significant differences (p-value > 0.05), while **** represent significant differences (p-value < 0.0001).

As expected, the results are similar to those obtained with the ClosTron mutant: the number of transconjugants increased by 29-fold in the *C. sporogenes* Δc06750 strain relative to wild type. The conjugation efficiencies per donor and per recipient were also higher in the mutant strain by 23-fold (Figure 3.12).

3.3.1.3 Testing an *E. coli* strain that avoids methylation in cytosine nucleotides: sExpress

Some *E. coli* strains are known to have *dcm* and/or *dam* genes that methylate cytosine and adenosine residues, respectively. The strain used as a conjugative donor above, *E. coli* CA434, is known to have both these systems[112]. The putative type IV system found in *C. sporogenes* is hypothesised to recognise methylated cytosine residues and cleave the modified DNA[103], so any Dcm methylated DNA from *E. coli* CA434 could be degraded by the product of the CLSPO_c06750 gene.

The use of a donor strain lacking Dcm methylation was hypothesised to be an alternative to the deletion of the gene encoding the type IV restriction system in the recipient strain. A strain created in the SBRC lab for this purpose, *E. coli* sExpress[112], was selected to test this theory. Accordingly, the conjugation assays described above were repeated and the vector pMTL83151 was transferred from *E. coli* CA434 or *E. coli* sExpress into *C. sporogenes* WT.

The results presented in Figure 3.13 show that the number of total CFUs after conjugation and the number of CFUs for each donor strain were the same for all the replicates (unpaired t-test; p-value > 0.05). The number of transconjugants, on the other hand, was two orders of magnitude higher when using sExpress as the donor strain in comparison to CA434. This increase was also observed after calculating the conjugation efficiencies per donor and per recipient (Figure 3.13).

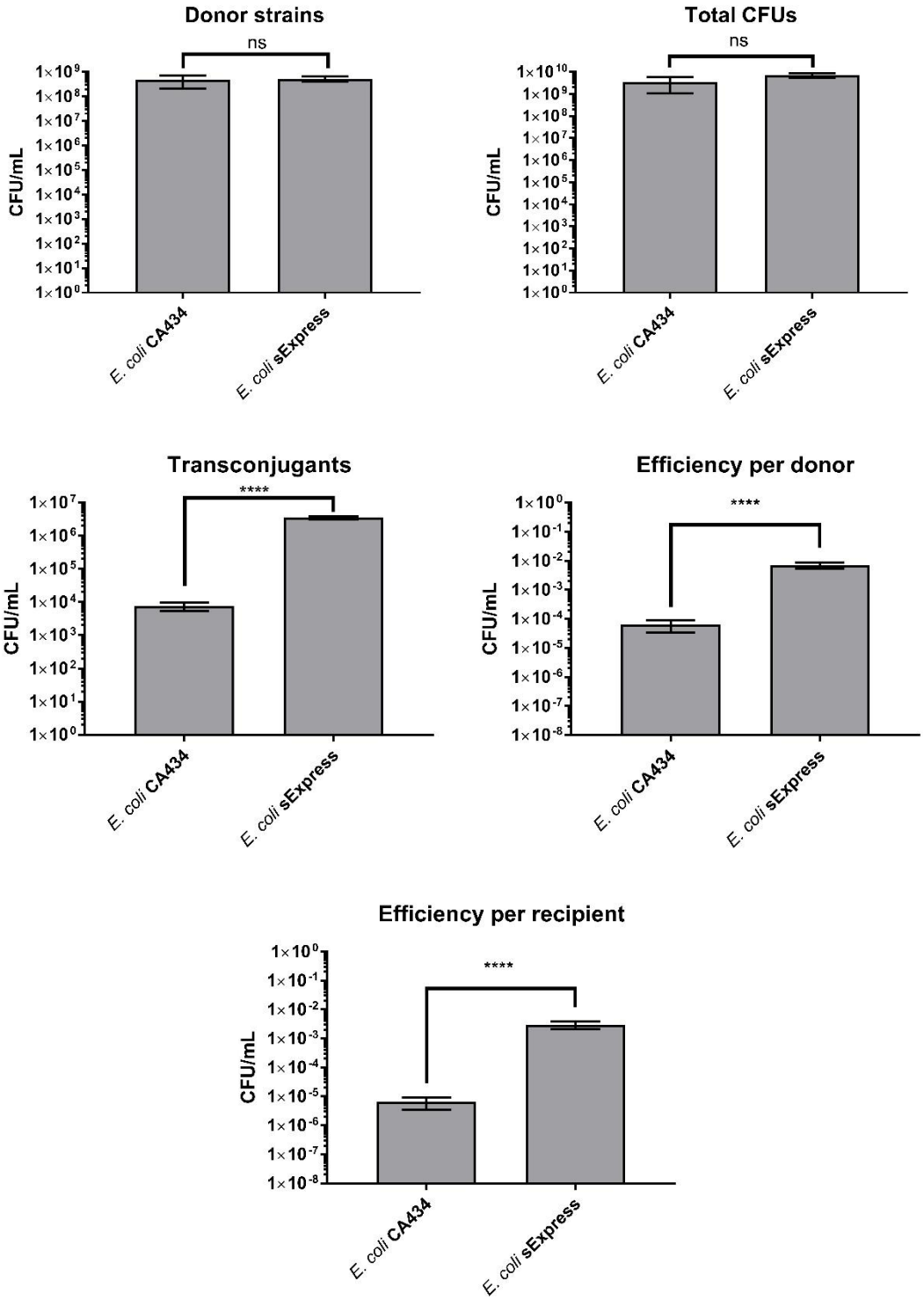


Figure 3.13: Comparison between conjugation efficiencies using *E. coli* CA434 or *E. coli* sExpress as donor strains. The number of transconjugants is presented, as well as the number of total CFUs and donor strains. Data presented shows the mean and error bars represent the standard deviation of three biological and three technical replicates. Unpaired t-tests were performed using GraphPad 7 to compare both donor strains: ns represent non-significant differences (p-value > 0.05), while **** represent significant differences (p-value < 0.0001).

3.3.1.4 Comparison between the different approaches to overcome the type IV restriction system

The increase in conjugation efficiency obtained with the sExpress donor strain was comparable to the increase in conjugation efficiency obtained with the CLSPO_c06750::CT strain. This observation suggests that the product CLSPO_c06750 really is a type IV restriction system that targets methylated cytosine residues, and that this system can be evaded by avoiding Dcm methylation in the donor strain. In order to confirm this, all the approaches described above were compared in the same conjugation assay. Accordingly, two donor strains (*E. coli* CA434 and *E. coli* sExpress) and three recipient strains (*C. sporogenes* WT, *C. sporogenes*-c06750-1240a::CT and *C. sporogenes* Δc06750) were included in the same assay in order to compare their conjugation efficiencies.

The results presented in Figure 3.14 suggest that the effect of the use of sExpress as a donor strain is the same as the mutation of the gene CLSPO_c06750, either using ClosTron or CRISPR/Cas9. Indeed, the conjugation efficiencies per donor were 28-fold higher when using sExpress compared to CA434 (unpaired t-test; p-value < 0.0001); likewise, conjugation efficiencies per recipient were 10 and 21-fold higher in the ClosTron and CRISPR mutants (respectively) compared to wild type (unpaired t-test; p-value < 0.0001).

The fact that the increase in efficiency of the ClosTron mutant was 55-fold in the previous assay and 10-fold in this one is a consequence of variability between assays, as evidenced by the error bars obtained when counting transconjugants (Figure 3.8 and Figure 3.14). The error bars are consistently bigger for wild type, possibly because environmental variations have a bigger impact on conjugation when the number of transconjugants is low.

Overall, the results presented above appear to confirm that the CLSPO_c06750 gene encodes a type IV restriction system. This system can be avoided by knocking-out the gene or by eliminating Dcm methylation in the donor strain. By using either approach, conjugation efficiencies can be increased by at least 10-fold (comparing to conjugations from *E. coli* CA434 into *C. sporogenes* WT)[112].

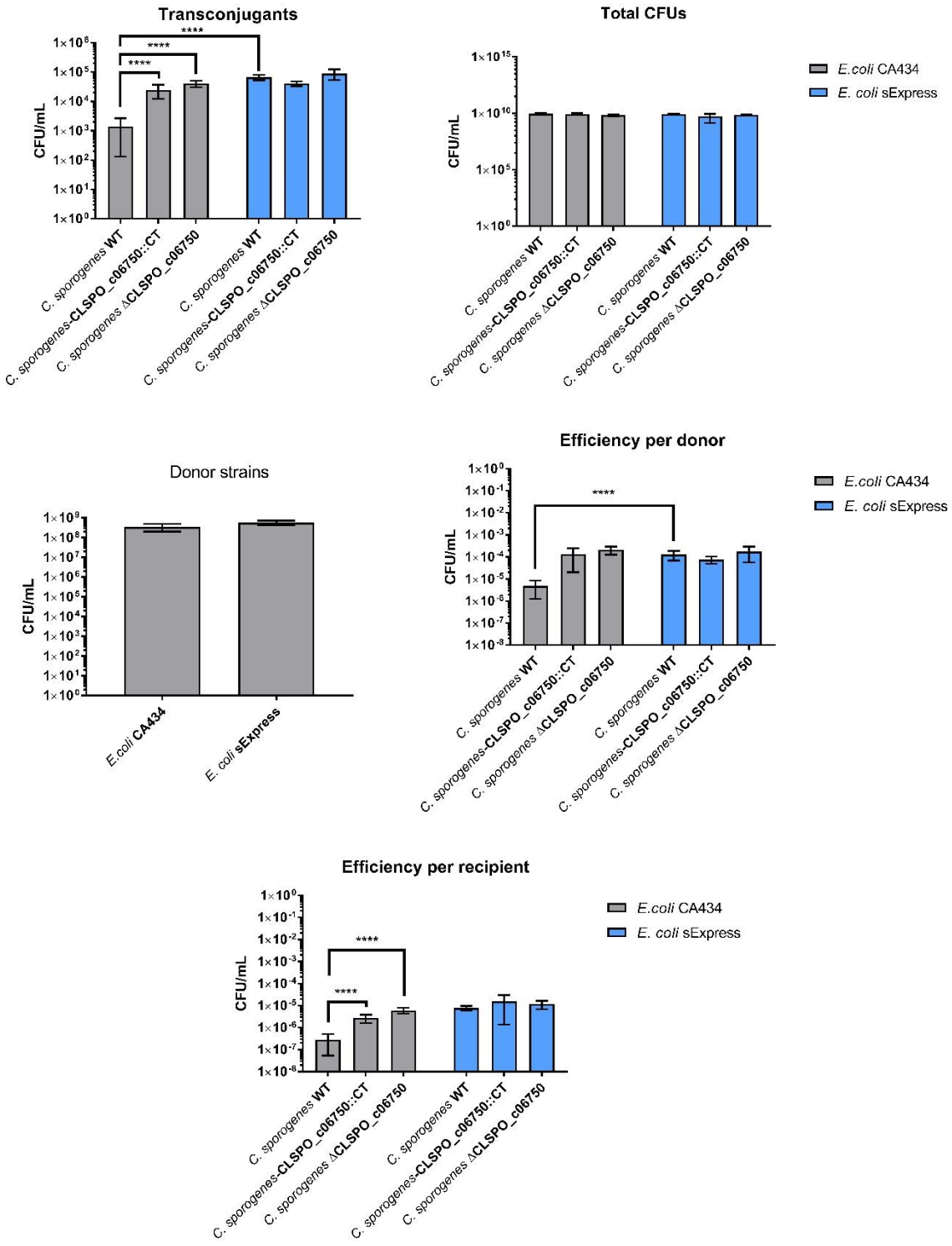


Figure 3.14: Comparison between conjugation efficiencies using *E. coli* CA434 or *E. coli* sExpress as donor strains, and *C. sporogenes* WT, *C. sporogenes*-c06750-1240a::CT or *C. sporogenes* Δc06750 as recipient strains. The number of transconjugants is presented, as well as the number of total CFUs and donor strains. Data presented shows the mean and error bars represent the standard deviation of three biological and three technical replicates. Unpaired t-tests were performed using GraphPad 7 to compare the donor strains (efficiency per donor) or mutants to wild type (efficiency per recipient). Accordingly, **** represent significant differences (p-value < 0.0001).

3.3.2 Overcoming the type I restriction system

While the type IV system contains only one gene, the type I restriction system in *C. sporogenes* appears to include three: CLSPO_c17010, CLSPO_c17020 and CLSPO_c17030. As described above, the gene CLSPO_c17010 appears to encode the modification subunit (MjaXP), CLSPO_c17020 the specificity subunit (PrrB) and CLSPO_c17030 the restriction subunit. These genes form an operon and can be referred to as *hsdM*, *hsdS* and *hsdR*, respectively.

The first step to overcome the type I restriction system was to target the *hsdR* gene using ClosTron, as the HsdR subunit is the only one with restriction activity[95]. Then, when the CRISPR/Cas9 system became available, it was used to create an in-frame deletion of the gene. Finally, HsdS and HsdM were expressed in the conjugative donor strain in order to mimic the methylation pattern of *C. sporogenes* and avoid restriction.

3.3.2.1 Targeting the type I system using ClosTron

The gene encoding the HsdR subunit, CLSPO_c17030, will be referred to as *hsdR*. In order to target this gene using the ClosTron approach, the plasmid pMTL007::*Csp-hsdR*-1537a was obtained from DNA2.0 and used to generate three independent biological replicates of the strain *C. sporogenes- hsdR*1537a::CT. The strategy used to confirm the single insertion of the intron at the target gene was the same used for the type IV restriction system: colony PCR (Figure 3.15 – A), inverse PCR (Figure 3.15 – B) and Sanger sequencing. The inverse PCR produced a strong band (approx. 2.5 Kb) and a very faint band (approx. 5Kb), and Sanger sequencing revealed that both bands correspond to insertions into the *hsdR* gene. As described in Chapter 2, the inverse PCR is done by digesting the genomic DNA of the mutants with *Hin*dIII, ligating the resulting fragments and amplifying them using specific primers. It is possible that the second band was obtained due to inefficient digestion with *Hin*dIII. Overall, these results indicate that the ClosTron intron was inserted into the *hsdR* gene and nowhere else in the genome.

3.3.2.1.1 Sporulation assay

A sporulation assay was conducted and revealed that the mutation of the type I system had no relevant impact on sporulation (Figure 3.16).

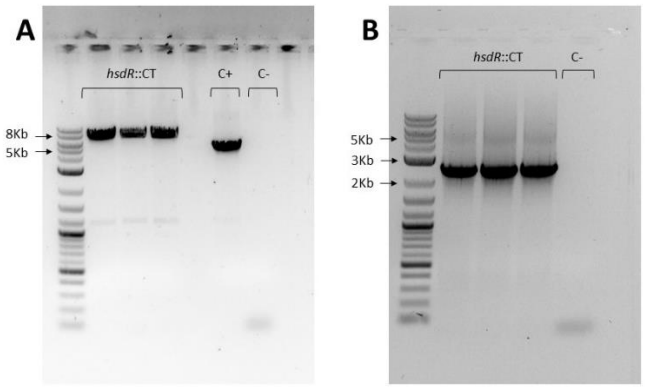


Figure 3.15: A: Gel electrophoresis of colony PCR samples demonstrating the insertion of the ClosTron into *hsdR*. A: The strains were screened using primers hsdR_scr_F and hsdR_scr_R. Wild type was also included (C+), as well as a negative control (water as template; C-). **B:** Gel picture of inverse PCR showing the number of ClosTron insertions in each biological replicates of the strain *C. sporogenes-hsdR*1537a::CT. Primers erm_inv_F and erm_inv_R were used to amplify the DNA. A negative control (water as template; C-) was included.

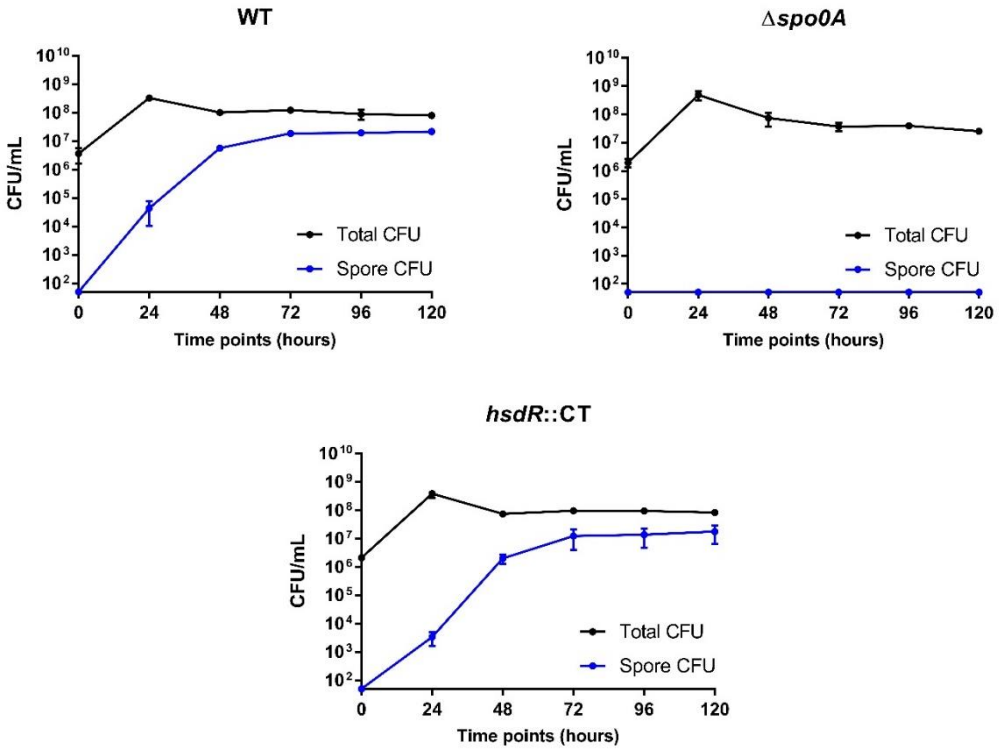


Figure 3.16: Sporulation profile of the *C. sporogenes-hsdR*1537a::CT strain. *C. sporogenes* WT and *C. sporogenes* Δ*spo0A* were included as positive and negative controls, respectively. All the cultures were synchronised at the beginning of the assay, and samples were collected at 24 hour intervals during a total period of 120 hours. For the total CFU count, the cultures were diluted and spotted into TYG agar. For spore CFU count, the cultures were heat-treated at 80°C for 20 minutes prior to the dilutions. Data presented shows the mean and error bars represent the standard deviation of three biological and three technical replicates. The Y axis starts at the limit of detection (50 spores/mL).

3.3.2.1.2 Determination of conjugation efficiency

The mutation of the HsdR subunit of the type I restriction system was expected to have an impact on conjugation efficiency. In order to assess this impact, a conjugation assay was conducted as described above.

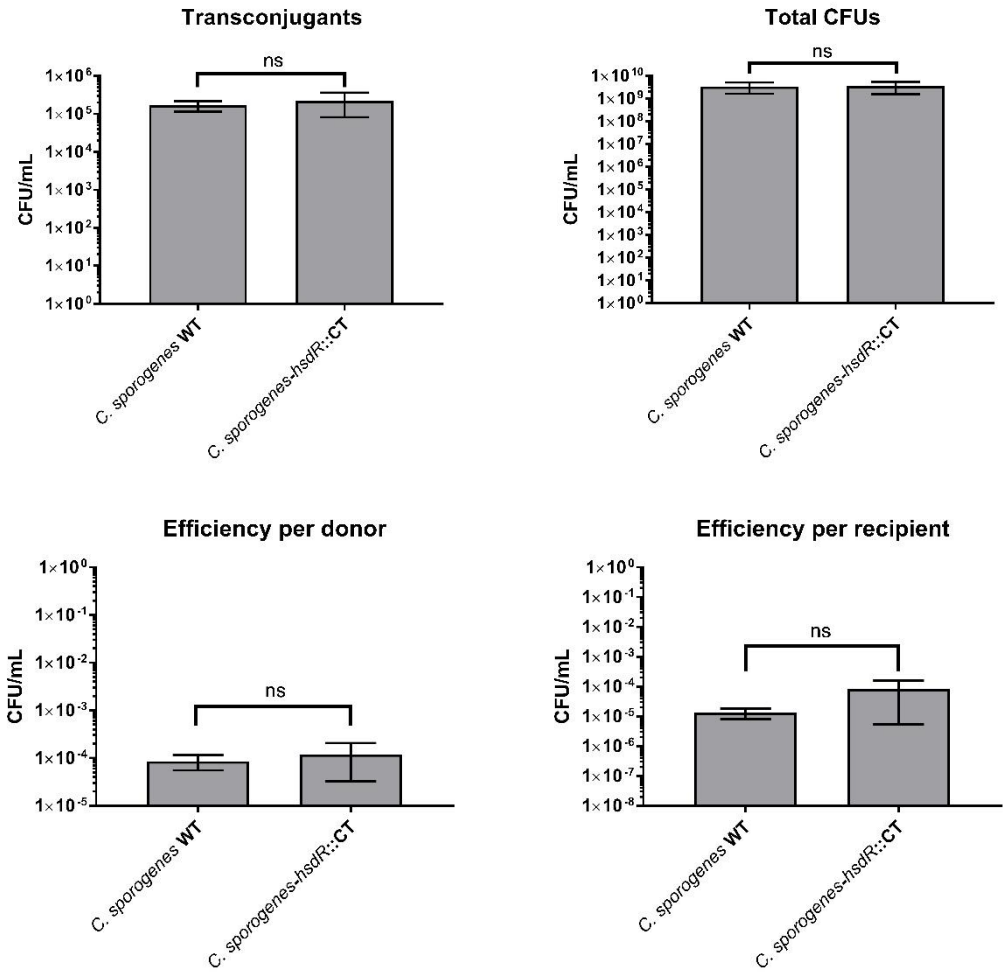


Figure 3.17: Comparison between conjugation efficiencies in *C. sporogenes* WT and *C. sporogenes-hsdR*1537a::CT. The number of transconjugants is presented, as well as the number of total CFUs. Data presented shows the mean and error bars represent the standard deviation of three biological and three technical replicates. Unpaired t-tests were performed using GraphPad 7 to compare the mutant to wild type: ns represent non-significant differences (p-value > 0.05).

The results indicate that there are no significant differences between the conjugation efficiencies of *C. sporogenes* WT and *C. sporogenes- hsdR*1537a::CT (Figure 3.17; unpaired t-test). This was unexpected, as mutations of the HsdR subunit have reduced DNA restriction in other species[96, 97]. Two hypotheses can be proposed to explain these results: i) the insertion of the intron in the gene was not sufficient to affect the restriction activity of the HsdSMR protein complex; or ii) the vector pMTL83151 does not contain the recognition sequence targeted by the same complex. The first hypothesis can be verified by creating an in-frame deletion of the *hsdR* gene or the

whole *hsdM-hsdS-hsdR* operon. On the other hand, the second hypothesis can be tested by identifying the target sequence of the type I system and including it in the pMTL83151 vector. Both approaches were attempted, as described below.

It was also observed that *C. sporogenes-hsdR*1537a::CT strain had a different phenotype than wild type: cells aggregated when growing in broth and buried into the agar when growing on agar plates. This challenging trait makes it difficult to re-streak and pick colonies, which affects the usefulness of the mutants for techniques such as TraDIS. This phenotype can be a result of the mutation itself, or a polar effect from the ClosTron insertion. In-frame deletion of the gene was therefore attempted to help elucidate the true cause of this undesired trait.

3.3.2.2 Targeting the type I system using CRISPR/Cas9

In order to address the problems mentioned above, the *hsdR* gene and the *hsdM-hsdS-hsdR* operon were targeted using the CRISPR/Cas9 system[86]. However, several issues affected this process and the in-frame deletions of the gene and the operon were not obtained.

The first issue was the difficulty in generating the pMTL43151-CLSPO-*hsdR* and pMTL43151-CLSPO-*hsdMSR* vectors. The vectors are modular, and a similar vector was successfully used to target the type IV system. One of the differences is the 20 nucleotide seed sequence of the sgRNA fragment. Hence, four different seed sequences were selected (one targeting the ClosTron intron, as described above, and others targeting the *hsdR* gene). None of these allowed the generation of the vector, not even the seed sequence targeting the ClosTron insertion. The other difference is the gene editing template containing the two homology arms and a small part of the target gene (two first and three last codons). The only portion of the gene editing template common to both pMTL43151-CLSPO-*hsdR* and pMTL43151-CLSPO-*hsdMSR* is the right homology arm (RHA). Consequently, it was hypothesised that the cloning issues might be due to potential read-through from the P_{araE} promoter into the gene editing template, and that the RHA might encode a product toxic to *E. coli*. In order to test this theory, the size of the RHA was reduced from 1 Kb to 500 bp. In addition, the high-copy number Gram-negative replicon ColE1 was replaced by the low-copy number p15a in the modular vector[113]. Finally, only one *E. coli* TOP10 CFU containing the pMTL43121-CLSPO-*hsdR* vector was obtained amongst dozens screened (data not shown). The vector contained a gene editing template with a 500 bp

long RHA, and a sgRNA targeting the *hsdR* gene. The vector pMTL43121-CLSPO-*hsdMSR* was never obtained.

The vector pMTL43121-CLSPO-*hsdR* was transformed into *E. coli* CA434 and then conjugated into *C. sporogenes* WT. Many conjugations were attempted and all screened transconjugants corresponded to wild type. Nevertheless, some colonies appeared to bury into the agar, the same phenotype observed with the ClosTron mutant. In fact, the phenotype appeared to be more intense in these transconjugants, and picking those colonies was nearly impossible. Any attempts to screen those colonies using colony PCR originated bands that had the same size as wild type. Finally, one of the colonies generated two bands in the PCR: one corresponding to wild type and a smaller one, slightly smaller than would be expected for a mutant (Figure 3.18).

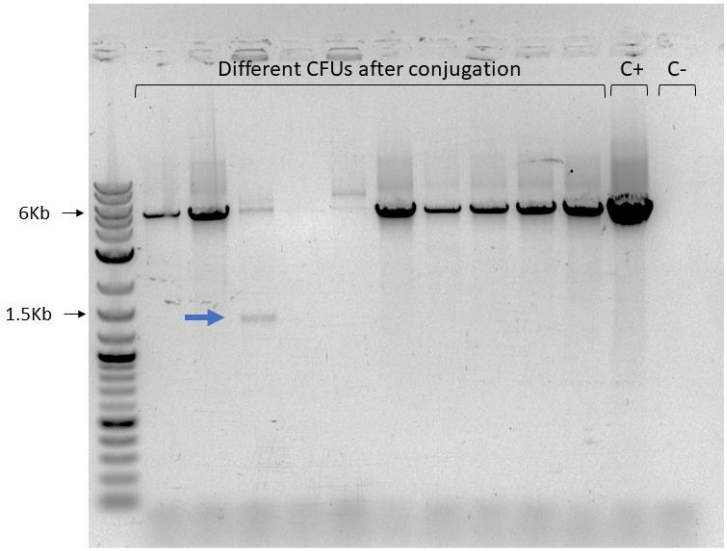


Figure 3.18: Gel electrophoresis of colony PCR samples screening the in-frame deletion of the *hsdR* gene. Different colonies were screened following conjugation of the CRIPR/Cas9 plasmid into *C. sporogenes*, using primers hsdR_scr_F and hsdR_scr_R. The blue arrow shows a band that might correspond to the size expected on deletion of the *hsdR* gene. Wild type was also included (C+), as well as a negative control (water as template; C-).

It is possible that, since the colonies bury into the agar, the only cells that were picked for the colony PCR were actually wild type cells mixed with the mutant. This would explain the two bands visible in Figure 3.18. Several attempts were made to isolate the colonies with the burying phenotype, including i) re-streaking those colonies to single colonies on agar plates; and ii) sub-culturing the colonies in broth, followed by serial dilutions of the culture onto plates to obtain isolated colonies. None of the approaches were successful, as only bands corresponding to wild type or no bands at all were obtained after screening by PCR (data not shown).

The fact that some *C. sporogenes* colonies started burying in the agar after conjugation with the pMTL43121-CLSPO-*hsdR* vector suggest that those CFUs were mutants and that the deletion of the *hsdR* gene is the cause of the phenotype. However, this could not be demonstrated due to the issues described above.

3.3.2.3 Changing the methylation pattern of the DNA before transfer to avoid restriction

As the conjugation efficiency of the *C. sporogenes-hsdR*1537a::CT strain was the same as wild type and the *C. sporogenes* Δ*hsdR* strain was not obtained, a different approach to circumvent the type I restriction system was tested. The approach consisted of changing the methylation pattern of DNA before transfer in order to avoid restriction, which can be achieved by expressing the specificity (HsdS) and modification (HsdM) subunits in the donor strain. Indeed, these subunits are both required for the methylation of DNA, but HsdR is not necessary for this process[95]. An expected issue when using this method is the methylation of the *E. coli* genome by the HsdS-HsdM enzyme complex, which could have toxic effects in this species.

The modular vector pMTL80221 was selected to express HsdS and HsdM, as it shows the following characteristics:

- it does not have a Gram-positive replicon, as the vector will only be used in *E. coli* and not be transferred into *C. sporogenes*;
- it confers resistance to erythromycin, so it can be selected in cells already carrying a vector conferring resistance to thiamphenicol, such as pMTL83151;
- the Gram-negative replicon p15a ensures a low-copy number of the plasmid in the donor cell, which could help prevent HsdS and HsdM toxicity in *E. coli*.

3.3.2.3.1 Testing different promoters to drive the expression of hsdM-hsdS

Two different promoters were selected to drive the expression of the *hsdM-hsdS* operon in *E. coli*:

i) P_{hsd}, the native promoter of the *hsdM-hsdS-hsdR* operon in *C. sporogenes*;

ii) P_{BAD}, an inducible promoter that responds to the presence of L-arabinose.

The use of an inducible promoter provides an additional measure to protect the *E. coli* cells from the potential toxic effects of HsdS and HsdM. The vector containing the

P_{BAD} promoter also included the gene *araC*, required to repress transcription in the absence of L-arabinose.

The expression level obtained with each promoter was assessed by cloning them in front of the gene encoding a red fluorescence protein (RFP). The fluorescence levels in *E. coli* TOP10 were determined using the CLARIOstar plate reader and normalised to the optical density of the cultures (OD_{600}). Three biological and three technical replicates were measured for each promoter. A vector with no promoter driving the expression of RFP (P_{neg}) was included as a negative control. The P_{BAD} promoter and the negative control (P_{neg}) were tested in the absence (0 mM) and presence (100 mM) of the inducer L-arabinose.

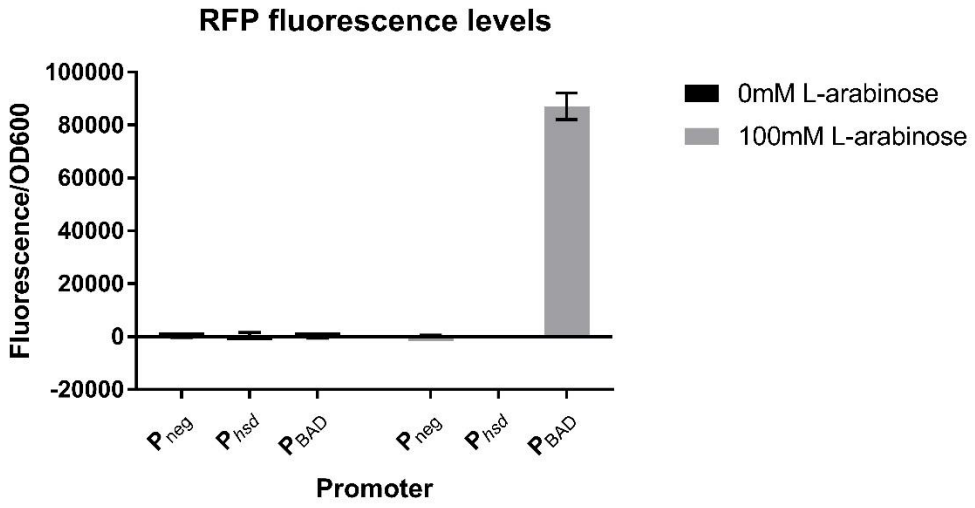


Figure 3.19: Fluorescence levels of RFP driven by different promoters: no promoter (P_{neg}), the native promoter of the *hsdM-hsdS-hsdR* operon in *C. sporogenes* (P_{hsd}) and an inducible promoter (P_{BAD}). The fluorescence levels in *E. coli* TOP10 after 16 hours of growth were determined using the CLARIOstar plate reader and normalised to the OD_{600}. The data presented shows the means and the error bars represent the standard deviation from three biological and three technical replicates.

The results presented in Figure 3.19 show no or very low fluorescence for P_{neg} (in the absence or presence of inducer) or P_{hsd}, indicating that the gene encoding RFP is not being expressed (or expressed at very low levels). The results from P_{neg} indicate that the expression of RFP is not affected by transcriptional read-through. The results from P_{hsd}, on the other hand, suggest that the *E. coli* RNA polymerase does not recognise this promoter. The P_{BAD} promoter appears to be well repressed, as almost no fluorescence was detected in the absence of the inducer. However, repression appears to be alleviated in the presence of 100 mM L-arabinose, resulting in expression of RFP and increase of fluorescence (Figure 3.19).

The inducible promoter P_{BAD} was selected to drive the expression of *hsdS* and *hsdM*, as it is well repressed in the absence of inducer and presents good expression levels in the presence of L-arabinose.

3.3.2.3.2 Determination of conjugation efficiency

Following the selection of the best promoter, the vector pMTL80221-P_{BAD}-*hsdMS* was obtained. The vector includes the P_{BAD} promoter driving the expression of the *hsdM-hsdS* operon, as well as the gene encoding the AraC repressor.

In order to determine the conjugation efficiencies using this approach, two vectors were transformed into the donor *E. coli* sExpress: pMTL80221-P_{BAD}-*hsdSM* and pMTL83151. The vector pMTL80221-P_{BAD}-*hsdSM* should methylate pMTL83151 in *E. coli* sExpress, protecting it from HsdR restriction before its transfer into *C. sporogenes*. The conjugation efficiencies were determined as described above. Cells containing the pMTL80221-P_{BAD}-*hsdSM* vector were grown in the presence of 100 mM L-arabinose to induce the expression of *hsdM-hsdS*.

The results presented in Figure 3.20 indicate that the presence of two vectors affects the growth of the *E. coli* sExpress donor, as suggested by the statistically significant decrease in the number of donor strain CFUs in *E. coli* sExpress carrying pMTL80221-P_{BAD}-*hsdSM* relative to *E. coli* sExpress (unpaired t-test; p-value = 0.0043). The former was also carrying pMTL83151 and was grown in the presence of two antibiotics: thiamphenicol to maintain pMTL83151 and erythromycin to maintain pMTL80221-P_{BAD}-*hsdSM*. The increased toxicity expected in the presence of two antibiotics might have contributed to the lower number of CFUs obtained (Figure 3.20).

It was also observed that the expression of HsdS and HsdM in the donor strain did not affect conjugation efficiency (Figure 3.20). These results, together with the results obtained with the ClosTron mutant (Figure 3.17), suggest that the vector pMTL83151 might not contain the sequence recognised by the type I restriction system.

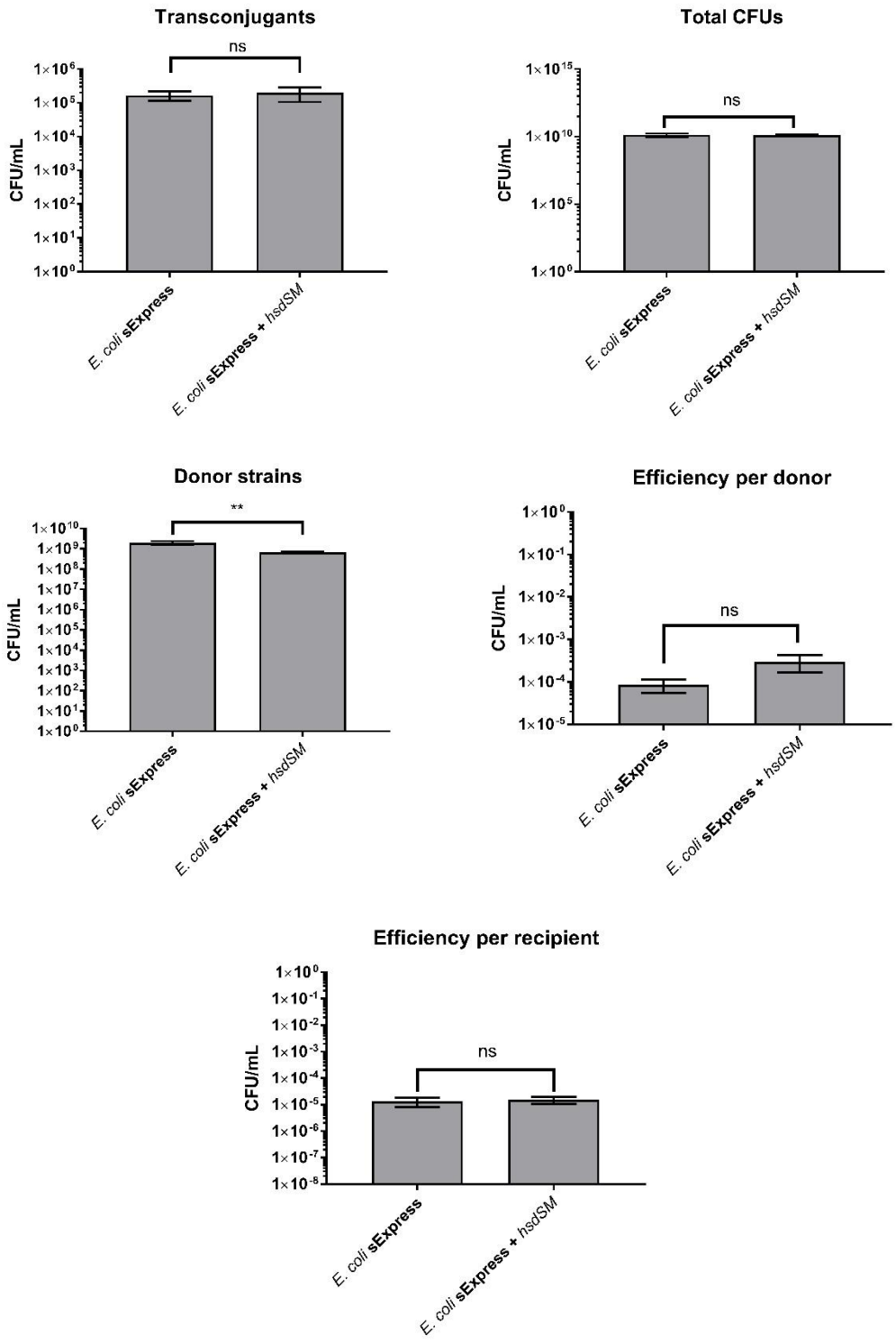


Figure 3.20: Comparison between conjugation efficiencies using *E. coli* sExpress or *E. coli* sExpress carrying pMTL80221-P$_{BAD}$-*hsdSM* as donor strains. The number of transconjugants is presented, as well as the number of total CFUs and donor strains. Data presented shows the mean and error bars represent the standard deviation of three biological and three technical replicates. Unpaired t-tests were performed using GraphPad 7 to compare both donor strains: ns represent non-significant differences (p-value > 0.05) and ** represent significant differences (p-value < 0.005).

3.3.2.4 Identification of potential type I recognition sites

As mentioned in the introduction to this chapter, the recognition sequence of the type I restriction system found in *C. sporogenes* is not known. The vector pMTL83151 was used in the conjugation efficiency assays, but the results indicate that this vector does not contain that recognition sequence and, consequently, was not targeted by the type I system. As a last attempt to determine the impact of the type I restriction system on gene transfer, some potential recognition sequences were determined *in silico*.

In the SBRC group, it was found that the pathogen *C. botulinum* ATCC 19397 has a type I system that recognises the sequence 5'-CTANNNNNNRTGAA-3' [101] (Daphne Groothuis, personal communation). The HsdS subunit from *C. botulinum* ATCC 19397 presents 19.73% identity to HsdS from *C. sporogenes* (blastp; e-value = $1e^{-5}$) and its recognition sequence was found twice in the *trcas9* gene used in the CRISPR/Cas9 system. Moreover, the specificity subunit from *C. sporogenes* is 100% identical to a specificity subunit from *C. botulinum* MFBjulcb1 (blastp; e-value = 0), and this subunit appears to recognise the sequence 5'-AGTNNNNNNRTGC-3' [101].

Therefore, three potential recognition sites were designed in order to be added to the pMTL83151 vector:

i) Recognition site 1 (RS1): 5'-CTAAATACGATGAA-3'
ii) Recognition site 2 (RS2): 5'-CTAGTCATTATGAA-3'
iii) Recognition site 3 (RS3): 5'-AGTTATATAATGC-3'

Part of the *trcas9* gene containing the recognition sequences 1 and 2 was added to the pMTL83151 vector, resulting in pMTL83151_RS1/2. These sequences were tested despite the low homology between the two proteins because the CRISPR/Cas9 system is routinely used to create in-frame deletions of genes. If these sequences are targeted in *C. sporogenes*, then the method can be improved by avoiding restriction and increasing conjugation efficiency. Likewise, the recognition sequence 3 was added to pMTL83151, resulting in pMTL83151_RS3.

3.3.2.5 A final attempt

A conjugation assay was performed as described above. The vectors pMTL83151, pMTL83151_RS1/2 and pMTL83151_RS3 were selected to be conjugated from the donor *E. coli* sExpress into the recipient *C. sporogenes* WT.

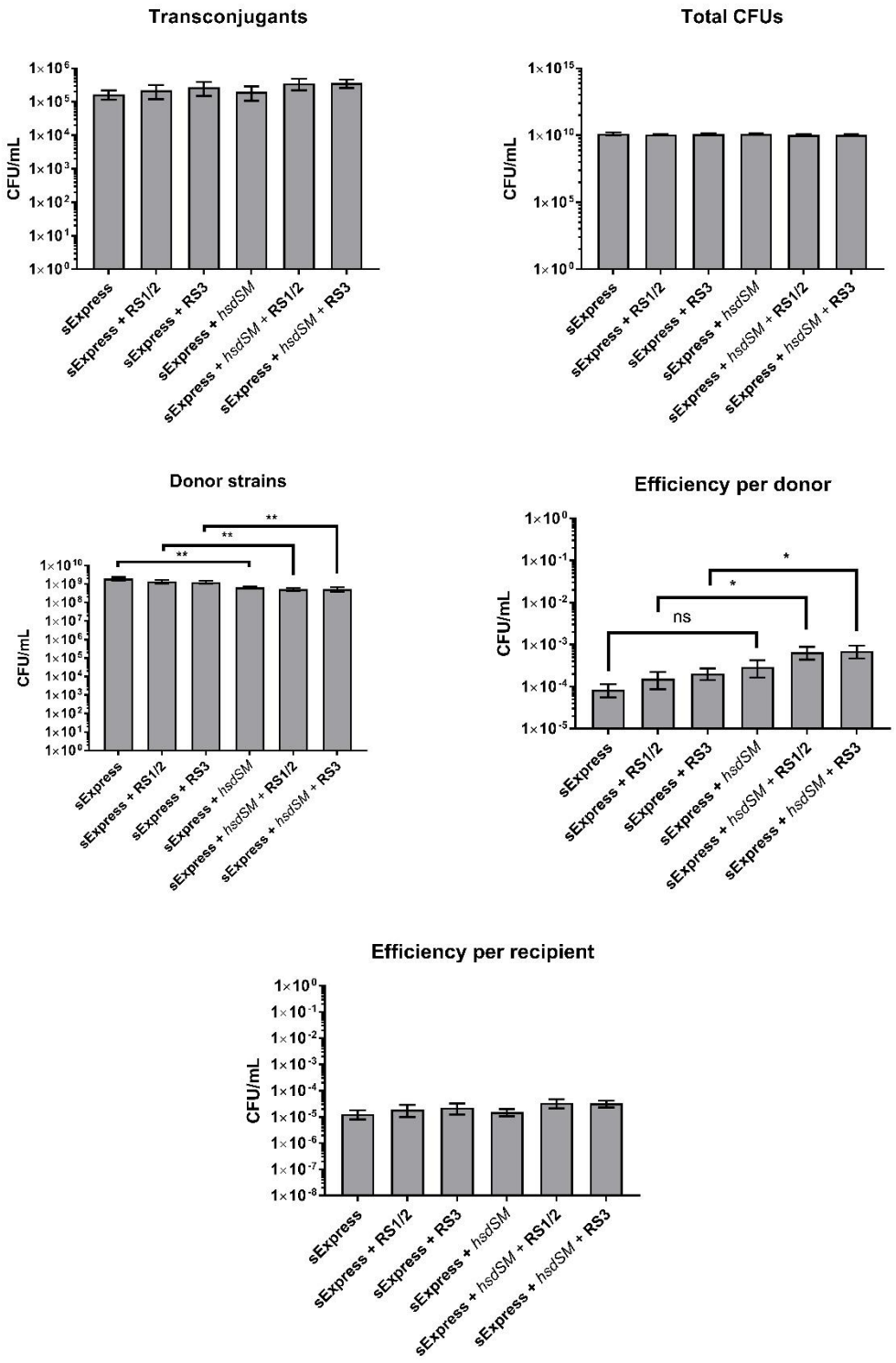


Figure 3.21: Comparison between conjugation efficiencies using *E. coli* sExpress or *E. coli* sExpress carrying pMTL80221-PBAD-*hsdSM* as donor strains. Different vectors were conjugated from each *E. coli* donor into *C. sporogenes* WT: pMTL83151 (no label), pMTL83151_RS1/2 (RS1/2) and pMTL83151_RS3 (RS3). The number of transconjugants is presented, as well as the number of total CFUs and donor strains. Data presented shows the mean and error bars represent the standard deviation of three biological and three technical replicates. Unpaired t-tests were performed using GraphPad 7 to compare both donor strains: ns represent non-significant differences (p-value > 0.05); * (p-value < 0.05) and ** (p-value < 0.005) represent significant differences.

In order to test if the methylation pattern had any impact on type I restriction, the *E. coli* sExpress strain carrying pMTL80221-P_{BAD}-*hsdSM* was also used as donor. The *C. sporogenes-hsdR*1537a::CT strain was not tested because the fact that it buries into the agar compromises its use in techniques such as TraDIS.

The results presented in Figure 3.21 indicate no significant difference between the different donor strains and transferred vectors, when considering the number of transconjugants, total CFUs and efficiency per recipient (One-way ANOVA performed using GraphPad 7). Similar to what was observed above, the presence of two vectors and consequent use of two antibiotics had a significant negative impact on the growth of the donor strains (unpaired t-test; p-value < 0.005). Nevertheless, the number of transconjugants obtained was similar, so the conjugation efficiency per donor was higher when using the vectors containing the recognition sequences in combination with the HsdSM-carrying donor (unpaired t-test; p-value < 0.05).

Despite the apparent increase in conjugation efficiency per donor, the results indicate that the recognition sequences are not targeted by the type I restriction system in *C. sporogenes*. If they were, a decrease in conjugation efficiency relative to pMTL83151 would be expected, and that was not observed (Figure 3.21). It appears that the slight decrease in the number of donor CFUs is not sufficient to negatively affect the number of transconjugants obtained, and that is possibly the reason why conjugation efficiency per donor was higher.

3.4 Discussion

The spore former *C. sporogenes* has two restriction systems that might act as a barrier to gene transfer, and the work described above allowed the increase of transfer frequency by escaping one of those systems. The methods normally used to transfer DNA into *C. sporogenes* are transformation and conjugation. It has been shown above that higher efficiencies are obtained with conjugation than with transformation, so the former was selected as the favoured method for DNA transfer. It should be noted that transformations were performed using freshly prepared *C. sporogenes* competent cells, and that transformation efficiency can be even lower if stored competent cells are used (Dr Aleksandra Kubiak, personal communication).

In addition, the results herein presented appear to confirm that the product of the gene CLSPO_c06750 behaves as a type IV restriction enzyme that cleaves DNA containing methylated cytosine residues. This has been demonstrated by the increase in conjugation efficiency that was observed when the gene was mutated, either by disruption (ClosTron) or by in-frame deletion (CRISPR/Cas9). Furthermore, the use of a donor strain that lacks the Dcm protein (*E. coli* sExpress) resulted in a similar increase in conjugation efficiency, possibly due to the absence of methylation in cytosine nucleotides. The fact that no cumulative effect was observed when DNA was conjugated from sExpress into the CLSPO_c06750 mutants indicates that both strategies increase conjugation efficiency by avoiding the type IV restriction system[112]. Although both approaches have the same result, sExpress is preferred as it allows the use of wild type recipient strain (the mutation of the CLSPO_c06750 gene appears to affect sporulation, which is not desirable).

Regarding the type I restriction system, the *hsdR* ClosTron mutant did not present a significant increase in conjugation efficiency. The same result was observed when the specificity and modification subunits (HsdS and HsdM, respectively) were expressed in the donor strain in order to mimic the methylation pattern of the recipient. The most likely explanation is that the vector used to test conjugation efficiency did not contain the specific sequence targeted by this restriction system. An attempt to identify that specific sequence was unsuccessful, but further research might allow the identification of new potential targets. An alternative explanation to the lack of increase in conjugation efficiency is that the type I system is not functional in *C. sporogenes*, but further studies would be required to explore this hypothesis.

The mutation of the *hsdR* gene appeared to cause the cells to bury in the agar. The reason for this is unknown, but the phenotype has been observed in other mutants (Dr Inés Cañadas Blasco, personal communication). The use of different media or agar concentrations might help to alleviate this phenotype and allow the manipulation of the cells. However, the expression of HsdS and HsdM in the donor cells appears to be a better alternative because it allows the use of wild type strains.

Finally, it is worth mentioning that the seed sequence used to target the ClosTron intron and create an in-frame deletion of the CLSPO_c06750 gene can be a valuable tool to create in-frame deletions in any ClosTron mutant. This can be particularly relevant in situations where the target gene lacks efficient and/or specific sgRNA targets.

Overall, the results obtained in this chapter indicate that *E. coli* sExpress might be a good candidate to replace *E. coli* CA434 as a conjugation donor. Both strains will be compared again in Chapter 4 to determine if the DNA transfer frequencies are enough to allow the use of high-throughput techniques such as TraDIS.

Chapter 4: Identification of genes essential for sporulation or germination in *Clostridium sporogenes*

4

4.1 Background

The identification of genes that are essential for sporulation or germination is an important step towards understanding these complex developmental processes. In addition, their identification can play a major role in the optimisation of *C. sporogenes* as a delivery vehicle for tumour therapy[30] (CDEPT, described in Chapter 1). One of the concerns when using this therapy is that, following colonisation of a patient's tumour, the genetically modified strain might be released to the external environment. Should the GMO remain as vegetative cells, then in the unlikely event that release does occur the oxygen sensitive cells would be killed upon exposure to air. Spores, in contrast, would be resistant and could, therefore, mediate dissemination of the GMO. One solution to counter this risk would be to create a *C. sporogenes* strain which only produces spores under defined conditions – a conditionally sporulating strain. This would allow the generation of high spore titres under laboratory conditions, but once germinated in the tumour the clostridia would remain as vegetative cells and be unable to sporulate again. Such a sophisticated form of disablement will ensure that dissemination of clostridia is avoided, thereby satisfying any concerns of regulatory agencies. The required conditionally sporulating strain could be generated by placing a gene essential for sporulation under the control of an inducible promoter. Then, the gene would only be expressed in the presence of an inducer and, consequently, the ability of the cells to sporulate would be dependent on that inducer.

The aim of the work presented in this chapter is to identify genes essential for survival, vegetative growth, sporulation and germination in *C. sporogenes*. TraDIS was the approach selected to achieve this goal, and is described below.

4.1.1 TraDIS: a high-throughput approach to identify essential genes

TraDIS (Transposon Directed Insertion-site Sequencing) is a valuable tool that can be used to identify genes that are essential for survival. This approach involves the creation of a library of mutants generated by random transposon insertion. The library needs to be big enough (usually one million mutants) to saturate the whole genome, so that genes with no insertions can be identified as essential for survival[92, 93]. Once the

library is created (mutants are obtained and pooled), it has to be grown under the selective conditions that are being tested. Then, genomic DNA is extracted and prepared for Illumina sequencing. The sequencing primers align to the transposon, so the reads obtained during sequencing allow the identification of the site in the genome where the transposon was inserted. Finally, the analysis of the sequencing results indicates which genes are essential for survival under the tested conditions, by identifying genes with few or no reads mapped (the insertion index is determined taking into account the number of reads that map to the gene and the size of the gene)[93, 114] (Figure 4.1).

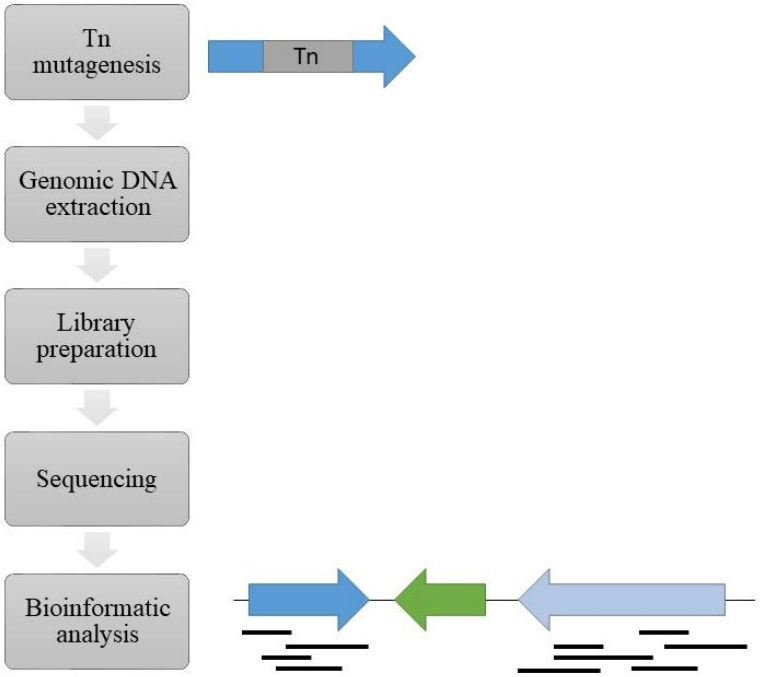


Figure 4.1: Schematic representation of the steps required for TraDIS. A mutant library is created by random insertions of transposons (Tn) into genes (blue) and grown under conditions to select for a specific phenotype, *e.g.*, ability to sporulate. The genomic DNA of the whole library is extracted, prepared for sequencing and then sequenced using Illumina technology. Bioinformatic analysis allows the identification of genes essential for the process being investigated (green), as these do not have reads (black lines) mapped to them. Adapted from van Opijnen *et al.*[93].

4.1.1.1 Use of tetracycline inducible promoters to regulate the expression of relevant genes

One of the problems associated with the creation of transposon libraries is the occurrence of early transposition events. This can be problematic if the transposon is inserted elsewhere in the vector: if the transposon disrupts the transposase, for instance, the vector will no longer be suitable for the creation of transposon libraries.

Furthermore, early transposition events can lead to a "dominance" of certain mutants, which would be a limitation for TraDIS, where the whole genome should be covered to assess gene essentiality. These issues can be prevented by the use of an inducible system to control the expression of the transposase. Using this approach, transfer of the vector into the cells and occurrence of transposition events are sequential and not simultaneous, resulting in a random transposon mutant library.

The transposon mutant library must be stable, which means that the transposon must be inserted in a single gene and stay in that gene. This can only happen if the plasmid is lost following transposition – without the transposase, the transposon will no longer be mobile and the mutant will be stable. Moreover, the presence of high amounts of plasmid during sequencing might lead to most reads aligning to the plasmid rather than the genome, affecting the outcome of the experiment.

The tetracycline inducible system (Tet system) allows the control of the transposase and the induction of plasmid loss simultaneously. This system relies on a resistance mechanism originally found in the *E. coli* Tn*10* transposon against the antibiotic tetracycline (Tc)[115]. Tetracycline binds to the small ribosomal subunit of the cells and impedes its interaction with aminoacyl-tRNA, inhibiting protein synthesis[115]. In order to avoid this, *E. coli* transports the antibiotic out of the cell in a process mediated by the membrane-associated protein TetA. However, TetA also transports other cations and affects the membrane potential, which can be lethal to the cell[116]. Therefore, the expression of *tetA* is tightly controlled by the presence of a repressor, TetR. The promoters that drive the expression of *tetA* and *tetR* are divergent and contain two palindromic operators (tetO1 and tetO2)[117]. In the absence of Tc, TetR binds to the operator sequence and blocks transcription. When Tc is present, it binds to TetR (in association with Mg^{2+}) and causes a conformational change in the DNA-binding domain that decreases the affinity of the repressor to the operator sequence, allowing transcription of *tetA* and *tetR*[117] (Figure 4.2). This system can also be induced by anhydrotetracycline (aTc), a non-antibiotic analogue[118].

The anhydrotetracycline inducible system has been previously used in TraDIS experiments[114]. The authors used anhydrotetracycline not only to control the expression of the transposase, but also to induce the loss of the plasmid used to create the library. This was achieved by placing the origin of replication downstream of the promoter driving the expression of *tetR*. Since there was no terminator between *tetR*

and the replicon, transcriptional read-through allowed the induction of plasmid instability by the presence of aTc[114].

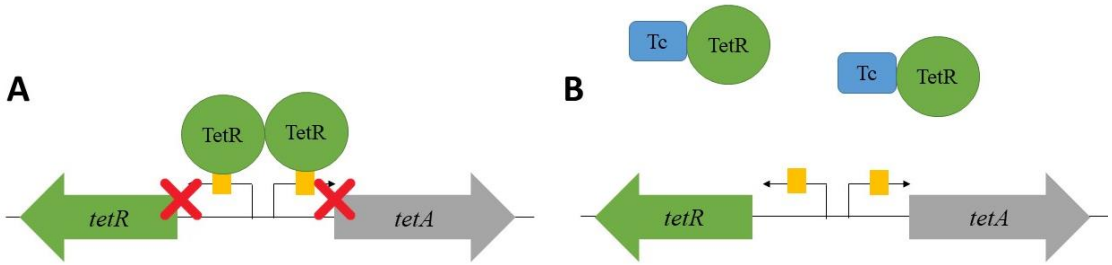


Figure 4.2: Schematic representation of the tetracycline inducible system. A: in the absence of tetracycline, the repressor TetR prevents transcription of *tetR* and *tetA*. **B:** when tetracycle (Tc) is present it binds to TetR releasing the operators (yellow) and allowing transcription of *tetR* and *tetA*.

4.2 Chapter overview

The main goal of the work presented in this chapter was to identify genes essential for sporulation and germination in *C. sporogenes*. The strategy used to achieve this goal was similar to the one previously used in *C. difficile*[114].

The first step was to determine the best plasmid to mediate the insertion of the transposon into the *C. sporogenes* genome. Different configurations of tetracycline inducible promoters were tested to determine which one provided better repression in the absence of anhydrotetracycline and/or better expression in the presence of the inducer. The results allowed the selection of the best promoters to regulate the expression of the transposase and to trigger plasmid loss.

The creation of a transposon mutant library required increased conjugation efficiencies to allow the transfer of the selected plasmid into *C. sporogenes*. The conjugative donors *E. coli* CA434 and *E. coli* sExpress (selected as the best candidate in Chapter 3) were both tested to determine the number of conjugations required to obtain one million mutants. Then, as the identification of genes essential for sporulation/germination required the isolation of pure spores, two spore purification methods were tested and compared.

Once all the optimisation steps were completed, the transposon mutant library was obtained and allowed the identification of genes essential for the survival of *C. sporogenes*. The mutants were submitted to starvation conditions to induce sporulation, and then the spores were used to inoculate nutrient-rich broth that triggered germination. These conditions allowed the identification of genes essential for sporulation and germination.

4.3 Results

4.3.1 Selection of the best plasmid to mediate transposon mutagenesis

The creation of a transposon mutant library required the random insertion of a transposon into the genome of *C. sporogenes*. The machinery required to mediate the transposition events was delivered on a plasmid, and the characteristics of that plasmid had to be carefully selected. First, the plasmid had to contain a transposon. That transposon was composed of a selection marker (*catP*, conferring resistance to thiamphenicol) flanked by two inverted terminal repeat sequences (ITRs). The ITRs are the sequences recognised by the transposase, and *catP* allowed the selection of cells where the transposon was inserted in the genome. A transposase was also required to mediate the mobility of the transposon. The *Mariner/Himar1* transposase was selected because it mediates integration into TA sites, which is advantageous for random mutagenesis in clostridia (TA rich genomes)[119]. The plasmid also contained another marker that allowed the selection of the plasmid itself (*ermB*, conferring resistance to erythromycin), a high-copy Gram-negative replicon (ColE1) and a stable Gram-positive replicon (pCB102).

A plasmid with the characteristics described above, pMTL-YZ14, has been previously used to mediate transposon mutagenesis in *C. sporogenes*[119]. In this plasmid, the transposase is under the control of the P_{tcdB} promoter from *C. difficile*, which only allows transcription in the presence of the *C. difficile* sigma factor TcdR. Consequently, the *C. sporogenes* strain used by Zhang *et al.*[119] had to be genetically modified to express *tcdR* in order to allow transposon mutagenesis using pMTL-YZ14. In the present study, the use of the wild type *C. sporogenes* strain was preferred, so the use of tetracycline inducible promoters was explored as an alternative to drive the expression of the transposase. The other advantage of this system is its ability to control both the expression of the transposase and the loss of the plasmid[114].

4.3.1.1 Optimisation of expression of transposase and induction of plasmid loss

The tetracycline inducible system used by Dembek *et al.*[114] consisted of two divergent promoters, each containing the operator sequences recognised by TetR. One of the

promoters controls the expression of the transposase, while the other controls the expression of *tetR*. The second promoter also controls plasmid loss due to transcriptional read-through into the Gram-positive replicon (Figure 4.3).

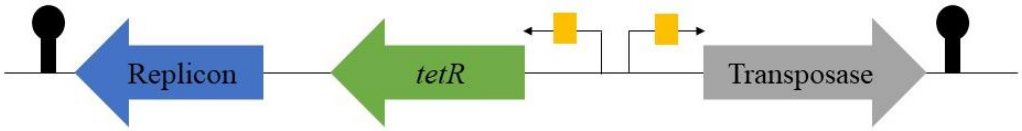


Figure 4.3: Schematic representation of the inducible system used by Dembek *et al.*[114]and in the present study. The two promoters in this system are divergent and contain operator sequences recognised by TetR (yellow). One of the promoters regulates the expression of the transposase required to mediate the mobility of the transposon. The other promoter controls the expression of the repressor TetR, as well as plasmid loss.

Different promoters can be used to control the expression of the transposase and *tetR*, provided they contain the operator sequences recognised by TetR. At the beginning of this work, some strains carrying vectors containing tetracycline inducible promoters were available in the SBRC culture collection. Those constructs had the inducible promoters driving the expression of *catP*, a system that can be used as a reporter to assess gene expression levels through an enzymatic assay of chloramphenicol acetyltransferase.

One of the vectors available in the culture collection (pMTL-tet3nO) was designed by Dr David Walker as part of his PhD thesis[120]. The pMTL-tet3nO construct uses the ferredoxin promoter (P_{fdx}) to control the expression of *catP* and the thiolase promoter (P_{thl}) to control the expression of *tetR*. Both promoters were obtained from *C. sporogenes*. An operator (*tetO*1) was placed in the P_{fdx} promoter to make it inducible. Furthermore, a 46 bp region containing a second ribosome binding site (RBS) was placed upstream of the target gene. In this construct, the *tetR* gene was codon-optimised for *C. difficile* (Figure 4.4).

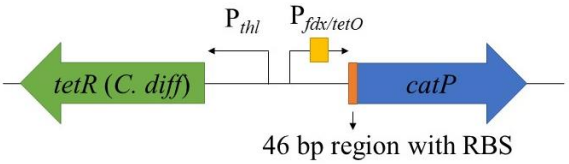


Figure 4.4: Schematic representation of the Tet inducible system from pMTL-tet3nO. The P_{thl} promoter drives the expression of a *tetR* gene that has been codon optimised for *C. difficile*. The P_{fdx} promoter contains the operator *tetO* that makes it inducible (yellow), and it drives the expression of the *catP* gene. An additional 46 bp region with a second ribosome binding site (RBS) is present between $P_{fdx/tetO}$ and *catP*.

The other construct available in the SBRC culture collection (pMTL-tet3nO_RFtetR) was obtained by Ed Farries and is based on the construct used to create the *C. difficile* transposon mutant library[114]. In this vector, the xylose promoter (P_{xylA}) from *B. subtilis* controls the expression of the target gene. This promoter does not have an RBS, but the construct includes the same 46 bp region found in pMTL-tet3nO upstream of the target gene, and this region contains the RBS needed for the translation of the gene. The native *tetR* gene from *Staphylococcus aureus* was used in this construct, and its expression was driven by its native promoter (only two base pairs were changed in the -10 box to match the consensus sequence from *B. subtilis*)[118]. Both promoters contain *tetO* operators, so both of them can be induced (Figure 4.5).

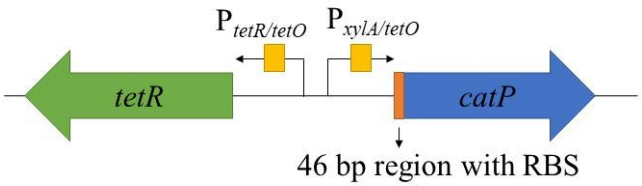


Figure 4.5: Schematic representation of the Tet inducible system from pMTL-tet3nO_RFtetR. The $P_{tetR/tetO}$ inducible promoter drives the expression of the *tetR* gene. The $P_{xylA/tetO}$ inducible promoter drives the expression of the *catP* gene. An additional 46 bp region with a ribosome binding site (RBS) is present between $P_{xylA/tetO}$ and *catP*.

The 46 bp region found upstream of the target gene is necessary in the pMTL-tet3nO_RFtetR vector, due to the lack of an RBS in the promoter. However, in the pMTL-tet3nO construct, the presence of two RBS sequences so close together (only 40 bp between the sequences) might interfere with translation. Hence, a new version of pMTL-tet3nO was created for this study: the 46 bp region was eliminated as described in Chapter 2. This new version was named pMTL-tet3nO_RBS.

4.3.1.2 Enzymatic assay of chloramphenicol acetyltransferase

The ideal inducible system should be well repressed in the absence of an inducer and provide good expression levels when the inducer is present. This can be optimised by using different promoters and/or operators driving the expression of the target gene. In order to assess the best promoter/operator combination, an enzymatic assay of chloramphenicol acetyltransferase (CAT assay) was conducted as described in Chapter 2. The assay tested the expression levels obtained with pMTL-tet3nO (P_{thl} driving *tetR* and $P_{fdx/tetO}$ driving *catP*), pMTL-tet3nO_RBS (same promoters, but only one RBS in front of *catP*) and pMTL-tet3nO_RFtetR ($P_{tetR/tetO}$ driving *tetR* and $P_{xylA/tetO}$ driving *catP*). A promoterless construct was included as a negative control, and a

construct with P_{fdx} driving the expression of *catP* was included as a positive control. Another construct was available that had *catP* under the control of a lactose inducible system, and this vector was also included for comparison. Three biological replicates of *C. sporogenes* strains carrying each plasmid were grown for eight hours. For each construct, two concentrations of anhydrotetracycline were added after four hours of growth: 100 ng/mL and 200 ng/mL. The strain carrying the lactose inducible system was induced with 10 mM lactose also after four hours of growth.

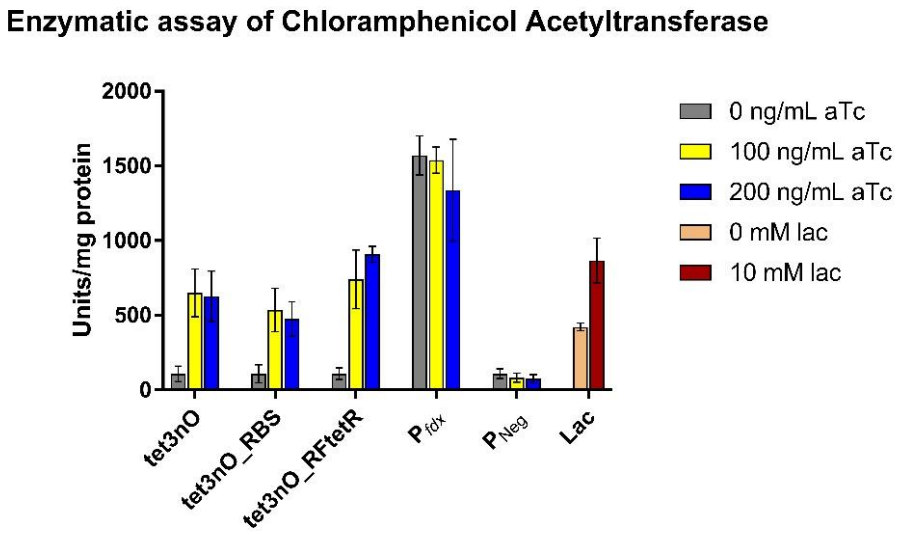


Figure 4.6: CAT assay comparing enzymatic activity resulting from expression of *catP*. *C. sporogenes* cells carrying the different plasmids were grown for eight hours in TYG medium, and the inducers were added after four hours of growth. All cultures were induced with 100 ng/mL or 200 ng/mL anhydrotetracycline, except the cells carrying the lactose inducible promoter, which were induced with 10 mM lactose. Data shows the means and error bars represent standard deviations of three biological replicates.

The results presented in Figure 4.6 show that the enzymatic activity is high in cells carrying constructs based on pMTL-tet3nO that have been induced with anhydrotetracycline. This indicates that the expression levels obtained with those promoters are high when induced. Moreover, the low levels of activity in the absence of the inducer indicate that the system is tightly regulated. Although the expression levels obtained with the tetracycline inducible promoters and the lactose inducible promoter appear to be similar, the tetracycline system is better repressed. The results also suggest that the tetracycline systems are not dose-dependent, and that the presence of the 46 bp region containing the second RBS does not affect translation in pMTL_tet3nO.

Overall, the tetracycline inducible promoters appeared to be good candidates to control the expression of the transposase in the TraDIS experiments. The system used in

pMTL-tet3nO_RF showed the highest expression levels after induction. Moreover, the presence of the second *tetO* operator in the promoter controlling expression of *tetR* allows the induction of plasmid loss due to transcriptional read-through into the Gram-positive replicon. Hence, the $P_{xylA/tetO}$ promoter was selected to drive the expression of the transposase and $P_{tetR/tetO}$ was selected to control the expression of *tetR* and plasmid loss.

4.3.1.3 Description of the plasmid used to generate the transposon mutant library

The vector used in the present study to generate the transposon mutant library was constructed by Dr Craig Woods. The vector, named pMTL-CW21 (Figure 4.7), has $P_{xylA/tetO}$ driving the expression of the *Mariner/Himar1* transposase and $P_{tetR/tetO}$ driving the expression of *tetR* and reading into the Gram-positive replicon pCB102 - as described in section 4.3.1.1. The *catP*-based transposon is flanked by ITRs and by I-*Sce*I recognition sequences. The I-*Sce*I sites are present in the vector but are not transferred to the genome, so they allow further elimination of any vector remaining after plasmid loss by digestion with the restriction enzyme I-*Sce*I. The vector also contains the ColE1 Gram-negative replicon and confers resistance to erythromycin (*ermB*).

Another trait of the pMTL-CW21 vector is the presence of a strong promoter (P_{thl}) on one side of the mini-transposon and the *codA* gene on the other side. These are part of a mechanism to prevent transposition in the *E. coli* donor strains: when the mini-transposon is present, the *fdx* terminator inside the transposon prevents transcription of the *codA* gene by P_{thl}; when the transposon "jumps" to another position, transcription of the *codA* gene is enabled. The product of the *codA* gene is toxic in the presence of 5-fluorocytosine[121], thus the system present in pMTL-CW21 leads to the elimination of any *E. coli* cells where transposition occurs.

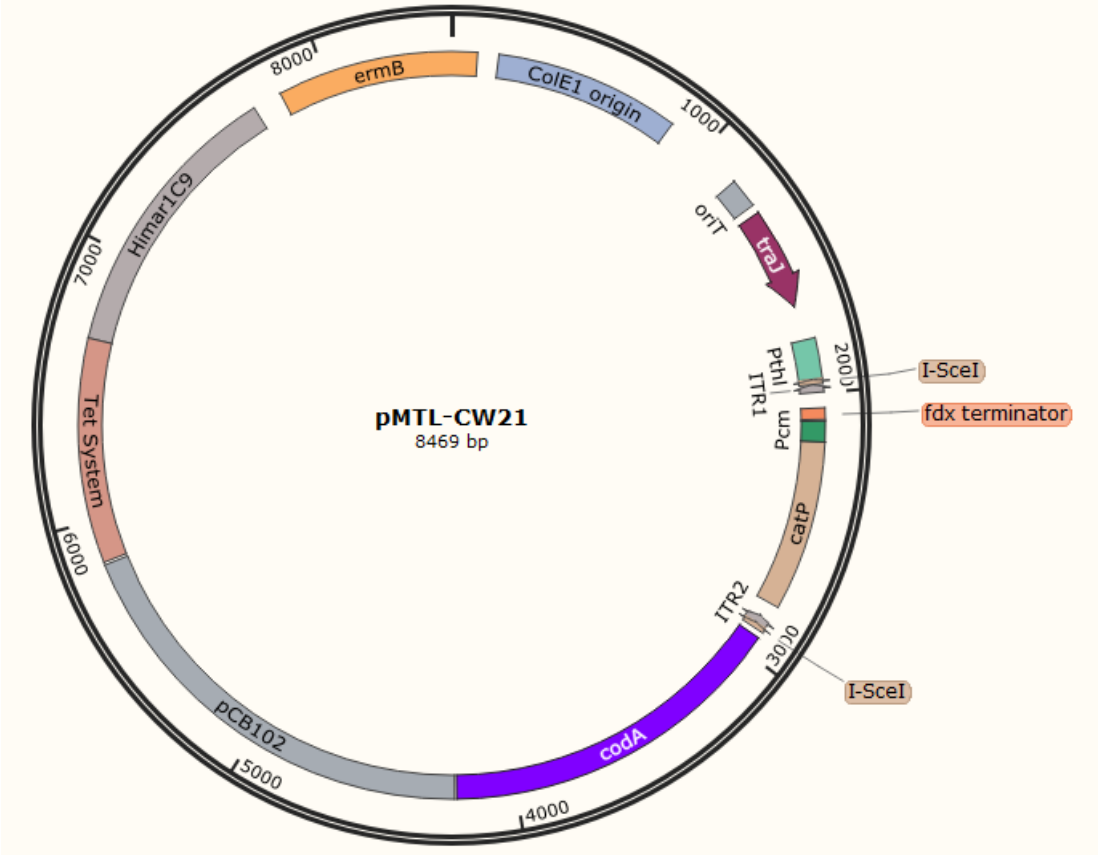


Figure 4.7: pMTL-CW21 vector used for the generation of the transposon mutant library.

4.3.2 Comparison between two donor strains and determination of the number of conjugations required to obtain one million mutants

In Chapter 3, it was demonstrated that the use of the *E. coli* sExpress strain as a conjugative donor leads to increased conjugation efficiencies relative to *E. coli* CA434. It was hypothesised that this increase was due to the avoidance of the type IV restriction system from *C. sporogenes*. Both donor strains have been tested with the modular vector pMTL83151 in Chapter 3. However, after the pMTL-CW21 vector was selected to mediate transposon mutagenesis, the assays to determine conjugation efficiency were repeated. Accordingly, the pMTL-CW21 vector was electroporated into *E. coli* CA434 and *E. coli* sExpress donors and then conjugated into *C. sporogenes* WT. After conjugation, *C. sporogenes* cells were plated onto TYG medium with or without erythromycin selection to count transconjugants and total CFUs, respectively. The donor strains were plated onto LB medium with erythromycin selection. All conjugations were done with three biological and three technical replicates.

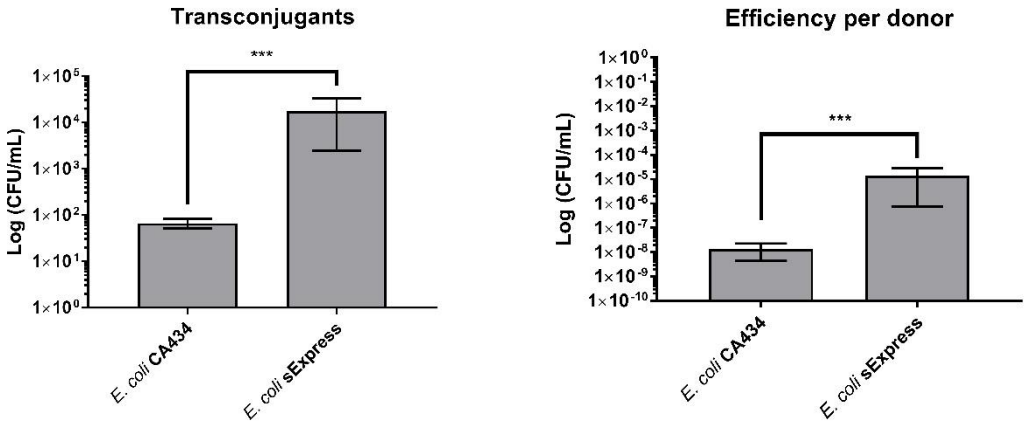


Figure 4.8: Comparison between conjugation efficiencies using *E. coli* CA434 or *E. coli* sExpress as donor strains to transfer pMTL-CW21 into *C. sporogenes* WT. The number of transconjugants is also presented. Data show the means and error bars represent the standard deviation of three biological and three technical replicates. Unpaired t-tests were performed on log-transformed data using GraphPad 7: *** represent significant differences (p-value < 0.005).

The results confirmed that sExpress was a better donor strain than CA434 (Figure 4.8). Indeed, the conjugation efficiency obtained with sExpress was 1070-fold higher than the efficiency obtained with CA434 (p-value < 0.005; unpaired t-test).

A preliminary TraDIS experiment was performed where, after conjugation, *C. sporogenes* cells were grown under different selective conditions. First, the cells were grown with erythromycin to select cells that incorporated the pMTL-CW21 vector. Then, cells were transferred to erythromycin plates containing aTc – the inducer should cause plasmid loss, so the number of cells growing on these plates should be reduced. The erythromycin-resistant cells were also plated onto medium containing thiamphenicol and aTc – the anhydrotetracycline should induce transposition events and plasmid loss, while the thiamphenicol should select *C. sporogenes* cells that incorporated the transposon in their genome (mutants). Both *E. coli* CA434 and *E. coli* sExpress were tested as donors, and the conjugations were undertaken in triplicates. The transposition ratio and percentage of plasmid loss were determined using the formulas:

$$Transposition\ ratio = \frac{CFU/mL\ (Tm + aTc) - CFU/mL\ (Erm + aTc)}{CFU/mL\ (Erm)}$$

$$Plasmid\ loss = \frac{CFU/mL\ (Erm) - CFU/mL\ (Erm + aTc)}{CFU/mL\ (Erm)} \times 100\%$$

Table 4.1: Transposition ratio and percentage of plasmid loss obtained in the preliminary TraDIS experiment. The vector pMTL-CW21 was transferred from *E. coli* CA434 or *E. coli* sExpress into *C. sporogenes*. The conjugated cells were grown in different selective media, allowing the calculation of transposition ratio and plasmid loss. The number of mutants obtained per conjugation was determined, as well as the number of conjugations required to obtain one million mutants. The results presented show the average of three independent conjugations.

Donor strain	Transposition ratio	Plasmid loss	Mutants/ conjugation	Conjugations for 1×10^6 mutants
***E. coli* CA434**	4.42×10^{-4}	99.9998%	2.07×10^{3}	483.87
***E. coli* sExpress**	1.82×10^{-4}	99.9999%	5.67×10^{4}	17.65

The results presented in Table 4.1 indicate that good plasmid loss and transposition ratios were achieved regardless of which donor strain was used, indicating that the tetracycline system efficiently induced transposition and plasmid loss. Moreover, it appears that almost 500 conjugations would be required to obtain one million mutants using CA434 as donor strain, while less than 20 conjugations should be sufficient using sExpress. The results presented above confirmed that *E. coli* sExpress was the best donor strain to conjugate pMTL-CW21 into *C. sporogenes*, and that the tetracycline inducible system enabled good transposition and plasmid loss.

4.3.3 Optimisation of a spore purification protocol

Following the optimisation of the steps required to generate the transposon mutant library, the protocol to identify genes essential for sporulation/germination was the next issue to consider. The identification of such genes requires the library to be exposed to starvation conditions, as these are known to induce sporulation in *C. sporogenes*. Then, the goal was to extract genomic DNA only from the cells that are able to sporulate, in order to assess which genes could not tolerate transposon insertions – those are the genes essential for sporulation. This required the separation of the spores from any remaining vegetative cells that might have survived the starvation conditions.

Previous work carried out by Dr Aleksandra Kubiak allowed the optimisation of a spore purification method for *C. sporogenes* using a sucrose gradient. However, Dembek *et al.*[114] used a HistoDenz gradient to separate *C. difficile* spores from vegetative cells in their TraDIS work. Thus, both methods were tested and compared in the present study.

Accordingly, three independent cultures of *C. sporogenes* WT were grown in TYG for ten days to allow the formation of heat-resistant spores. The samples were incubated at 80°C for 20 minutes to eliminate vegetative cells and then washed several times. The resulting cells were separated in a gradient (50% w/v sucrose or 50% w/v HistoDenz). The spores were treated with DNase to ensure that any remaining vegetative cells would be killed and that their DNA would not affect the final sequence analysis. Finally, the spores were washed and observed under the microscope. The quality of the pictures obtained using phase-contrast microscopy was poor, so the spores were also stained using the method developed by Schaeffer *et al.*[122].

The pictures shown in Figure 4.9 indicate that both purification methods allowed the elimination of most vegetative cells, but that some cellular debris remained in the samples regardless of the purification method used. It also appeared that the amount of spores obtained with a HistoDenz gradient was lower than the amount of spores obtained using a sucrose gradient. In fact, the purification methods were tested several times and, on some occasions, no spores at all could be found after purification with HistoDenz (data not shown). The potential loss of spores during purification might lead to false positives in the TraDIS results: the lack of reads after sequencing would be a consequence of the spore being lost rather than the gene being essential. Thus, the sucrose purification method was selected, as it resulted in relatively pure spores without the loss of sample observed with HistoDenz.

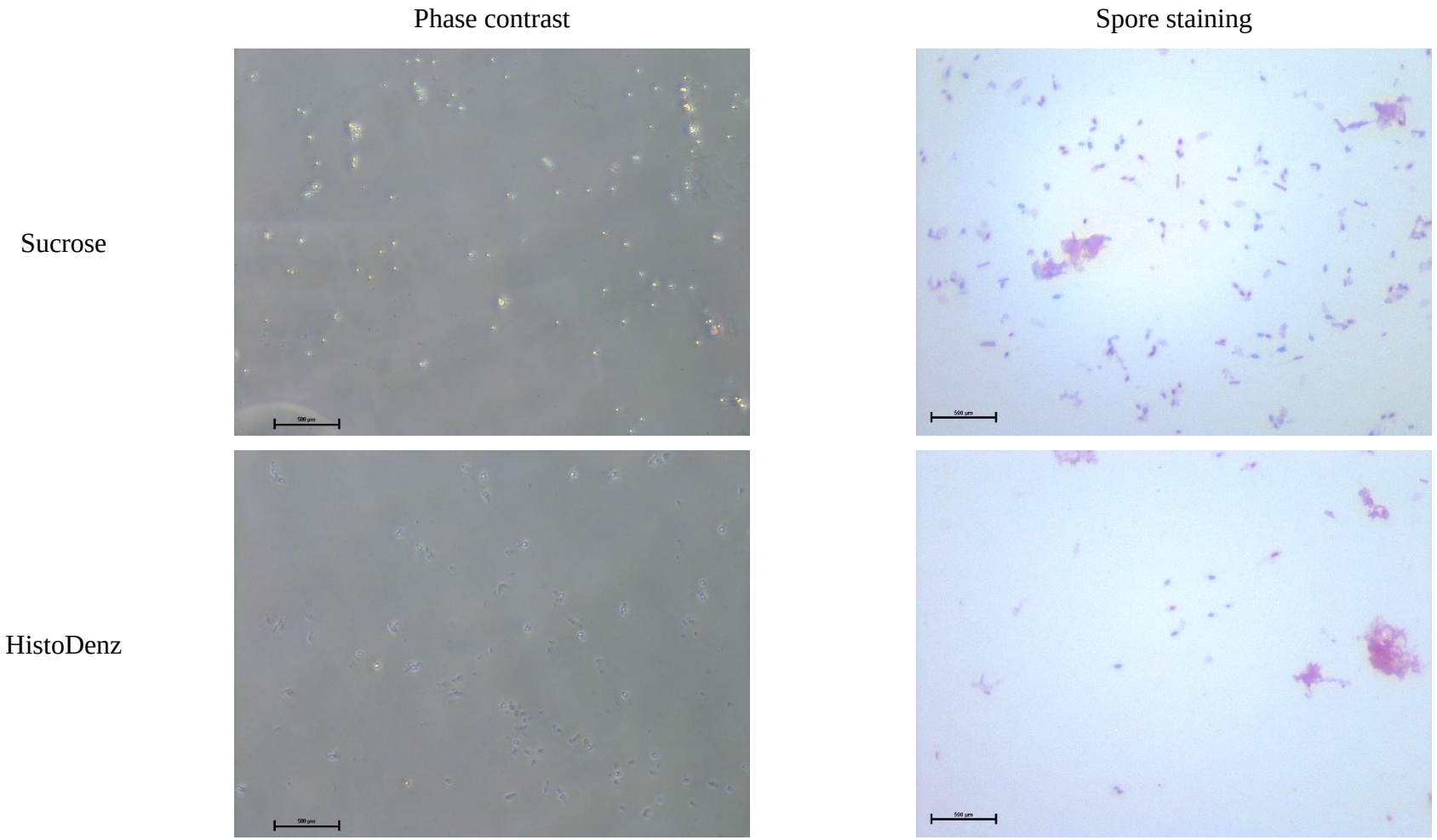


Figure 4.9: Microscopic visualisation of *C. sporogenes* spores following spore purification using Sucrose or HistoDenz gradients. The spores were observed using phase-contrast and stained using the Schaeffer *et al.*[122] method.

4.3.4 Generation of the transposon mutant library

Once all the optimisation steps were completed, the transposon mutant library was finally obtained. The library was generated by conjugating the vector pMTL-CW21 from *E. coli* sExpress into *C. sporogenes* WT. The *C. sporogenes* cells were first grown on plates containing erythromycin to select cells that incorporated the plasmid. Then, the cells were submitted to three different selection conditions: erythromycin, erythromycin with anhydrotetracycline, and thiamphenicol with anhydrotetracycline. The number of cells growing under each condition allowed the determination of the transposition ratio and percentage of plasmid loss, as described above. Using this approach, a total of 1.16×10^{6} mutants (growing on thiamphenicol with anhydrotetracycline) were obtained from 20 conjugations (as estimated above). The transposition ratio was 2.65×10^{-4} and plasmid loss was 99.99994%. The number of mutants and plasmid loss percentage appeared to be good enough to assess gene essentiality, so the mutants were pooled into a stock that contained the entire transposon mutant library.

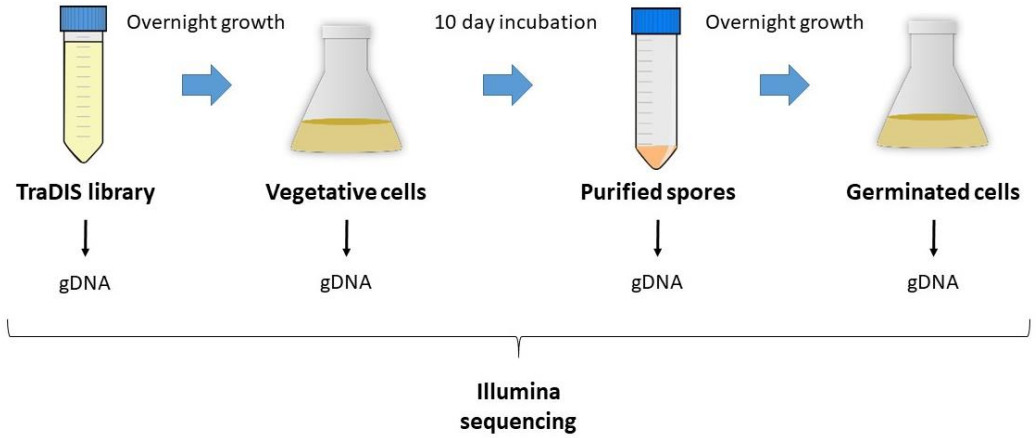


Figure 4.10: Schematic representation of the strategy used to identify genes essential for sporulation and germination in *C. sporogenes*. The transposon mutant library contained approx. one million mutants to allow determination of gene essentiality, and genomic DNA was extracted from the library stock to allow the identification of genes that are essential for the survival of *C. sporogenes*. The library was then used to inoculate fresh broth. Genomic DNA extracted after overnight growth allowed the identification of genes essential for vegetative growth. The remaining culture was incubated for 10 days to induce sporulation by starvation. The spores were purified to remove vegetative cells, and genomic DNA was extracted from the spores to allow the identification of genes essential for sporulation. Finally, the purified spores were used to inoculate fresh broth, triggering germination. Genomic DNA extracted from the germinated spores allowed the identification of genes essential for germination.

The process used to identify genes essential for survival, vegetative growth, sporulation and germination is represented in Figure 4.10. First, genomic DNA was

extracted from the transposon mutant library stock to allow the identification of genes that are essential for the survival of *C. sporogenes*. Then, cells from the stock were used to start three independent cultures - thiamphenicol selection was maintained to avoid the growth of wild type cells, and anhydrotetracycline was added to induce plasmid loss in any cells that might have retained the pMTL-CW21 vector. The cultures were grown overnight and genomic DNA was extracted from the vegetative cells. The remaining cultures were incubated for ten days to induce sporulation by starvation and, after that period, the spores from the three cultures were heat-treated and purified using a sucrose gradient. The purified spores from each culture were divided in two aliquots; one of the aliquots was used to extract genomic DNA from spores, and the other was used to inoculate fresh TYG medium (with thiamphenicol selection) and induce germination. Following overnight growth, the germination samples were sub-cultured onto fresh TYG medium, as it was previously determined that one sub-culture was enough to allow germination of all the spores (data not shown). Finally, genomic DNA was extracted from the germinated cells. All the genomic DNA samples were prepared for sequencing by Dr Craig Woods and sent to the University of Nottingham DeepSeq facility for Illumina sequencing. Following this approach, it was expected that:

i) Genes that had no insertions in the "stock library" corresponded to genes essential for survival of the organism;
ii) Genes that had insertions in the "stock library" but no insertions in the "vegetative growth library" should be required for vegetative growth;
iii) Genes that had insertions in the "vegetative growth library" but not in the "sporulation library" should be essential for sporulation;
iv) Genes that had insertions in the "sporulation library" but not in the "germination library" should be essential for germination.

4.3.5 Analysis of the TraDIS results

The sequencing of the libraries resulted in a large amount of data, and Dr Craig Woods did the processing and bioinformatic analysis of that data. A summary of the results obtained from the MiSeq Illumina Sequencing is presented in Table 4.2. The data indicates that the stock library contained 118,894 unique insertion sites (UIS). Given the size of the *C. sporogenes* genome (4,141,984 bp), this means that there was an average of one transposon insertion every 35bp. The number of reads that mapped to

the plasmid was very low in almost all libraries, as was expected given the high plasmid loss rate that was achieved and the I-*Sce*I digestion prior to sequencing. However, one of the libraries (vegetative cells, biological replicate 3) had a higher number of reads mapping to the plasmid, possibly due to the fact that there were several insertions in the region that contains the components of the tetracycline inducible system (data not shown). This system is responsible for the induction of plasmid loss, so the insertions might have compromised this mechanism in this sample. Nevertheless, it appears that the number of reads that mapped to the genome was still high enough to saturate it and assess gene essentiality.

In order to obtain accurate results, the number of insertions in each gene was normalised to the size of the gene, and the resulting value was the insertion index. The distribution of insertion indexes is bimodal, with the first peak corresponding to essential genes and the second broader peak corresponding to non-essential genes (Figure 4.11 - this figure only includes the plot corresponding to the stock library; the plots from the remaining libraries can be found in the Appendix, Figures S7 to S9). Accordingly, genes with an insertion index below 0.0106 were considered essential for survival and genes with an insertion index above 0.0149 were considered non-essential. All the genes with an insertion index between those two values were in the ambiguous zone. This analysis allowed the identification of 471 genes that appear to be essential for the survival of *C. sporogenes* under the conditions that were tested, while 170 are ambiguous.

A KEGG (Kyoto Encyclopedia of Genes and Genomes) analysis of the results obtained from the stock library allowed the identification of pathways that are essential for survival. Accordingly, several essential genes are involved in metabolic pathways (113 genes), RNA translation (49 genes), secondary metabolite biosynthesis (41 genes), biosynthesis of amino acids (25 genes), carbon metabolism (22 genes) and aminoacyl-tRNA biosynthesis (21 genes), among others presented in Table 4.3. These results are in line with what was observed in *C. difficile*[114] and are not surprising, as these are processes that are important for the cell.

Table 4.2: Summary of the data obtained from the MiSeq Illumina sequencing. The total number of reads obtained per biological replicate is shown. The number of unique insertion sites (UIS) in the genome allowed an estimation of the frequency of the insertions (genome size/UIS), indicating the average number of base pairs between each transposon insertion.

Sample	Biological replicate	Total reads	Unique insertion sites (UIS)	Genome size (bp)/UIS
TraDIS stock library	1	1,153,662	118,894	34.84
Vegetative growth library	1	2,269,833	133,348	31.06
	2	792,622	103,727	39.93
	3	1,024,833	65,832	62.92
Sporulation library	1	474,199	64,110	64.61
	2	3,664,847	123,387	33.57
	3	2,108,883	114,629	36.13
Germination library	1	4,337,340	113,675	36.44
	2	2,002,100	99,316	41.71
	3	1,588,887	88,688	46.70

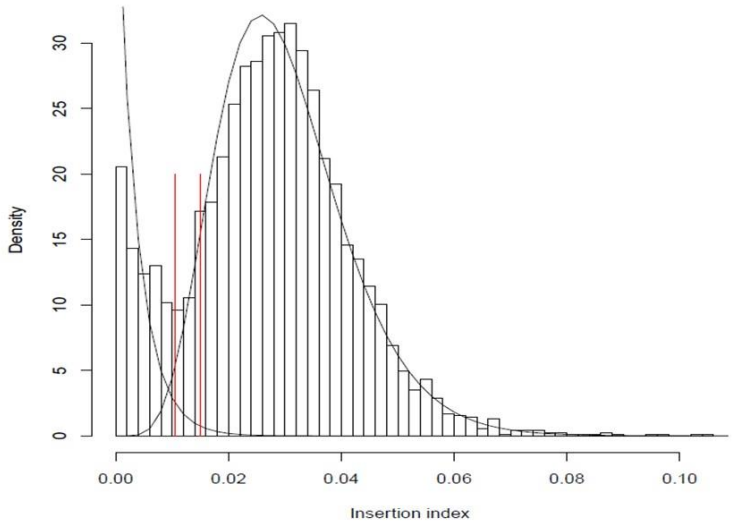


Figure 4.11: Plot demonstrating the identification of genes essential for survival of *C. sporogenes*. The plot shows the insertion index distribution from the TraDIS stock library. The black lines represent the fitted gamma distributions. The peak on the left represents essential genes, and the first red line represents the insertion index cut-off for essential genes. The broad peak to the right of the second red line represents genes that contain several insertions and are not essential. The second red line represents the cut-off for ambiguous genes.

Table 4.3: KEGG pathway analysis of essential genes. The number of essential genes identified in the stock and vegetative growth libraries are presented and compared to the number of essential genes identified in *C. difficile* by Dembek *et al.*[114].

KEGG Pathway	Number of essential genes		
	***C. difficile*[114]**	**Stock library**	**Vegetative growth library**
Ribosomal	46	49	2
Aminoacyl-tRNA biosynthesis	23	21	1
Metabolic pathways	104	113	24
Fatty acid metabolism	12	5	6
Peptidoglycan biosynthesis	12	13	1
DNA replication	9	9	1
Purine metabolism	18	8	1
Terpenoid backbone biosynthesis	9	7	0
RNA polymerase	4	4	0
Protein export	7	7	0
Lysine biosynthesis	9	8	0
Pyrimidine metabolism	14	6	1
Homologous recombination	8	13	2
Secondary metabolite biosynthesis	37	41	7
Oxidative phosphorylation	8	10	1
Mismatch repair	7	8	4
Folate biosynthesis	5	1	0
Carbon metabolism	18	22	5
D -Glutamine and D -glutamate metabolism	3	3	0
Glycerophospholipid metabolism	6	3	0
Microbial metabolism in diverse environments	25	36	8
Biotin metabolism	4	2	2
Propanoate metabolism	6	4	4

Table 4.3: KEGG pathway analysis of essential genes (continuation).

KEGG Pathway	Number of essential genes		
	***C. difficile*[114]**	**Stock library**	**Vegetative growth library**
Pantothenate and coenzyme A biosynthesis	5	5	0
Biosynthesis of amino acids	nd	25	0
Quorum sensing	nd	8	0
RNA degradation	nd	6	1
Bacterial secretion system	nd	6	0
Sulphur metabolism	nd	2	0
Sulphur relay system	nd	2	0

In addition to the assessment of the genes essential for survival (no insertions in the stock library), the genes essential for vegetative growth were also assessed. These were the genes that had insertions in the stock library but no insertions after overnight growth. Accordingly, 142 genes appear to be essential for vegetative growth while 106 are in the ambiguous zone. The essential genes are involved in microbial metabolism in diverse environments (eight genes), secondary metabolite biosynthesis (seven genes), fatty acid metabolism (six genes) and metabolic pathways (24 genes) (Table 4.3). The full lists of candidate genes essential for survival and for vegetative growth is presented in the Appendix (Tables S4 and S5).

4.3.5.1 Identification of genes essential for sporulation

Following the analysis of the genes essential for survival or vegetative growth, the genes required for sporulation were identified. These were the genes that had few or no insertions in the sporulation library but had several insertions in the stock and vegetative growth libraries. A total of 116 genes satisfied those conditions and were identified as essential for sporulation. Moreover, the analysis of the mutant frequencies allowed the identification of 154 genes that had significantly fewer insertions in the sporulation library relative to the vegetative cells library ($log_2FC < -2$; q-value < 0.05) (Figure 4.12). Combined, these results indicate that 183 genes might be essential for sporulation in *C. sporogenes*.

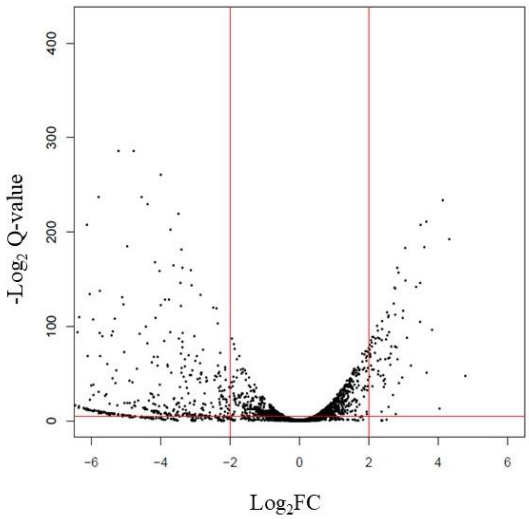


Figure 4.12: Plot showing the difference in mutant frequencies in the spores relative to vegetative cells. The vertical red lines indicate the cut-off selected for the $\log_2$ fold change (Log_2FC), which were -2 and 2. The horizontal line represents the cut-off for significant results (0.05), inferred from the Log_2 Q-value.

The lists overlap, with some genes being essential for sporulation and, not surprisingly, having significantly less insertions than in the vegetative cells library. However, some genes that have $log_2FC < -2$ do not satisfy the essentiality criteria. Nevertheless, as these genes can tolerate fewer insertions in the sporulation library, they are potentially essential for sporulation. The complete list of candidate genes essential for sporulation is presented in the Appendix (Table S6).

A partial list of candidate genes essential for sporulation is presented in Table 4.4 and includes the master regulator of sporulation, *spo0A* (CLSPO_c18650), as well as the genes encoding the sigma factors σ^E (*sigE* - CLSPO_c26280), σ^G (*sigG* - CLSPO_c26270) and σ^K (*sigK* - CLSPO_c26360). Other genes that have a known role in sporulation were also identified, such as *spoIID* (CLSPO_c01900), *spoIIM* (CLSPO_c18600), the *spoIIIA* operon (CLSPO_c18810 to CLSPO_c18880), *spoIVA* (CLSPO_c26120), *cotS* (CLSPO_c01350), *ytaF* (CLSPO_c16610), *yqfC* (CLSPO_c30420) and *ftsK* (CLSPO_c25070). The proteins SpoIID and SpoIIM have an important role in sporulation, as they are required for proper engulfment and prevent the formation of a second septum in the mother cell[34]. The products of the *spoIIIA* operon, on the other hand, appear to be relevant for the formation of the channel between the mother cell and the forespore, as well as the activation of the late-stage sigma factor σ^G [10, 34]. Proteins SpoIVA and CotS play a role in the formation of the coat layer[123-125], while YqfC is possibly involved in the formation of the cortex[126]. The

DNA translocase FtsK is part of the same family as SpoIIIE, and at least one of these proteins must be present to ensure the transfer of the entire chromosome from the mother cell to the forespore[34, 127]. Some putative spore coat proteins (CLSPO_c01400 and CLSPO_c30310) have been also identified, as well as a stage II sporulation protein (CLSPO_c30560), a sporulation domain-containing protein (CLSPO_c18220) and a putative sporulation protein YtaF. These results appear to confirm the role of these proteins in sporulation, but more studies are required to further elucidate that role. It should be noted that the stage II sporulation protein encoded by CLSPO_c30560 is identical to the gene encoding SpoIIP in *C. sporogenes* strain DSM 795 (blastn; 100% identity; e-value = 0). The fact that a sensor histidine kinase (CLSPO_c09720) appears to be essential for sporulation is also relevant, as this kinase might be involved in the phosphorylation and activation of Spo0A.

Other genes had an insertion index above the essentiality cut-off but significantly fewer insertions in spores relative to vegetative cells. These include *spoIIR* (CLSPO_c01510), *spoIIGA* (CLSPO_c26290), *spoIIAA* (CLSPO_c31830), *spoIVB* (CLSPO_c18660), *spoIIID* (CLSPO_c01920) and *yabQ* (CLSPO_c37140). Some of these genes are involved in the activation of the sigma factors, such as *spoIIAA* (activation of σ^F), *spoIIR*, *spoIIGA* (activation of σ^E) and *spoIVB* (activation of σ^K)[34]. The protein SpoIIID also has a regulatory activity by activating or repressing the expression of some genes from the σ^E regulon[34]. Other genes are possibly involved in the formation of the cortex, such as *yabQ*[128].

The genes *spoIID*, *spoIIM* and *ftsK* have been targeted with directed mutagenesis in *C. sporogenes* and the respective mutants did not produce any heat-resistant spores (Rubab Rizvi, personal communication). These results confirm that the products of these genes are essential for sporulation in this organism.

Table 4.4: Partial list of candidate genes essential for sporulation in *C. sporogenes*. The insertion index obtained for each biological replicate is shown, as well as the cut-off for essential and ambiguous genes (values from the sporulation library). Bold numbers correspond to an insertion index below the essentiality cut-off, and numbers with * correspond to values from the ambiguous region. The last column indicates the log_2 fold change of the mutant frequencies in the sporulation library relative to the vegetative growth library. Bold numbers represent significant differences ($log_2FC < -2$ and q-value < 0.05).

		Insertion index			
		Replicate 1	Replicate 2	Replicate 3	
Essential cut-off		0.0033	0.0069	0.0063	
Ambiguous cut-off		0.0046	0.0092	0.0084	
Gene ID	**Gene name/description**				**Log_2FC**
CLSPO_c01350	*cotS*	**0.0022**	0.0165	0.0088	**-4.55**
CLSPO_c01400	putative spore coat protein	**0.0011**	**0.0044**	0.0066*	**-6.61**
CLSPO_c01510	*spoIIR*	0.0140	0.0281	0.0316	**-2.85**
CLSPO_c01900	*spoIID*	**0.0021**	**0.0062**	**0.0052**	**-3.72**
CLSPO_c01920	*spoIIID*	0.0130	0.0174	0.217	**-2.78**
CLSPO_c09720	Sensor histidine kinase	**0.0000**	**0.0008**	**0.0008**	**-5.54**
CLSPO_c16610	*ytaF*	**0.0018**	0.0071*	0.0142	-1.76
CLSPO_c18220	sporulation domain-containing protein	**0.0009**	**0.0019**	**0.0019**	**-6.61**
CLSPO_c18600	*spoIIM*	0.0051	**0.0051**	0.0068*	**-2.32**
CLSPO_c18650	*spo0A*	**0.0000**	**0.0000**	**0.0000**	**-6.13**
CLSPO_c18660	*spoIVB*	0.0048	0.0113	0.0065*	**-2.02**
CLSPO_c18810	*spoIIIAH*	**0.0021**	**0.0042**	**0.0021**	**-5.41**
CLSPO_c18820	*spoIIIAG*	**0.0019**	**0.0058**	**0.0019**	0.48
CLSPO_c18830	*spoIIIAF*	**0.0019**	**0.0019**	**0.0038**	**-5.05**
CLSPO_c18840	*spoIIIAE*	**0.0019**	**0.0067**	0.0086	-1.89
CLSPO_c18850	*spoIIIAD*	**0.0029**	**0.0058**	**0.0029**	0.15
CLSPO_c18860	*spoIIIAC*	**0.0000**	0.0168	0.0112	**-3.44**
CLSPO_c18880	*spoIIIAA*	0.0036*	**0.0048**	0.0072*	-1.89
CLSPO_c25070	*ftsK*	**0.0010**	**0.0029**	**0.0034**	**-3.16**

Table 4.4: Partial list of candidate genes essential for sporulation in *C. sporogenes* (continuation).

Gene ID	Gene name/description	Insertion index			Log_2FC
		Replicate 1	Replicate 2	Replicate 3	
CLSPO_c26120	*spoIVA*	**0.0015**	**0.0022**	**0.0030**	**-4.15**
CLSPO_c26270	*sigG*	**0.0028**	0.0086*	0.0129	**-2.84**
CLSPO_c26280	*sigE*	**0.0031**	0.0125	0.0110	**-2.83**
CLSPO_c26290	*spoIIGA*	0.0070	0.0209	0.251	**-3.76**
CLSPO_c26360	*sigK*	**0.0031**	**0.0047**	**0.0047**	**-3.48**
CLSPO_c30310	spore coat protein-like protein	**0.0021**	0.0145	0.0124	**-2.14**
CLSPO_c30420	*yqfC*	**0.000**	**0.000**	**0.000**	**-6.77**
CLSPO_c30560	stage II sporulation protein	0.0041*	0.0081*	**0.0051**	**-2.28**
CLSPO_c31830	*spoIIAA*	0.0066	0.0297	0.0231	**-2.69**
CLSPO_c33910	Hypothetical protein	**0.000**	**0.000**	**0.000**	**-9.71**
CLSPO_c37140	*yabQ*	0.0111	0.0222	0.0167	**-3.41**

Finally, the list of genes essential for sporulation includes 15 hypothetical proteins, while 24 have significantly fewer insertions in spores relative to vegetative cells. Amongst the essential hypothetical proteins was one (CLSPO_c33910) that did not have any insertions in any of the biological replicates.

4.3.5.2 Identification of genes essential for germination

The genes essential for germination were ones that tolerated insertions in the sporulation library but not in the germination library. Furthermore, similar to what was described above, the frequency of mutants in the germination library was compared to the frequency of mutants in the sporulation library (Figure 4.13). Accordingly, 56 genes had an insertion index below the essentiality cut-off, and 132 had significantly fewer insertions in germinated cells relative to spores. Combined, these results indicate that 142 genes might be essential for germination in *C. sporogenes*.

A list containing some of the candidate genes essential for germination is presented in

Table 4.5, and the complete list is can be found in the Appendix, Table S7. The gene encoding the early-stage sigma factor σ^F (CLSPO_c31810) appears to be essential for germination, as well as the gene encoding one of the proteins required for the activation of σ^E [34] (*spoIIE* - CLSPO_c37040).

The list of genes essential for germination also includes genes directly or indirectly involved in the protection of the spore DNA, such as genes from the SpoVA family[129] (CLSPO_c31780, CLSPO_c31790, CLSPO_c31800) and a gene encoding a small acid-soluble protein[1] (CLSPO_c17810). The gene encoding SpoVT (CLSPO_c37200), known to regulate genes from the σ^G operon[130], is also part of this list, as well as two putative cortex-related proteins: YabP (CLSPO_c37150) and DacB (CLSPO_c18470)[34, 131, 132]. A M24 family metallopeptidase (CLSPO_c22010) also appears to be essential for germination, although the reason for this is not clear.

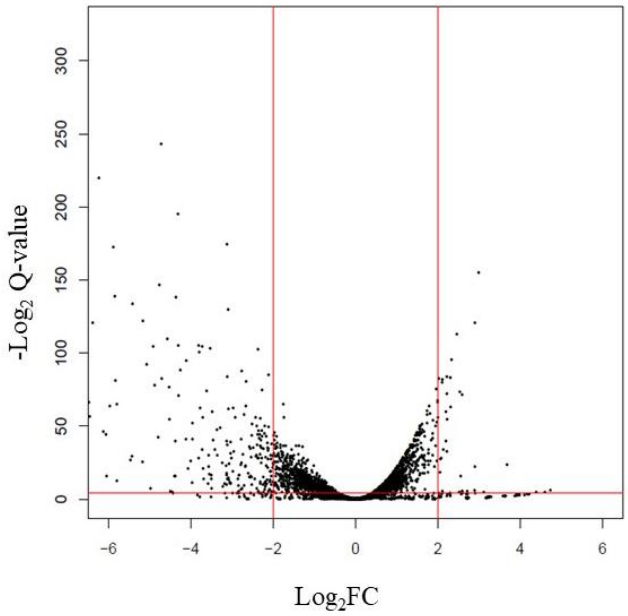


Figure 4.13: Plot showing the difference in mutant frequencies in germinated cells relative to spores. The vertical red lines indicate the cut-off selected for the $\log_2$ fold change (Log_2FC), which were -2 and 2. The horizontal line represents the cut-off for significant results (0.05), inferred from the Log_2 Q-value.

Table 4.5: Partial list of candidate genes essential for germination in *C. sporogenes*. The insertion index obtained for each biological replicate is shown, as well as the cut-off for essential and ambiguous genes (values from the germination library). Bold numbers correspond to an insertion index below the essentiality cut-off, and numbers with * correspond to values from the ambiguous region. The last column indicates the log_2 fold change of the mutant frequencies in the germination library relative to the sporulation library. Bold numbers represent significant differences ($log_2FC < -2$ and q-value < 0.05).

		Insertion index			
		Replicate 1	Replicate 2	Replicate 3	
Essential cut-off		0.0063	0.0036	0.0027	
Ambiguous cut-off		0.0088	0.0051	0.0039	
Gene ID	**Gene name/description**				**Log_2FC**
CLSPO_c01480	*gerKA4*	0.0190	0.0152	0.0106	**-4.31**
CLSPO_c15000	*divIVA*	**0.0051**	**0.0034**	0.0034*	-1.63
CLSPO_c17810	SASP beta	0.0110	**0.0000**	**0.0000**	**-2.19**
CLSPO_c18470	*dacB1*	**0.0029**	**0.0020**	**0.0020**	**-5.07**
CLSPO_c22010	M24 family metallopeptidase	**0.0051**	**0.0000**	0.0051*	**-3.58**
CLSPO_c30570	*gpr*	0.0068*	0.0091	0.0057	**-6.23**
CLSPO_c31780	*spoVA*	**0.0031**	0.0062	**0.0000**	**-7.61**
CLSPO_c31790	*spoVAD*	**0.0000**	**0.0000**	**0.0000**	**-7.61**
CLSPO_c31800	*spoVAC*	**0.0000**	**0.0000**	**0.0000**	**-5.96**
CLSPO_c31810	*sigF*	**0.0000**	**0.0029**	**0.0000**	**-7.71**
CLSPO_c37040	*spoIIE*	**0.0019**	**0.0019**	**0.0009**	**-7.81**
CLSPO_c37150	*yabP*	**0.0000**	**0.0000**	**0.0000**	**-8.41**
CLSPO_c37190	*spoVB2*	**0.0029**	**0.0022**	**0.0007**	**-8.18**
CLSPO_c37200	*spoVT*	0.0080*	**0.0020**	0.0040	**-4.52**

The gene *divIVA* (CLSPO_c15000) also appears to be essential for germination. The product of this gene is important for the localization of RacA to the cell pole, and RacA connects the chromosome to the pole of the cell. This is relevant for stage I of sporulation, and *B. subtilis divIVA* mutants have been shown to be defective in their ability to sporulate[34]. Moreover, an in-frame deletion of the gene encoding DivIVA has been created in *C. sporogenes* and the resulting mutant appears to be highly affected in its ability to form heat-resistant spores (Rubab Rizvi, personal

communication). Observation of the mutant under the microscope will determine whether the gene is involved in sporulation or germination.

Finally, a germinant receptor (CLSPO_c01480) and a germination protease (CLSPO_c30570) have significantly fewer insertions in germinated cells relative to spores. These genes might not be essential, as there are other proteins in the genome that have similar functions. Moreover, 9 hypothetical proteins appear to be essential for germination and 21 have fewer insertions in germinated cells than in spores.

4.4 Discussion

The objective of this chapter was to identify genes essential for sporulation and germination in *C. sporogenes*. Several optimisation steps were required to achieve this goal, including the selection of the plasmid to deliver the transposon machinery, selection of the best donor strain to conjugate that plasmid into *C. sporogenes* at high transfer rates and selection of the best gradient to purify the *C. sporogenes* spores. Once these optimisation steps were successfully completed, the TraDIS library was obtained and 471 genes were identified as essential for the survival of the organism. This corresponds to almost 13% of the annotated genes, which is in line with *C. difficile* (11% of the genes are essential[114]), *Burkholderia pseudomallei* (8.5% of the genes are essential[133]) and *Salmonella enterica* serovar Typhimurium (7.8% of the genes are essential[134]).

The results herein presented suggest that several genes are required for the formation of functionally heat-resistant spores. These include several genes that have been also identified as essential for sporulation/germination in *C. difficile*[114] and suggested to be part of the minimal set of genes required for sporulation[6]. The genes that are essential for sporulation/germination in *C. difficile* were identified using TraDIS[114]. On the other hand, the minimal set of genes required for sporulation was determined by identifying genes that are conserved in Bacilli and clostridia[6]. The results obtained in this chapter were compared to those studies (Table 4.6). The comparison shows that 33 genes were identified in the three studies, so the products of these genes are probably very important for the sporulation and/or germination processes. Moreover, 14 genes appear to be conserved[6] and relevant for sporulation in *C. sporogenes*, but these were not identified in the *C. difficile* study[114]. Despite the similarities, some genes identified as essential for sporulation in the previous studies were not essential for sporulation in *C. sporogenes*, as described below.

The sigma factor σ^H is known to regulate the expression of several sporulation-related genes, and the *sigH* gene appears to be essential for sporulation in other species[6, 114]. This sigma factor appears to be essential for the survival of *C. sporogenes*, as the gene *sigH* (CLSPO_c36620) did not contain insertions in any of the tested libraries. The same was observed with the 4-hydroxy-tetrahydrodipicolinate synthases DapA (CLSPO_c32570) and DapB (CLSPO_c32580), with the peptidyl-tRNA hydrolase Pth (CLSPO_c37230), and with the stage V sporulation proteins SpoVB1

(CLSPO_c14630), SpoVD1 (CLSPO_c14870), spoVE (CLSPO_c30870) and SpoVS (CLSPO_c25020), which appear to be essential for sporulation in *B. subtilis*[6]. Further investigations are required to confirm that these genes are in fact essential for survival, as these results might be false positives due to polar effects.

In addition, some proteins that appear to be essential for sporulation in *B. subtilis*[6] could not be found in the genomes of *C. sporogenes* and/or *C. difficile* and are not represented in Table 4.6. This could be due to differences in sporulation between Bacilli and clostridia, to low homology between the orthologues and/or to poor annotation of the proteins. Other genes that are conserved in spore formers[6] are present in the genome of *C. sporogenes* but have several transposon insertions in the sporulation and germination libraries. It is possible that the proteins encoded by these genes are involved in the sporulation process, but that other proteins perform similar functions. Thus, the mutants are able to form heat-resistant spores despite having insertions in those genes.

A surprising result was the observation that genes important for early stages of sporulation had an insertion index above the essentiality cut-off in the sporulation library. Those genes are responsible for the transcription and activation of early-stage sigma factors (*sigF*, *spoIIAA*, *spoIIE*, *spoIIGA* and *spoIIR*). Given the fact that σ^F and σ^E control the expression of the genes involved in early stages of sporulation, these genes were expected to be essential for sporulation. Nonetheless, these genes had significantly fewer insertions in spores relative to vegetative cells and/or appear to be essential for germination, so they might be required for the process of sporulation.

Overall, the results discussed above appear to confirm that the sporulation process is complex and that some of the mechanisms are conserved in different species. Nevertheless, it appears that some genes that are relevant for sporulation in other species are not present or not essential for sporulation in *C. sporogenes*. Further studies are required to fully elucidate how spore formation occurs in *C. sporogenes*, and this work provides the basis for those studies. Finally, several genes essential for sporulation or germination were identified and this information can potentially be used to create a conditionally sporulating strain. In order to confirm that these genes are required for sporulation, they need to be targeted with a directed mutagenesis approach and the ability of the mutants to form heat resistant spores needs to be evaluated. Accordingly, some of the genes identified above were selected for further characterisation and this will be the focus of Chapter 6.

Table 4.6: List of genes that are conserved and relevant for sporulation in *B. subtilis*, *C. sporogenes* and *C. difficile*. The genes in the first column are conserved in bacilli and clostridia, and some of them appear to be essential for sporulation in *B. subtilis*[6]. The genes in the *C. sporogenes* section were identified in this study. The data from *C. difficile* corresponds to genes that appear to be required for sporulation and was published by Dembek *et al.*[114]. Bold numbers indicate significant results (q-value < 0.05).

Conserved genes	*C. sporogenes*				*C. difficile*		
	Gene ID	**Description**	**Log_2FC (spo)**	**Log_2FC (ger)**	**Gene ID**	**Description**	**Log_2FC**
Genes that are conserved and relevant for sporulation/germination in *C. sporogenes* and *C. difficile*							
spo0A	CLSPO_c18650	stage 0 sporulation protein A	**-6.13**		CDR20291_1052	stage 0 sporulation protein A	**-10.53**
sigF	CLSPO_c31810	RNA polymerase sigma-F factor SigF	-0.27	**-7.71**	CDR20291_0701	RNA polymerase sigma-F factor	**-5.75**
sigE	CLSPO_c26280	RNA polymerase sigma-E factor SigE	**-2.83**	**-3.13**	CDR20291_2531	RNA polymerase sigma-E factor	**-11.16**
sigG	CLSPO_c26270	RNA polymerase sigma-G factor SigG	**-2.84**	**-5.17**	CDR20291_2530	RNA polymerase sigma-G factor	**-11.24**
spoIIAA	CLSPO_c31830	anti-sigma F factor antagonist	**-2.69**	**-3.09**	CDR20291_0699	anti-sigma F factor antagonist	**-10.92**
spoIIAB	CLSPO_c31820	anti-sigma F factor	**-1.51**	-1.53	CDR20291_0700	anti-sigma F factor	**-9.83**
spoIID	CLSPO_c01900	stage II sporulation protein D	**-3.72**	**-5.80**	CDR20291_0123	stage II sporulation protein D	**-8.99**
spoIIE	CLSPO_c37040	stage II sporulation protein E	**-1.30**	**-7.81**	CDR20291_3327	stage II sporulation protein E	**-8.57**
spoIIGA	CLSPO_c26290	putative sporulation sigma-E factor-processing peptidase	**-3.76**	**-3.81**	CDR20291_2532	sporulation sigma-E factor processing peptidase	**-12.62**

Table 4.6: List of genes that are conserved and relevant for sporulation in *B. subtilis, C. sporogenes* and *C. difficile* (continuation).

Conserved genes	*C. sporogenes*				*C. difficile*		
	Gene ID	**Description**	**Log_2FC (spo)**	**Log_2FC (ger)**	**Gene ID**	**Description**	**Log_2FC**
spoIIP	CLSPO_c30560	stage II sporulation protein	**-2.28**	**-7.56**	CDR20291_2362	putative stage II sporulation protein P	**-11.89**
spoIIR	CLSPO_c01510	stage II sporulation protein R	**-2.85**	**-7.63**	CDR20291_3401	stage II sporulation protein	**-7.94**
spoIIIAA	CLSPO_c18880	stage III sporulation protein AA	**-1.89**	**-6.45**	CDR20291_1030	stage III sporulation protein AA	**-12.55**
spoIIIAB	CLSPO_c18870	stage III sporulation protein AB	1.54	**-8.74**	CDR20291_1031	stage III sporulation protein AB	**-10.93**
spoIIIAD	CLSPO_c18850	stage III sporulation protein AD	0.15	**-6.12**	CDR20291_1033	stage III sporulation protein AD	**-10.06**
spoIIIAE	CLSPO_c18840	stage III sporulation protein AE	**-1.89**	**-7.41**	CDR20291_1034	stage III sporulation-related protein	**-13.13**
spoIIIAF	CLSPO_c18830	stage III sporulation protein AF	**-5.05**	-3.84	CDR20291_1035	putative stage III sporulation protein AF	**-5.49**
spoIIIAG	CLSPO_c18820	stage III sporulation protein AG	0.48	**-8.29**	CDR20291_1036	stage III sporulation protein AG	**-10.13**
spoIIID	CLSPO_c01920	stage III sporulation protein D	**-2.78**	**-6.47**	CDR20291_0125	stage III sporulation protein D	**-10.12**
spoIIIE	CLSPO_c25070	DNA translocase FtsK	**-3.16**	**-3.06**	CDR20291_3458	putative conjugative transposon FtsK_SpoIIIE-related protein	**-8.24**

Table 4.6: List of genes that are conserved and relevant for sporulation in *B. subtilis, C. sporogenes* and *C. difficile* (continuation).

Conserved genes	*C. sporogenes*				*C. difficile*		
	Gene ID	**Description**	**Log_2FC (spo)**	**Log_2FC (ger)**	**Gene ID**	**Description**	**Log_2FC**
spoIVA	CLSPO_c26120	stage IV sporulation protein A	**-4.15**	**-2.86**	CDR20291_2513	stage IV sporulation protein A	**-10.12**
spoIVB	CLSPO_c18660	SpoIVB peptidase	**-2.02**	**-2.29**	CDR20291_1051	stage IV sporulation protein B	**-4.07**
spoVAC	CLSPO_c31800	stage V sporulation protein AC	**-1.19**	**-5.96**	CDR20291_0702	stage V sporulation protein AC	**-11.01**
spoVAD	CLSPO_c31790	stage V sporulation protein AD	**-0.60**	**-7.61**	CDR20291_0703	stage V sporulation protein AD	**-6.19**
spoVD	CLSPO_c26370	stage V sporulation protein D	**-1.19**	-0.22	CDR20291_2544	stage V sporulation protein D	**-9.76**
spoVG	CLSPO_c37290	putative septation protein SpoVG	-0.13	**-1.16**	CDR20291_3353	stage V sporulation protein G	**-11.35**
spoVT	CLSPO_c37200	stage V sporulation protein T	**-1.20**	**-4.52**	CDR20291_3336	stage V sporulation protein T	**-6.94**
cwlD	CLSPO_c36050	N-acetylmuramoyl-L-alanine amidase	**0.68**	**-5.17**	CDR20291_1463	N-acetylmuramoyl-L-alanine amidase	**-2.65**
dacB	CLSPO_c18470	D-alanyl-D-alanine carboxypeptidase DacB	-0.15	**-5.07**	CDR20291_2048	D-alanyl-D-alanine carboxypeptidase	**-4.92**
yqfC	CLSPO_c30420	sporulation protein YqfC	**-6.77**	-1.83	CDR20291_2335	putative sporulation protein yqfc	**-9.00**
gpr	CLSPO_c30570	germination protease Gpr	**1.21**	**-6.23**	CDR20291_2363	germination protease	**-10.87**

Table 4.6: List of genes that are conserved and relevant for sporulation in *B. subtilis, C. sporogenes* and *C. difficile* (continuation).

Conserved genes	*C. sporogenes*				*C. difficile*		
	Gene ID	**Description**	**Log_2FC (spo)**	**Log_2FC (ger)**	**Gene ID**	**Description**	**Log_2FC**
lgt (gerF)	CLSPO_c33110	prolipoprotein diacylglyceryl transferase Lgt	**-4.38**	**-1.71**	CDR20291_2547	prolipoprotein diacylglyceryl transferase	**-4.02**
ssp	CLSPO_c17810	small, acid-soluble spore protein beta	-0.41	**-2.19**	CDR20291_3107	small acid-soluble spore protein B	**-6.99**
Genes that are conserved and relevant for sporulation/germination in *C. sporogenes*							
sigK	CLSPO_c26360	RNA polymerase sigma-28 factor SigK	**-3.48**	**-5.43**	-	-	-
spoIIM	CLSPO_c18600	stage II sporulation protein M	**-2.32**	**-7.32**	-	-	-
spoIIIAC	CLSPO_c18860	stage III sporulation protein AC	**-3.44**	-4.43	-	-	-
spoIIIAH	CLSPO_c18810	stage III sporulation protein AH	**-5.41**	-1.26	-	-	-
spoVAF	CLSPO_c15920	stage V sporulation protein AF	0.43	**-1.55**	-	-	-
spoVB	CLSPO_c14640	stage V sporulation protein B	-0.37	**-1.92**	-	-	-
	CLSPO_c37190		0.28	**-8.18**	-	-	-
pth (spoVC)	CLSPO_c37230	peptidyl-tRNA hydrolase Pth	**-4.12**	0.04	-	-	-

Table 4.6: List of genes that are conserved and relevant for sporulation in *B. subtilis, C. sporogenes* and *C. difficile* (continuation).

Conserved genes	*C. sporogenes*				*C. difficile*		
	Gene ID	**Description**	**Log_2FC (spo)**	**Log_2FC (ger)**	**Gene ID**	**Description**	**Log_2FC**
cotS	CLSPO_c01350	CotS family spore coat protein	**-4.55**	**-4.07**	-	-	-
obgE	CLSPO_c30770	GTPase Obg	**-3.37**	0.07	-	-	-
yabP	CLSPO_c37150	sporulation protein YabP	**-1.30**	**-8.41**	-	-	-
yabQ	CLSPO_c37140	spore cortex biosynthesis protein YabQ	**-3.40**	**-3.12**	-	-	-
alr (yncD)	CLSPO_c35540	alanine racemase 1	**-3.71**	0.05	-	-	-
gerKA	CLSPO_c01480	spore germination protein KA	0.39	**-4.31**	-	-	-
	CLSPO_c19770		-0.19	**-2.06**	-	-	-
	CLSPO_c28720		**0.58**	**-3.80**	-	-	-
gerAC	CLSPO_c28700	spore germination protein, GerAC	0.05	**-4.11**	-	-	-

Chapter 5: Regulation of the sporulation cascade by the sigma factors σ^E, σ^F, σ^G and σ^K

5

5.1 Background

5.1.1 Activation of sigma factors and their role in regulation of sporulation

The sigma factors σ^F, σ^E, σ^G and σ^K and their involvement in the regulation of the sporulation cascade have been briefly mentioned in Chapter 1. This regulation is spatially and temporally specific, so that the sigma factors are inactive and are only activated in the correct compartment and at the correct time. The early stage sigma factors appear to be activated after the formation of the septum between the mother cell and the forespore, *i.e.*, after the asymmetric division of the cell into two compartments. This is not surprising, as the activity of each sigma factor is compartmentalised, with σ^F being active only in the forespore and σ^E only in the mother cell. The activity of late stage sigma factors, on the other hand, appears to be coupled to engulfment[34, 58].

The forespore-specific σ^F is the first sigma factor to be activated. This sigma factor is then involved in the activation of σ^E in the mother cell and in the transcription of *sigG* (encoding σ^G) in the forespore. The forespore-specific σ^G is activated in response to a signal from σ^E. In turn, *sigK* transcription is regulated by σ^E and the encoded σ^K is activated in response to a signal from σ^G [34] (Figure 5.1). This criss-cross regulation is important for the correct compartment- and time-specific activation of each sigma factor, as well as to the coupling of regulation to morphological events.

In *B. subtilis*, *sigF* and *sigE*, respectively encoding σ^F and σ^E, are regulated by Spo0A[13]. Consequently, and as mentioned above, the sigma factors involved in the regulation of early sporulation are produced in the pre-divisional cell but only activated when the septum is fully formed and asymmetric division is complete[34]. The activation of these sigma factors involves several proteins (SpoIIAA, SpoIIAB and SpoIIE for σ^F; SpoIIGA and SpoIIR for σ^E)[37] and will be described in Chapter 6. Briefly, SpoIIAB inactivates σ^F by its binding and sequestration into an inactivation complex[135]. When septum formation is complete, SpoIIE dephosphorylates SpoIIAA which, in turn, induces the release of σ^F from SpoIIAB[136]. Then, the active σ^F induces the transcription

of *spoIIR*[60], and the product of this gene signals SpoIIGA to cleave a pro-sequence from σ^E, leading to its activation[137].

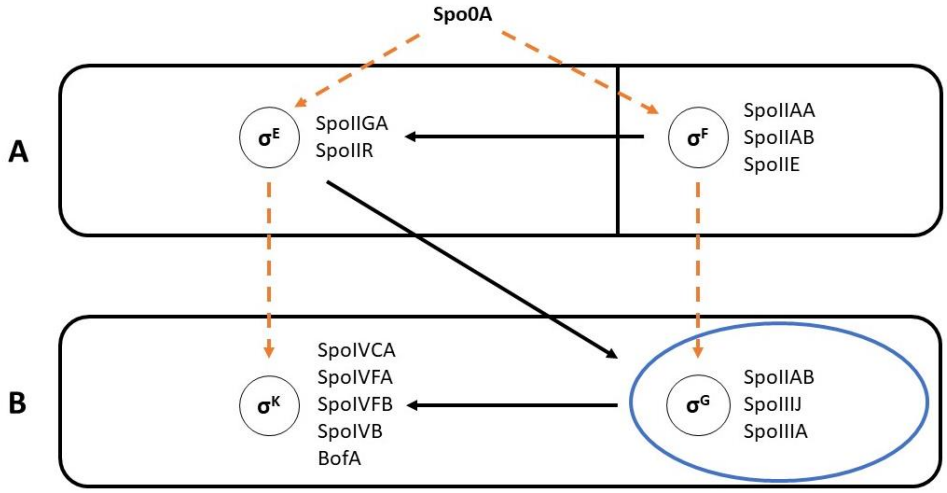


Figure 5.1: Schematic representation of the criss-cross regulation of sigma factors. Orange, dotted arrows represent transcription regulation, while black, solid arrows represent post-translational regulation. Next to each sigma factor is a list of proteins involved in their activation. Note that the proteins are next to the sigma factor they activate, but not necessarily in the compartment where they are expressed/active. **A:** The mother cell and forespore compartment appear separated by the septum, represented as a black line. σ^F is activated in the forespore upon asymmetric division completion. Then, σ^E is activated in the mother cell in response to the σ^F-mediated expression of SpoIIR in the forespore. Both σ^F and σ^E are transcriptionally regulated by Spo0A. **B:** The forespore is represented as a blue smaller cell inside the bigger mother cell. σ^F is responsible for the transcription of the gene encoding σ^G in the forespore. Following engulfment, σ^G is activated due to a signal from SpoIIIA (synthesised in the mother cell). Likewise, the gene encoding σ^K is transcribed by σ^E in the mother cell and activated due to a signal from the forespore (SpoIVB).

During the late stages of sporulation, the expression of *sigG* is positively regulated by both σ^F and σ^G and is consequently confined to the forespore compartment[138]. Surprisingly, the mother cell-specific σ^E is also required for the transcription of *sigG*[139]. The late sigma factor σ^G is inactive before engulfment and is seemingly activated by SpoIIIJ and SpoIIIA[37, 140, 141]. It appears that following translation, σ^G is inhibited by the anti-sigma factor SpoIIAB, which is also responsible for the inhibition of σ^F [141, 142]. It has been hypothesised that a signal from the mother cell is sent to the forespore and induces the release of σ^G from SpoIIAB. The proposed signalling pathway involves the products of the *spoIIIA* operon in the mother cell side of the forespore membrane, and the putative lipoprotein SpoIIIJ on the forespore side[34, 140, 141]. Indeed, SpoIIIA appears to be responsible for the removal of SpoIIAB from the forespore following engulfment[142]. It has also been suggested that SpoIIQ has a role in the signalling pathway due to its involvement in the formation of the feeding tube[143], and that the σ^F-dependent CsfB prevents the premature activation of σ^G [144]. Overall,

the activation of σ^G is a complex process involving several proteins, and has yet to be fully elucidated.

Similar to what is observed with the other sigma factors, σ^K is inactive in the *B. subtilis* mother cell following *sigK* expression, which is regulated by σ^E and σ^K. Activation of σ^K requires the removal of a *skin* (sigma K intervening) element from *sigK*, as well as cleavage of a pro-peptide from pro-σ^K. The removal of the *skin* element is mediated by SpoIVCA (regulated by σ^E and SpoIIID), while the cleavage of the pro-peptide is mediated by SpoIVFB, which is activated by SpoIVB (regulated by σ^F and σ^G)[10]. The processing of pro-σ^K by SpoIVFB appears to occur within the cell membrane and to be inhibited by SpoIVFA and BofA. It has been suggested that this inhibition stops once SpoIVB, which is synthesised in the forespore, degrades SpoIVFA[34, 37]. The redundant mechanisms of σ^K activation (*skin* element and processing) are not observed in all spore formers. The gut pathogen *C. difficile*, for instance, contains an insertion with different characteristics than the *skin* element and does not contain a pro-sequence. Other species contain the pro-sequence but no *skin* element[34].

Previous studies on *C. acetobutylicum, C. botulinum, C. perfringens* and *C. difficile* show differences between the different sigma factor mutants regarding spore morphogenesis[10]. Moreover, *sigK* appears to have a role in early sporulation in *C. perfringens*[145] and *C. botulinum*[146]. These differences highlight the fact that although the sigma factors are conserved across spore formers, their regulatory role appears to be diverse in different species. Nevertheless, some functions are conserved in the regulons of sigma factors from *B. subtilis*[34, 58, 126] and *C. difficile*[147]:

i) σ^F-dependent genes regulate other genes (directly or indirectly) and spore morphogenesis;
ii) σ^E-dependent genes are involved in engulfment, as well as initial stages of spore coat and cortex formation;
iii) σ^G-dependent genes are responsible for the protection of the spore DNA and for the expression of germinant receptors;
iv) σ^K-dependent genes encode spore coat proteins and are involved in spore maturation.

5.1.1.1 Co-regulation of genes by more than one sigma factor/regulatory gene

Some genes can be regulated by more than one sigma factor, or by a sigma factor and another regulatory protein. Co-regulation of genes by σ^F and σ^G in the forespore has been reported, as well as co-regulation by σ^E and σ^K in the mother cell[34, 147]. Furthermore, SpoIIID and GerE also regulate gene expression in the mother cell and activate or repress the transcription of certain genes. The proteins SpoIIQ, RsfA, SplA and SpoVT appear to have a similar role in the forespore, although the mechanism of regulation is less clear[34, 130]. The regulation by more than one regulatory protein ensures the correct timing of gene expression, repressing transcription when the gene is no longer needed[34].

5.1.2 RNA-Seq: a high-throughput approach to transcriptomics

The determination of the regulons of the sigma factors requires the study of gene expression and transcriptomics. The transcriptome has been defined by Wang *et al.*[148] as "the complete set of transcripts in a cell, and their quantity, for a specific developmental stage or physiological condition". Hence, transcriptomics focuses on the study of coding RNAs and how their expression levels change in response to intra or extra-cellular signals[148]. While the genome is stable and does not change with the environment, the transcriptome is constantly changing and sheds light on the function of the genes and the processes in which they are involved.

RNA-Seq is characterised by the study of transcriptomes by sequencing cDNA libraries synthesised from mRNA. This technique allows the quantification of gene expression levels based on the number of reads that map to each gene – more reads are indicative of higher expression. In order to obtain accurate results, the number of reads must be normalised to the size of the gene. One of the main advantages of RNA-Seq is that this high-throughput approach takes advantage of deep-sequencing technologies, allowing the study of the entire transcriptome. Moreover, it is a sensitive technique with a large dynamic range and low background noise[148, 149].

The study of transcriptomes is based on mRNA, which contains coding information, and the presence of tRNA and rRNA (more abundant) can compromise the success of RNA-Seq experiments. In eukaryotes, the poly-A tails of mRNA molecules are helpful to separate them from tRNA and rRNA[148]. However, since prokaryotic mRNA does

not contain poly-A tails, other methods are normally used to enrich the mRNA. These methods include the elimination of rRNA by digesting it with specific nucleases or by capturing it via hybridisation[149].

The several steps typically required to perform an RNA-Seq experiment in bacteria are listed in Figure 5.2. Although some of these steps might introduce biases, this approach is still more reliable and reproducible than previously used techniques. Moreover, advances in sequencing technologies are reducing these biases even further[149].

In addition to its usefulness as a tool to assess gene expression levels, RNA-Seq can also be used to determine transcriptional start sites, identify operons and small regulatory RNAs (sRNAs), and shed light on untranslated regulatory regions[149].

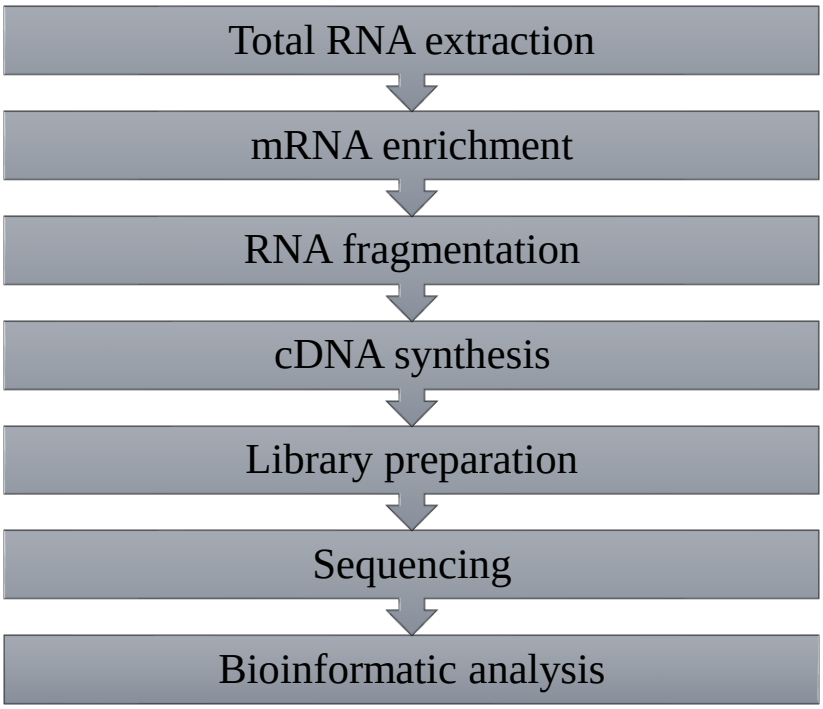


Figure 5.2: Flowchart representing the different steps usually required to perform RNA-Seq experiments in bacteria. Following total RNA extraction, the mRNA is enriched, fragmented and turned into cDNA. The preparation of the library depends on the sequencing platform to be used, but frequently involves cDNA amplification and size-selection. Adapted from Pinto *et al.*[150].

5.1.3 Other methods to assess gene expression

Approaches like RNA-Seq do not require prior knowledge about sequence or expression levels of the genes, and provide an overview of the entire transcriptome. However, when the goal is to investigate the expression levels of a specific gene, there are other techniques available. Similar to RNA-Seq, the first step of these PCR-based techniques is the conversion of RNA into cDNA. Then, the amount of cDNA can be

determined using typical end-point PCR amplification (RT-PCR) or measured in real-time (RT-qPCR).

In the first method (RT-PCR), the amount of DNA in the sample is inferred from the intensity of the bands on an agarose gel[151]. Hence, it does not provide an accurate measurement of cDNA concentration.

The RT-qPCR method, on the other hand, provides a sensitive, accurate and reproducible measurement of the initial concentration of DNA with a large dynamic range[152]. This method is fluorescence-based and relies on the fact that some DNA binding dyes have increased fluorescence when binding to double stranded DNA. The number of cycles required to reach a certain fluorescence threshold (C_T value; Figure 5.3) is inversely proportional to the initial amount of DNA in the sample (the C_T value is nowadays referred to as C_q value – quantification cycle)[152, 153].

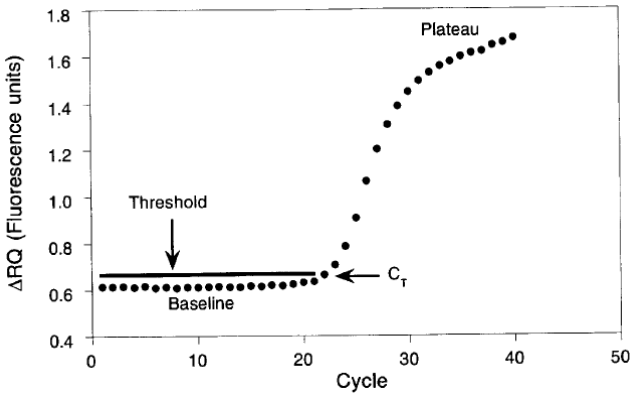


Figure 5.3: Representation of how fluorescence increases with cycle number in qPCR experiments. During the first few cycles, fluorescence is low and reflects the low efficiency of the PCR amplification. Then, when the efficiency of the reaction is optimal and the amount of DNA doubles with each cycle, fluorescence increases exponentially. In a third stage, efficiency decreases again due to reagent limitation and fluorescence reaches a plateau. Quantification of DNA is based on the cycle number at which fluorescence reaches a certain threshold (C_T). Source: Heid *et al.*[152].

The linearity between the C_q value and DNA quantity has been demonstrated by Heid *et al.*[152]. This linearity relies on a high efficiency of PCR amplification: the amount of DNA should double at each cycle. Several factors can influence the efficiency of the reaction, including the PCR target size. For best results, the amplicon should be small – usually 100 to 300 base pairs. Furthermore, in order to correct for other variations that might affect the accuracy of this method, a reference gene (frequently a housekeeping gene) is included to normalise the results[153].

5.1.4 RiboCas: a gene editing tool used to target the genes encoding the sigma factors

The gene-editing tool CRISPR/Cas9, already described in Chapter 3, was the method selected to knock-out the genes encoding the sigma factors. However, a different version of this system was used in the present chapter: the RiboCas[87]. The main difference relative to the system used in Chapter 3 is that the RiboCas uses a theophylline responsive riboswitch to control the expression of Cas9[87] rather than a constitutive promoter driving the expression of trCas9[86].

The use of the RiboCas method introduced some advantages relative to the previously used system. One of them is the fact that it allows the use of Cas9 instead of trCas9. The high toxicity of Cas9 usually makes it difficult to clone the vectors required to introduce the CRISPR/Cas9 system components into the target cells, and the less toxic trCas9 attenuated this problem. However, the use of riboswitches ensures low basal expression of Cas9, which facilitates the cloning step. Moreover, this low expression also allows more time for homologous recombination to occur before the introduction of the lethal double strand breaks, which only happens after induction of Cas9 expression. Once inside the target cell, the toxic Cas9 is more efficient than trCas9 at targeting and eliminating cells where recombination did not happen. These improvements increased the efficiency of gene editing[87].

5.2 Chapter overview

Four sigma factors are responsible for the regulation of the sporulation cascade in spore formers, and the goal of this chapter was to elucidate their role in *C. sporogenes*. With that aim, in-frame deletions of the four genes encoding these sigma factors were generated in the *C. sporogenes* genome. RNA-Seq was then used to compare the transcriptome of each mutant to that of the wild type, in order to establish the regulon of each sigma factor. This required the determination of optimal times for RNA extraction by RT-PCR and RT-qPCR.

5.3 **Results**

5.3.1 Deletion of the genes encoding σ^F, σ^E, σ^G and σ^K

In order to determine the regulon of each sigma factor, the genes *sigE*, *sigF*, *sigG* and *sigK* were independently targeted using the RiboCas system[87] described above. In addition, a bookmark was included in the resulting mutant strain to facilitate reversion of the gene to its original locus (see Chapter 6; Figure 5.4).

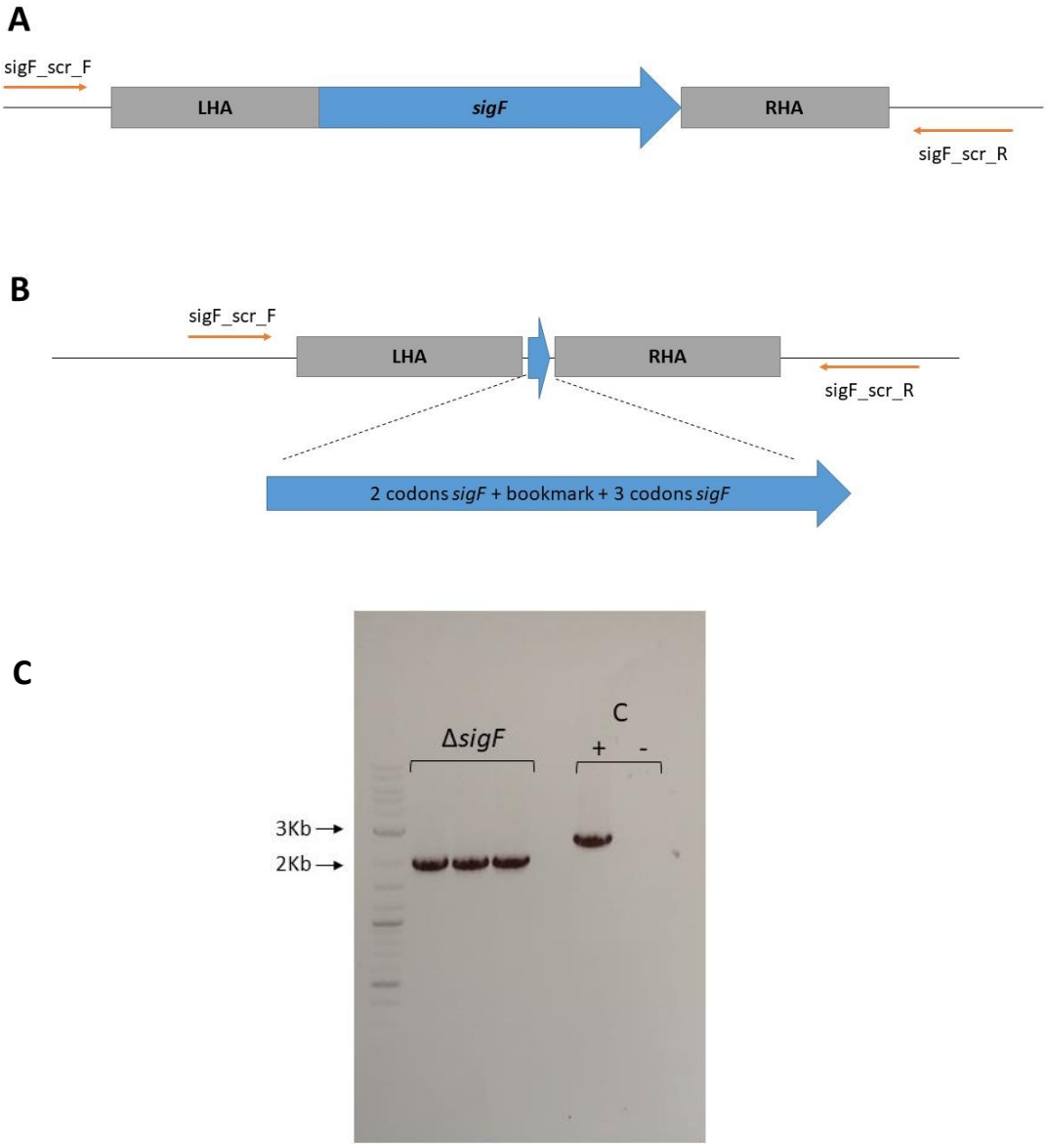


Figure 5.4: Representation of the colony PCR method used to confirm the deletion of the *sigF* gene, as an example. A and **B** represent the wild type strain and the mutant strain, respectively. The primers used in the colony PCR are represented, as well as the left homology arm (LHA) and the right homology arm (RHA). **C**: Gel electrophoresis picture of colony PCR samples demonstrating the deletion the *sigF* gene. The deletion was screened using sigF_scr_F and sigF_scr_R primers. The gene was also screened in wild type (C+), and a negative control (water as template) was included (C-).

The vectors pRECas1-*sigF*_BM, pRECas1-*sigE*_BM, pRECas1-*sigG*_BM and pRECas1-*sigK*_BM were used to knock-out *sigF*, *sigE*, *sigG* and *sigK*, respectively. The strains obtained using this method were named *C. sporogenes* Δ*sigF*, *C. sporogenes* Δ*sigE*, *C. sporogenes* Δ*sigG* and *C. sporogenes* Δ*sigK*. For each strain, three independent biological replicates were obtained. The deletion of the genes was confirmed by colony PCR, as exemplified in Figure 5.4, and by sequencing the region of the genome where the target gene originally resided.

5.3.2 Identification of the time points for RNA extraction

The main goal of this chapter was to determine the regulons of each sigma factor. The approach used to achieve this goal was RNA-Seq, a technique that allows the study of the entire transcriptome. Sigma factors are known to positively control the transcription of certain genes; consequently, the expression of those genes should be downregulated in the absence of the sigma factor that induces its transcription. Hence, the regulon of each sigma factor mutant can be determined by comparing the transcriptome of each mutant to wild type.

As described above, the sigma factors are usually only activated during sporulation: σ^F and σ^E activation is linked to septum formation, whereas σ^G and σ^K activities are coupled to completion of engulfment. Therefore, in order to determine the regulons of the sigma factors, the RNA must be extracted from wild type at times when the sigma factors are active. In a similar study in *C. difficile*, the authors extracted RNA 14 hours after inoculation to determine the regulons of σ^E and σ^F, 19 hours for σ^G and 24 hours for σ^K [147]. In *C. botulinum*, closely related to *C. sporogenes*, it was reported that *sigF*, *sigE* and *sigG* were expressed 11 hours after inoculation, while *sigK* was expressed 5 hours and then again 24 hours after inoculation[154].

The information found in the literature narrowed down the times of RNA extraction to the first 24 hours following inoculation. However, further experiments were necessary to determine the exact time points at which the sigma factors were active in *C. sporogenes*. This was achieved using RT-PCR and RT-qPCR, as described below.

5.3.2.1 RT-PCR

The time points for RNA extraction were first determined using RT-PCR and then confirmed using RT-qPCR. For the RT-PCR experiment, RNA was extracted from *C. sporogenes* WT at several time points: 5 hours, 8 hours, 10 hours, 14 hours,

17 hours, 20 hours and 24 hours (t = 0 hours corresponded to the moment of inoculation). Only two biological replicates were tested during this assay. The RNA was free from DNA contamination (confirmed by PCR) and its quality was high (see Appendix, Table S8). Finally, the RNA was converted into cDNA, which was then amplified using specific primers. The primers were carefully designed to amplify genes known to be regulated by each sigma factor in other species[34, 147]:

- *spoIID* (CLSPO_c01900) and *spoIIM* (CLSPO_c18600) for σ^E;
- *spoIIR* (CLSPO_c01510) and *gpr* (CLSPO_c30570) for σ^F;
- *spoIVB* (CLSPO_c18660) and *sspH1* (CLSPO_c08070) for σ^G;
- *cotA* (CLSPO_c08700) and *cotI* (CLSPO_c01380) for σ^K.

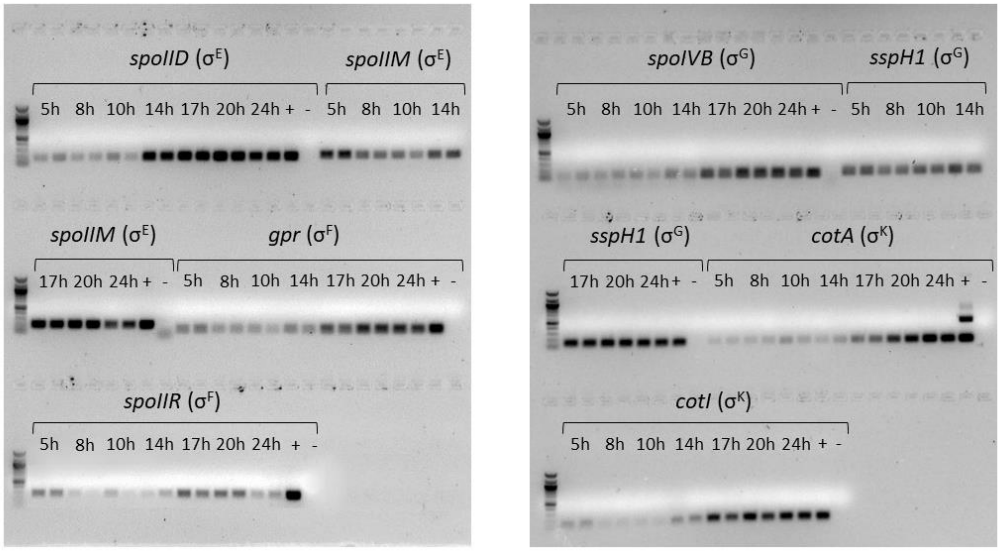


Figure 5.5: Gel electrophoresis showing changes in expression levels over time (RT-PCR). RNA was extracted from *C. sporogenes* WT (two biological replicates) at several time points (5 hours, 8 hours. 10 hours, 14 hours, 17 hours, 20 hours and 24 hours after inoculation). Then, the RNA was converted into cDNA and amplified using specific primers targeting genes known to be regulated by each sigma factor in other species. The intensity of the bands is proportional to the expression levels of each gene at each time point. A positive control (genomic DNA from *C. sporogenes* WT) and a negative control (water as template) were included for each gene.

The RT-PCR results (Figure 5.5) indicate that the expression levels of *spoIID* and *spoIIM* start to increase after 14 hours of growth, but the intensity of the bands is highest after 17 hours. The intensity of the bands also starts to increase after 14 hours for *gpr* and 17 hours for *spoIIR*. Given that the expression levels of these genes appear to be higher 17 hours after inoculation (except *gpr*, but intensity is already increasing), that was the time point selected for σ^E and σ^F. Regarding *spoIVB* and *sspH1*, the maximum expression of both genes appeared to be reached some 20 hours after inoculation, so that was the time point selected for σ^G. Finally, the time point selected

for σ^K was 24 hours, since expression of *cotA* and *cotI* appeared to be highest at this point. A second band is visible for the positive control of *cotA*. However, since the band is not present in the cDNA samples, it does not seem to affect the results.

5.3.2.2 RT-qPCR

The results obtained with RT-PCR allowed the selection of the time points for RNA extraction. However, only wild type samples were analysed and no comparisons were made between *C. sporogenes* WT and the different sigma factor mutants. In order to address this issue and confirm the different expression levels between wild type and each mutant, RNA was extracted from all the strains (three biological replicates each) at the selected time points:

- 17 hours: *C. sporogenes* WT, *C. sporogenes* Δ*sigE* and *C. sporogenes* Δ*sigF*;
- 20 hours: *C. sporogenes* WT and *C. sporogenes* Δ*sigG*;
- 24 hours: *C. sporogenes* WT and *C. sporogenes* Δ*sigK*.

The RNA was free from DNA contamination (confirmed by PCR) and its quality was assessed using Nanodrop and the 2100 Bioanalyzer instrument (see Appendix, Table S9). Then, the RNA was used as template for cDNA synthesis, and the relative expression levels were determined using RT-qPCR. This method was selected due to its high sensitivity, since the success of the RNA-Seq experiment depends on the correct determination of the time points for RNA extraction.

The RT-qPCR approach targeted the same genes described above (RT-PCR). A gene encoding 16S rRNA (*rrfA* - CLSPO_c00080) and another encoding a DNA gyrase (*gyrA* - CLSPO_c00070) were tested as reference genes, as they had been previously used in *C. sporogenes*[78]. However, only *rrfA* was used for normalisation because expression of *gyrA* was not sufficiently stable between the different strains. Calibration curves were obtained for each gene in order to determine the efficiency of each reaction (see Appendix, Table S10). Finally, relative expression of each gene was determined using the E-method[82].

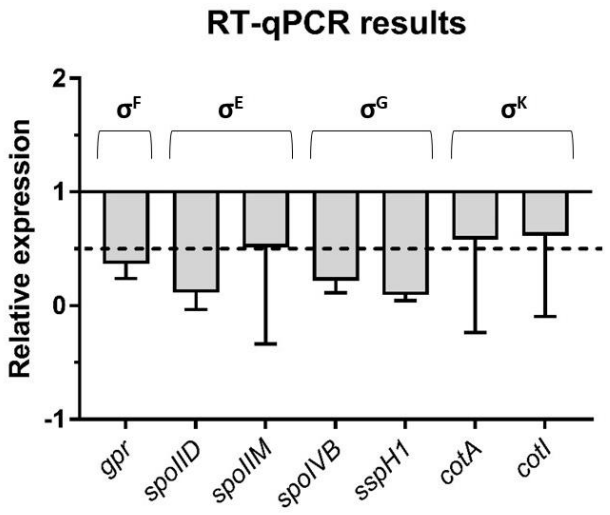


Figure 5.6: RT-qPCR results. Total RNA was isolated from *C. sporogenes* WT and each sigma factor mutant at specific time points. For each gene, the expression level of the respective mutant was compared to that of wild type (*e.g.*, expression of *spoIID* in the *sigE* mutant was compared to expression of *spoIID* in wild type). A relative expression below 0.5 was considered significant, and this threshold is indicated with a dotted line. Error bars represent standard deviations of three biological replicates.

The RT-qPCR results (Figure 5.6) compared the expression of each gene in the respective mutant to the expression of the same gene in wild type. All genes appear to be downregulated in the mutants, indicating that the genes are being expressed in wild type. Genes that had their expression levels reduced to half (relative expression lower than 0.5) were considered to be significantly downregulated. The genes *gpr* (regulated by σ^F), *spoIID* (regulated by σ^E), *spoIVB* and *sppH1* (regulated by σ^G) were all below this threshold. The genes *spoIIM* (regulated by σ^E), *cotA* and *cotI* (regulated by σ^K) were not significantly downregulated and showed some variation between the biological replicates. Nevertheless, these genes appear to be downregulated in the mutants. The relative expression of *spoIIR* (regulated by σ^F) is not shown because the primers were not specific: the presence of a second amplicon, revealed by the analysis of the melting curve, interfered with the quantification of gene expression.

Overall, the RT-PCR and RT-qPCR results indicate that the sigma factors are active at the selected time points.

5.3.3 Identification of the regulon of each sigma factor

Once the time points were confirmed, the RNA samples extracted above (and used for the RT-qPCR experiment) were sent to the University of Nottingham DeepSeq facility for RNA Sequencing (a summary of the sequencing data is shown in the Appendix,

Table S11). Similar to what was described above, gene expression levels in each sigma factor mutant were compared to wild type. However, rather than testing one gene at a time, the entire transcriptome was analysed. The log_2-transformed expression of genes was analysed, and genes with log_2-fold change (log_2FC) >1 were considered to be upregulated, while genes with $log_2FC < -1$ were considered to be downregulated. Only genes with an adjusted p-value ≤0.05 were included in the analysis. The different biological replicates were compared using PCA analysis (Appendix, Figures S10 to S13). Whole transcriptome heatmaps can also be found in the Appendix, Figure S14.

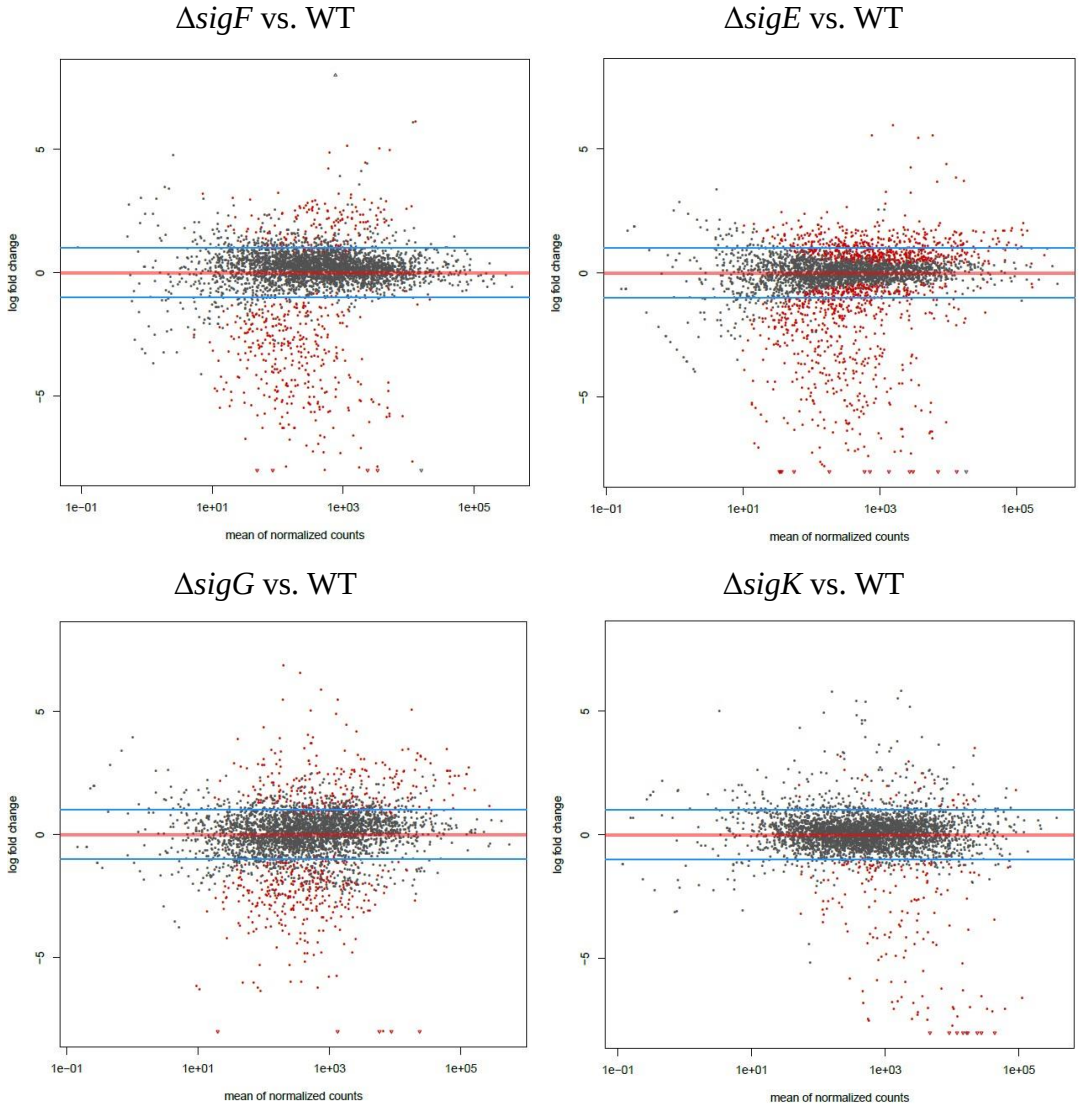


Figure 5.7: Overview of the differential gene expression between sigma factor mutants and wild type. Blue lines represent the threshold for significant differential expression: log_2FC >1 (positive fold change) indicates that the genes are upregulated, while $log_2FC < -1$ (negative fold change) indicates that genes are downregulated. Red dots represent genes that are significantly upregulated (log_2FC >1 and p-value ≤0.05) or downregulated ($log_2FC < -1$ and p-value ≤0.05). Genes represented in grey dots, on the other hand, represent genes with non-significant differential expression.

5.3.4 Regulation of expression and activity of the sigma factors

As mentioned in the introduction and above, some sigma factors affect the activity of others by transcriptional regulation, through post-translational modifications or by controlling the binding of inhibitors. In *C. sporogenes*, *sigF* and *sigE* appear not to be under the transcriptional regulation of any of the other sigma factors. The expression of the early stage sigma factors is possibly regulated by Spo0A, as is the case in *B. subtilis*[13]. Following asymmetrical division, SpoIIAA, SpoIIAB and SpoIIE activate σ^F in *B. subtilis*. The genes encoding these proteins are present in the *C. sporogenes* genome and appear to be essential for sporulation/germination, as shown in Chapters 4 and 6. These genes are not differentially expressed in any of the sigma factor mutants and are probably also regulated by Spo0A[13].

In the model organism *B. subtilis*, σ^E contains an inhibitory pro-sequence that is cleaved by SpoIIGA and SpoIIR (the mechanism is described in Chapter 6). In *C. sporogenes*, the gene encoding SpoIIGA is in an operon with *sigE*, is essential for sporulation (Chapter 4) and is probably regulated by Spo0A[13]. The gene encoding *spoIIR*, on the other hand, is regulated by σ^F in *B. subtilis*[60], *C. difficile*[147] and *C. sporogenes* (Table 5.1). Surprisingly, the *spoIIR* gene was also downregulated in the *C. sporogenes* Δ*sigG* strain. This result suggests that SpoIIR may have an additional role in late sporulation that has not been described in other species. Further studies are required to confirm the regulation of *spoIIR* by σ^G (such as RT-qPCR) and to determine its potential role in late sporulation.

The late stage sigma factor σ^G is transcriptionally regulated by σ^F in *B. subtilis*[34]. Intriguingly, *sigG* appears to be part of the σ^E regulon in *C. sporogenes* (Table 5.2). This is surprising, as σ^G is usually active in the forespore and σ^E in the mother cell. Nonetheless, it has been previously reported that σ^E activity is required for the transcription of *sigG* in *B. subtilis*, possibly as a means to delay σ^G activation[139]. While in other spore formers *sigG* is regulated by σ^F in the forespore[155, 156], *sigG* appears not to be significantly downregulated in *C. sporogenes* Δ*sigF* (Table 5.1). It is possible that *sigG* is in fact regulated by σ^F and that the results are not statistically significant due to variability between the different biological replicates (Appendix, Figure S10). The apparent lack of regulation of *sigG* by σ^F has been previously reported in *C. difficile*[147], and the authors suggest that *sigG* is under the control of two promoters: one recognised by σ^A and another recognised by σ^F and/or σ^G. The promoter

recognised by σ^A appears to be in front of the *spoIIGA-sigE* operon, which is immediately upstream of *sigG*; the promoter recognised by σ^F/σ^G appears to be in front of *sigG* itself[147]. In *C. sporogenes*, the *spoIIGA-sigE* operon is also upstream of *sigG*, so it is possible that the two-promoter hypothesis also applies in this species. However, further investigations are required to confirm how *sigG* is regulated in *C. sporogenes*. The activation of σ^G is mediated by SpoIIIA and SpoIIIJ in *B. subtilis*[140, 141]. The *spoIIIA* operon is also present in the *C. sporogenes* genome, where it appears to be regulated by σ^E (Table 5.2) and to be essential for spore formation (Chapter 4). The gene encoding SpoIIIJ, on the other hand, is absent from the *C. sporogenes* genome. A different protein might have a similar role in the signalling pathway to induce the activation of σ^G. Several putative *C. sporogenes* lipoproteins (*e.g.,* CLSPO_c01100, CLSPO_c03560, CLSPO_c03860) are good candidates for this role. Nevertheless, further investigations are required to elucidate this mechanism.

The mother cell-specific σ^E appears to regulate the expression of *sigK* in *C. sporogenes*, as the *sigK* gene is downregulated in the *sigF* and *sigE* mutants (Table 5.2). It is possible that the sigma factor is regulated only by σ^E, and that the down-regulation of the gene in the *sigF* mutant is a consequence of σ^E being partially dependent on σ^F. This hypothesis is supported by the fact that σ^F is active in the forespore and *sigK* is expressed in the mother cell. The absence of σ^K in *sigF* and *sigE* mutants has been reported in *C. difficile*[157]. The *sigK* gene is also positively regulated by SpoIIID and σ^K itself in *B. subtilis*[34] and *C. difficile*[147], but further studies are necessary to determine if the same occurs in *C. sporogenes*. The autoregulatory loop is also observed for σ^G in other spore formers[138, 147], but further evidence is required in *C. sporogenes*.

In *B. subtilis*, the *sigK* gene is interrupted by a 48 kb sequence named the *skin* element. Consequently, the activation of the sigma factor requires the removal of the *skin* element by SpoIVCA[34]. No *skin*-like element was found in the *C. sporogenes* genome, nor any *spoIVCA* homologue. This regulatory mechanism appears to be absent from *C. sporogenes*. Surprisingly, the genes encoding SpoIVFA and SpoIVFB, which are involved in pro-σ^K processing in *B. subtilis*, were also not found in the genome of *C. sporogenes*. On the other hand, *bofA* is present but does not appear to have an essential role in sporulation according to the TraDIS results (Chapter 4). The gene responsible for the coupling of σ^K activation to σ^G activity, *spoIVB*, is present in the *C. sporogenes* genome and appears to have a relevant role in sporulation (Chapter 4).

This gene appears to be part of the σ^G regulon in *C. sporogenes* (Table 5.3), which is in line with what has been observed in *B. subtilis*[34]. The full elucidation of the σ^K activation mechanism requires further investigations, as it appears to be divergent from other species.

5.3.4.1 The σ^F regulon

The analysis of the σ^F regulon of *C. sporogenes* revealed 39 genes downregulated in the *sigF* mutant but not in the *sigE*, *sigG* or *sigK* mutants. Some of these genes are presented in Table 5.1, and the complete σ^F regulon is presented in the Appendix, Table S12. Similar to what was observed in *B. subtilis*[58] and *C. difficile*[147], this is the smallest regulon of the four that were analysed.

A single-stranded DNA-binding protein (CLSPO_c04270) appears to be regulated by σ^F in *C. sporogenes* but by σ^E in *C. difficile*[147]. This protein might be involved in chromosome segregation during sporulation[158].

Other σ^F-dependent genes include a permease protein (CLSPO_c16540) and a Na^+ efflux pump (CLSPO_c19450). Similar membrane proteins are also regulated by σ^F in *C. difficile*[147]. Some proteins involved in cell wall metabolism are also σ^F-dependent in *C. difficile*[147] and *C. sporogenes* (Table 5.1), and the latter species includes a putative peptidoglycan binding protein (CLSPO_c23700).

Several genes encoding ethanolamine utilisation proteins were downregulated in the *sigF* mutant in *C. sporogenes* (CLSPO_c21190, CLSPO_c21230, CLSPO_c21240 and CLSPO_c21260) but upregulated in *C. difficile*[147]. Previous studies indicate that ethanolamine, a component of the phospholipid layer[159], can be used as a nutrient source during exponential growth[160] and that it is relevant for spore formation in yeast[161]. It can be hypothesised that the repression of the *eut* genes during sporulation in *C. difficile* might prevent the degradation of ethanolamine, which can then be used for membrane synthesis. It is unclear why those genes are positively regulated by σ^F in *C. sporogenes*.

Some other genes that are part of the σ^F regulon in other species (namely *B. subtilis* and *C. difficile*)[34, 147] are downregulated in the *C. sporogenes* Δ*sigF* strain, but also in other sigma factor mutants. These genes, which might be co-regulated by more than one sigma factor, include *spoIIR* and *sigG* (discussed above), *spoIVB* (involved in σ^K activation), *gpr* (involved in SASPs degradation) and *sleB* (involved in spore cortex hydrolysis). The downregulation of *gpr* in a *sigG* mutant has been observed in

C. difficile[157]. Interestingly, *gpr* and *spoIIR* appear not to be σ^F-dependent in *C. acetobutylicum*[155].

Table 5.1: Part of the σ^F regulon from *C. sporogenes*. Genes with log_2FC <-1 and adjusted p-value ≤0.05 were considered to be downregulated in the *sigF* mutant relative to wild type. * indicates genes with p-value >0.05.

Some genes from the σ^F regulon		
Gene ID	**Gene name/Description**	**Log_2FC (Δ*sigF*)**
CLSPO_c04270	*ssb*	-2.17
CLSPO_c16540	ABC transporter ATP-binding /permease protein	-1.77
CLSPO_c19450	Na^+-driven multidrug efflux pump	-1.09
CLSPO_c21190	ethanolamine utilization protein PduA/PduJ	-3.06
CLSPO_c21230	*eutQ*	-3.32
CLSPO_c21240	*eutN*	-3.28
CLSPO_c21260	*eutJ*	-3.42
CLSPO_c23700	putative peptidoglycan binding protein	-4.69
CLSPO_c26370	*spoVD2*	-2.31
Genes that might be co-regulated by σ^F and other sigma factors		
Gene ID	**Gene name**	**Log_2FC (Δ*sigF*)**	**Also downregulated in**
CLSPO_c01510	*spoIIR*	-1.19	Δ*sigG*
CLSPO_c02370	*sleB*	-4.99	Δ*sigE* and Δ*sigG*
CLSPO_c18660	*spoIVB*	-3.88	Δ*sigE* and Δ*sigG*
CLSPO_c26270	*sigG*	-0.89*	Δ*sigE*
CLSPO_c30570	*gpr*	-1.01*	Δ*sigG*

Notably, *spoIIQ* is part of the σ^F regulon in *B. subtilis*[58] and *C. difficile*[147] but could not be found in the *C. sporogenes* genome. The lack of SpoIIQ orthologs in some other clostridia has also been reported[162]. Other proteins might take its role in engulfment and in the formation of the channel between forespore and mother cell. Also surprising was the downregulation of *spoVD2* (CLSPO_c26370) in the *sigF* mutant, as this gene appears to be regulated by the late stage mother-cell specific σ^K in other species[34] and

is involved in spore maturation[163]. In addition, seven genes downregulated in the *sigF* mutant encode hypothetical proteins.

5.3.4.2 The σ^E regulon

The σ^E regulon contains 238 genes that are downregulated in the *sigE* mutant but not in the *sigG* or *sigK* mutants. Some of these genes are listed in Table 5.2 and the complete σ^E regulon is presented in the Appendix, Table S13. This appears to be the largest operon of the four. Interestingly, σ^E activity appears to be partially independent from σ^F, as 92 genes appear to be regulated exclusively by σ^E. This is in line with previous reports showing that *C. difficile* lacking σ^F is still able to process σ^E, although the amount of processed σ^E is highly reduced[147, 157]. On the contrary, σ^F appears to be required for complete processing of pro-σ^E in other species, such as *B. subtilis*[164] and *C. perfringens*[156]. Nevertheless, 146 genes are downregulated in both the *sigF* and *sigE* mutants. This indicates that the activity of σ^E is at least partially dependent on σ^F, possibly through a mechanism similar to *B. subtilis* (involving SpoIIR and SpoIIGA, as discussed above). It is possible that *spoIIR* expression is not only regulated by σ^F but also by Spo0A, similar to what was observed in *C. difficile*[147]. The low levels of SpoIIR present in the *sigF* mutant would be responsible for the processing of some pro-σ^E and consequent partial activity of σ^E. Further studies are required to confirm this in *C. sporogenes*.

It was mentioned in the introduction to this chapter that genes involved in engulfment are σ^E-dependent in other spore formers[58, 147]. That also seems to happen in *C. sporogenes*, since the engulfment-related genes *spoIID* (CLSPO_c01900) and *spoIIM* (CLSPO_c26370) are downregulated in the *sigE* mutant. A gene encoding a stage II sporulation protein (CLSPO_c30560) from *C. sporogenes* is identical to *spoIIP* from the *C. sporogenes* strain DSM 795 (blastn; 100% identity; e-value = 0). This gene is also involved in engulfment and it also appears to be positively regulated by σ^E.

Also in line with other spore formers[34, 147, 157], the *C. sporogenes* σ^E appears to regulate genes involved in coat formation, such as *spoIVA* (CLSPO_c26120), *yqfC* (CLSPO_c30420) and other putative spore coat proteins (CLSPO_c01400, CLSPO_c17960 and CLSPO_c30310)[123, 126]. Furthermore, this regulon includes proteins that play a role in cortex formation, such as a N-acetylmuramoyl-L-alanine amidase (CLSPO_c36050) and a polysaccharide deacetylase (CLSPO_c32630)[165].

Table 5.2: Part of the σ^E regulon from *C. sporogenes*. Genes with log_2FC <-1 and adjusted p-value ≤0.05 were downregulated in the *sigE* mutant relative to wild type. ** indicates that the CLSPO_c30560 is annotated as stage II sporulation protein but is identical to *spoIIP* from another *C. sporogenes* strain.

Some genes from the σ^E regulon

Gene ID	Gene name/Description	Log_2FC (Δ*sigE*)
CLSPO_c00970	*spmA*	-4.02
CLSPO_c00980	*spmB*	-4.04
CLSPO_c01400	putative spore coat protein	-8.13
CLSPO_c01900	*spoIID*	-5.82
CLSPO_c01920	*spoIIID*	-1.66
CLSPO_c17960	putative spore coat protein	-3.60
CLSPO_c18810	*spoIIIAH*	-6.04
CLSPO_c18820	*spoIIIAG*	-4.61
CLSPO_c18830	*spoIIIAF*	-4.66
CLSPO_c18840	*spoIIIAE*	-4.45
CLSPO_c18850	*spoIIIAD*	-5.31
CLSPO_c18860	*spoIIIAC*	-4.60
CLSPO_c18870	*spoIIIAB*	-5.03
CLSPO_c18880	*spoIIIAA*	-5.46
CLSPO_c26120	*spoIVA*	-6.38
CLSPO_c26270	*sigG*	-1.84
CLSPO_c26360	*sigK*	-2.96
CLSPO_c30310	spore coat protein-like protein	-7.12
CLSPO_c30420	*yqfC*	-4.39
CLSPO_c30560	*spoIIP***	-5.23
CLSPO_c32630	polysaccharide deacetylase	-4.39
CLSPO_c36050	N-acetylmuramoyl-L-alanine amidase	-3.57

Genes that might be co-regulated by σ^E and other sigma factors

Gene ID	Gene name	Log_2FC (Δ*sigE*)	Also downregulated in
CLSPO_c26370	*spoIIM*	-2.57	Δ*sigG*

The operon *spmA-spmB* (CLSPO_c00970 and CLSPO_c00980) plays a role in spore dehydration and, consequently, affects the spore's heat resistance[166]; this operon is

downregulated in *sigE* mutants in *C. sporogenes* (Table 5.2) and in *C. difficile*[147].Some σ^E-dependent genes also have regulatory functions, as they are involved in σ^K transcription (*sigK* - CLSPO_c26360) and in gene regulation (*spoIIID* - CLSPO_c01920). As already discussed, σ^E also regulates the expression of *sigG* (CLSPO_c26270) and of the *spoIIIA* operon (CLSPO_c18810 to CLSPO_c18880), involved in the formation of the feeding tube and σ^G activation. Finally, 40 genes regulated by σ^E have unknown function.

5.3.4.3 The σ^G regulon

The σ^G regulon contains 153 genes downregulated in the *sigG* mutant but not in the *sigK* mutant. Some of these genes are listed in Table 5.3, and the complete σ^G regulon is presented in the Appendix, Table S14. It was described above that the σ^E-dependent *spoIIIA* operon is involved in σ^G activation[141], suggesting that σ^G is also σ^E-dependent. Accordingly, 28 genes are downregulated in both the *sigE* and the *sigG* mutants. However, 70 genes are downregulated exclusively in the σ^G library. This suggests that the activity of σ^G is partially independent from σ^E, as previously reported for *C. difficile*[157]. Another explanation is the different timing of RNA extractions between the *sigG* (20 hours after inoculation) and the *sigF*/*sigE* mutants (17 hours after inoculation), as hypothesised by Saujet *et al.*[147]. It is possible that the earlier time point was not optimal for σ^G activity, as suggested by the RT-PCR experiments (Figure 5.5; bands corresponding to the *spoIVB* gene - σ^G dependent - are slightly fainter at 17 hours than at 20 hours). However, the RT-PCR results are not sensitive enough to demonstrate a significant difference in σ^G activity and further experiments are required to determine if σ^G is in fact partially independent from σ^E. It is also worth mentioning that the apparent activity of σ^G in the absence of σ^E supports the hypothesis that another sigma factor other than σ^E regulates the transcription of *sigG*.

Ten genes are downregulated in both the *sigF* and the *sigG* mutants. One explanation for this is the transcriptional regulation of *sigG* by σ^F, a hypothesis that was described above. According to this hypothesis, no σ^G would be synthesised in a *sigF* mutant, leading to the downregulation of σ^G-controlled genes in the *sigF* mutant. A second hypothesis is that some genes in the σ^F regulon control the activity of σ^G, such as *csfB* orthologs[147]. Finally, it is also possible that those genes are co-regulated by both sigma factors, as previously reported for *B. subtilis*[34] and *C. difficile*[147]. The fact that 79 genes were regulated by σ^F, σ^E and σ^G might be a result of the regulatory mechanisms

suggested above: transcriptional regulation of *sigG* by σ^E and possibly σ^F; post-translational activation of σ^E by σ^F; and/or post-translational activation of σ^G by σ^E.

One of the genes that is downregulated in the three mutants (*sigF*, *sigE* and *sigG*) is *spoVT* (CLSPO_c37200), which in turn regulates other genes in the σ^G regulon. The regulation of this gene by σ^F and σ^G has been previously reported in *B. subtilis*[34] and *C. difficile*[147, 157]. Saujet *et al.*[147] identified σ^F and σ^G consensus elements in the promoter of *spoVT*, indicating a co-regulation of the gene. However, a σ^E element was not mentioned despite the fact that the gene was downregulated in the *sigE* mutant[147], indicating that this is probably a result of the partial dependency of σ^E activity on σ^F. In addition to spoVT, σ^G also appears to regulate the expression of SpoIVB (CLSPO_c18660), which is involved in the activation of σ^K [34].

As indicated in Chapter 1, DNA from the spore needs to be protected from external threats. This is achieved with the help of small acid-soluble proteins (SASPs)[1], and the genes encoding those proteins are regulated by σ^G (*sspH1* - CLSPO_c08070; CLSPO_c17800; CLSPO_c17810 and CLSPO_c31420). Interestingly, the gene encoding another SASP (*sspH2* - CLSPO_c12520) appears to be regulated by both σ^G and σ^K. In addition to SASPs, dipicolinic acid (DPA) also plays a role in DNA protection. DPA is usually synthesised in the mother cell[1] and the *spoVA* operon is responsible for its uptake by the forespore[61]. In *C. sporogenes*, several genes encoding SpoVA proteins appear to be positively regulated by σ^G and are listed in Table 5.3. The *ssp* and *spoVA* genes are also σ^G-dependent in *B. subtilis*[34] and *C. difficile*[147, 157].

Several proteins involved in the formation of the spore cortex also appear to be regulated by σ^G in *C. difficile*[147, 157] and *C. sporogenes* (Table 5.3), such as D-alanyl-D-alanine carboxypeptidase (encoded by *dacF*). Other σ^G-dependent genes in *B. subtilis*[34], *C. difficile*[147] and *C. sporogenes* are germinant receptors. In *C. sporogenes*, these include *gerKA1* (CLSPO_c02990), *gerKA2* (CLSPO_c19770), *gerAB* (CLSPO_c19780) and others (Table 5.3). Several genes encoding hypothetical proteins (52 genes) are also downregulated in the *sigG* mutant.

Table 5.3: Part of the σ^G regulon from *C. sporogenes*. Genes with $\log_2$FC <-1 and adjusted p-value ≤0.05 were considered to be downregulated in the *sigG* mutant relative to wild type.

Some genes from the σ^G regulon

Gene ID	Gene name/Description	$\log_2$FC (Δ*sigG*)
CLSPO_c01480	*gerKA4*	-3.05
CLSPO_c01490	Germination protein	-2.23

Table 5.3: Part of the σ^G regulon from *C. sporogenes* (continuation).

Some genes from the σ^G regulon		
Gene ID	**Gene name/Description**	**Log_2FC (Δ*sigG*)**
CLSPO_c01500	Germination protein; GerC family	-2.43
CLSPO_c01510	*spoIIR*	-1.89
CLSPO_c02970	Germination protein; Ger(X)C family	-2.29
CLSPO_c02980	Germination protein	-2.81
CLSPO_c02990	*gerKA1*	-2.32
CLSPO_c08070	*sspH1*	-6.00
CLSPO_c08390	Germination protein	-2.29
CLSPO_c15920	*spoVAF*	-1.91
CLSPO_c17800	SASP alpha	-10.17
CLSPO_c17810	SASP beta	-8.00
CLSPO_c18520	*dacF2*	-2.83
CLSPO_c18660	*spoIVB*	-3.37
CLSPO_c19770	*gerKA2*	-1.73
CLSPO_c19780	*gerAB*	-2.67
CLSPO_c19790	*gerKA5*	-2.14
CLSPO_c19820	Spore germination protein	-2.26
CLSPO_c26370	*spoIIM*	-2.13
CLSPO_c29440	*gerKA6*	-3.51
CLSPO_c29450	Germination protein	-3.46
CLSPO_c30570	*gpr*	-1.40
CLSPO_c31420	SASP alpha/beta type	-8.03
CLSPO_c31780	SpoVA family protein	-1.87
CLSPO_c31790	*spoVAD*	-2.38
CLSPO_c31800	*spoVAC*	-3.65
CLSPO_c37200	*spoVT*	-4.80

Genes that might be co-regulated by σ^G and other sigma factors			
Gene ID	**Gene name**	**Log_2FC (Δ*sigG*)**	**Also downregulated in**
CLSPO_c12520	*sspH2*	-4.21	Δ*sigK*

5.3.4.4 The σ^K regulon

The σ^K regulon contains 140 genes, 38 of which are regulated exclusively by σ^K. Given that *sigK* is transcriptionally regulated by σ^E, it was expected that all the genes in the σ^K regulon would also be differentially expressed in the *sigE* mutant. Similar to what was discussed for σ^G, it is likely that the time point for RNA extraction of the *sigE* mutant was not optimal for σ^K activity. The RT-PCR results seem to support this hypothesis, as the intensity of the bands corresponding to *cotA* and *cotI* is slightly lower at 17 hours than at 24 hours. Further experiments are required to confirm this observation. Also in agreement with this hypothesis is the fact that three genes are regulated by σ^E and σ^K, and six genes by σ^F, σ^E and σ^K. Additionally, fifteen genes are regulated by σ^G and σ^K, five genes by σ^F, σ^G and σ^K, and seven genes by σ^E, σ^G and σ^K, indicating that σ^K activity is partially dependent on σ^G. These results also reflect the fact that σ^G activity is partially dependent on σ^F and σ^E, as described above. Finally, 66 genes appear to be regulated by all the sigma factors (σ^F, σ^E, σ^G and σ^K), highlighting the fact that σ^K activity can be affected by all the previous sigma factors. A partial list of the σ^K-dependent genes is provided in Table 5.4 and the complete list can be found in the Appendix, Table S15.

Similar to observations in *B. subtilis*[58] and *C. difficile*[147], σ^K appears to regulate several genes involved in the formation of the spore coat in *C. sporogenes*, such as *cotA* (CLSPO_c08700), *cotI* (CLSPO_c01380), *cotS* (CLSPO_c01350) and the *cotJ* operon (CLSPO_c12440 to CLSPO_c12460)[124, 125, 167]. Furthermore, an exosporium protein (CLSPO_c32280) also appears to be regulated by σ^K. This suggests the presence of an exosporium in the *C. sporogenes* strain NCIMB 10696, similar to what was observed in *C. sporogenes* ATCC 15579[168]. Nevertheless, the observation of spores of *C. sporogenes* NCIMB 10696 using transmission electron microscopy (TEM) is required to confirm the presence of the exosporium in this strain.

In addition, a N-acetylmuramoyl-L-alanine amidase (CLSPO_c22840) gene was downregulated in the *sigK* mutant and might be involved in the lysis of the mother cell following spore maturation[169], similar to previous observations in *C. difficile*[147].

The synthesis of DPA, involved in the protection of the spore DNA, usually occurs in the mother cell under the control of σ^K [1, 147]. Surprisingly, no genes involved in DPA synthesis were found in the *C. sporogenes* genome. However, since the σ^K regulon includes 61 genes encoding hypothetical proteins, it is possible that one of these is

responsible for the production of DPA. Finally, σ^K appears to induce the expression of an inhibitor of σ^G activity (CLSPO_c00800).

The results obtained with the RNA-Seq experiments indicate that σ^K only regulates the late stages of sporulation, which is in line with *B. subtilis*[34] and *C. difficile*[157] but not with *C. perfringens*[145] or *C. botulinum*[146]. In the two latter species, σ^K appears to be involved in regulation of early events during sporulation.

Table 5.4: Part of the σ^K regulon from *C. sporogenes*. Genes with log_2FC <-1 and adjusted p-value ≤0.05 were considered to be downregulated in the *sigK* mutant relative to wild type.

Some genes from the σ^K regulon

Gene ID	Gene name/Description	Log_2FC (Δ*sigK*)
CLSPO_c00800	inhibitor of sigma-G Gin	-2.08
CLSPO_c01350	*cotS*	-4.15
CLSPO_c01380	*cotI*	-1.71
CLSPO_c08700	*cotA*	-4.83
CLSPO_c12440	*cotJA*	-7.28
CLSPO_c12450	*cotJB*	-7.00
CLSPO_c12460	*cotJC*	-6.96
CLSPO_c22840	N-acetylmuramoyl-L-alanine amidase	-2.63
CLSPO_c32280	exosporium protein	-7.12

5.3.4.5 Other regulatory proteins

As mentioned above, SpoIIQ, RsfA, SplA and SpoVT appear to further regulate the σ^F and σ^G regulons in the forespore. The same is observed in the mother cell, where SpoIIID and GerE provide additional regulation to the σ^E and σ^K regulons[34, 130]. Similar to *C. difficile*[147], only SpoVT and SpoIIID could be found in the *C. sporogenes* genome. As mentioned above, *spoVT* is downregulated in both the *sigF* and *sigG* mutants (Table 5.3), indicating that this protein is present in the forespore. It is therefore possible that SpoVT also has a regulatory role in the forespore line of expression of *C. sporogenes*. Likewise, *spoIIID* is regulated by σ^E (Table 5.2) and could regulate the mother cell-specific genes.

Further studies are required to confirm the role of these proteins. For instance, Saujet *et al.*[147] mutated *spoVT*, and then tested the sporulation ability of the mutant to sporulate and observed it under a phase-contrast microscope. The authors also used

RT-qPCR to determine expression levels of selected genes in the mutant in order to confirm the regulatory role of SpoVT. On the other hand, the entire regulon of the *spoIIID* mutant was determined[147]. Similar approaches could be used in *C. sporogenes*.

5.4 Discussion

The results obtained in this chapter indicate that the regulation of sporulation in *C. sporogenes* presents some similarities to the model organism *B. subtilis*, but also some differences. Deviations from the model sporulation pathway have been reported for other *Clostridium* species[145, 157], highlighting the need to characterise each species individually. Furthermore, even though the sporulation and germination mechanisms in *C. sporogenes* are closer to other *Clostridium* species than to *B. subtilis* (*e.g.*, the apparent partially independent activity of the sigma factors), there are differences even between the different clostridia. For instance, a *sigF* mutant in *C. perfringens* is unable to produce any of the other sigma factors[156], while the same mutant in *C. difficile* is able to partially process σ^E and produce some σ^G [157]. In addition, the activation of σ^K in *C. perfringens* involves the removal of a pro-peptide[145] rather than the removal of the *skin* element observed in *C. difficile*[170]. Finally, σ^K appears to be active during early stages of sporulation in *C. perfringens*[145] and *C. botulinum*[146].

The results herein presented suggest that the activity of σ^E in *C. sporogenes* is partially dependent on σ^F. In addition, the transcription of *sigG* is regulated by σ^E and possibly σ^F, while the activity of σ^G is also partially dependent on σ^F and σ^E. Finally, the transcription of *sigK* is regulated by σ^E, and the activity of σ^K is partially dependent on σ^F, σ^E and σ^K. The partially independent activity between the different sigma factors might not be real, and several factors might contribute to this result:

i) Sporulation is frequently not synchronised in all the cells[157], potentially causing variability between different cultures/biological replicates (see Appendix, Figures S10 to S13). Hence, it is possible that some results obtained in the present study were not statistically significant due to that variability, causing some "false negatives".

ii) The regulation of the genes involved in sporulation is complex, and not all genes in a certain regulon are expressed at the same time. Several other proteins are known to regulate gene expression in addition to the sigma factors, such as SpoIIID and SpoVT, and these proteins might repress the expression of some genes when they are no longer needed[12]. It is possible that some genes were no longer being transcribed in wild type at the selected time points, so differential expression between mutants and wild type was no longer detected.

iii) The early sigma factors σ^F and σ^E are, as expected, active at earlier time points than σ^G and σ^K. This is evidenced by the RT-PCR results (Figure 5.5). Consequently, it is possible that some genes regulated by σ^G and σ^K were not downregulated in the *sigF* and *sigE* mutants (respectively) due to that time-specific activation.

Despite these issues, the RNA-Seq results shed some light on the mechanisms of regulation of sigma factors in *C. sporogenes*. Indeed, as mentioned above, the partially independent activity of some sigma factors deviates from the *B. subtilis* model[34] but is in accordance with some of the regulatory mechanisms reported for *C. difficile*[157]. Overall, it seems that the transcription and activation of *sigF* and *sigE* are similar to that of *B. subtilis*, although σ^F appears not to be absolutely essential for σ^E activation. The transcription of *sigG* is regulated by σ^F and σ^E, and the activation of σ^G is regulated by σ^E, possibly with the involvement of the "feeding tube" and the *spoIIIA* operon. The main difference relative to other species is the absence of SpoIIIJ – it is possible that another protein with similar functions takes its role in the signalling pathway, but further studies are necessary to elucidate this mechanism. Finally, the transcription of *sigK* follows the *B. subtilis* model and is σ^E-dependent. However, σ^K activation mechanism appears to be different: i) the *sigK* gene does not contain the *skin* element; and ii) some of the proteins usually involved in the processing of pro-σ^K could not be found in the genome of *C. sporogenes*, while others do not appear to have a relevant role in the sporulation process (TraDIS results – Chapter 4). Only the σ^G-dependent *spoIVB* appears to be involved in σ^K activation in *C. sporogenes*. Again, it is possible that other proteins are involved in this process but do not share homology to SpoIVFA and SpoIVFB. Further studies would allow the further elucidation of these mechanisms, such as RT-qPCR to confirm the transcriptional regulation of *sigG* and *sigK*, and Western blot to evaluate how σ^E and σ^K are processed in the different mutants.

Regarding the regulons of each sigma factor, the σ^F-dependent genes appear to show regulatory activities and to be involved in spore morphogenesis through membrane and cell wall metabolism. The mother cell-specific σ^E appears to regulate engulfment, the initial stages of coat formation and the translation/activation of the late stage sigma factors. After engulfment, σ^G is responsible for the protection of the spore DNA, and it regulates germinant receptors and cortex-related proteins. Finally, the σ^K regulon is involved in coat formation, spore maturation and the lysis of the mother cell.

Many genes were identified using the RNA-Seq technique, and not all of them were discussed in this chapter. The presence of genes of unknown function in the sigma factor regulons highlights the value of the approach taken and its ability to identify genes that have not previously been linked to sporulation. Overall, the results herein presented shed light on the sporulation process in *C. sporogenes*, namely on the regulatory mechanisms involving the four main sigma factors. Some of these mechanisms resemble those in other spore formers, but some differences were also identified, highlighting the need to study each species individually. This work answers some questions but it raises others, paving the way to further work that will allow the further elucidation of the regulatory mechanisms that control sporulation and germination in *C. sporogenes*.

Chapter 6: Characterisation of genes involved in sporulation and/or germination in *C. sporogenes*

6

6.1 Background

6.1.1 Brief overview of the selected genes

The high-throughput techniques used in this study allowed the identification of several genes involved in or even essential for sporulation or germination. The high number of genes identified and the fact that some of them have unknown functions highlights the power of high-throughput approaches. Despite these advantages, the data obtained is potentially compromised by the occurrence of false positives and/or polar effects. Consequently, validation of the results using other targeted and more sensitive approaches is usually required. In summary, the forward genetics tools allow the identification of potentially interesting genes and reverse genetics tools enable the further study of those genes.

With the purpose of validating the TraDIS and RNA-Seq libraries, some of the genes identified in the previous chapters were selected for further characterisation, and their contribution to forming viable spores was evaluated. A brief description of these genes and their functions is provided below.

6.1.1.1 Sigma factors

The genes encoding the sigma factors that regulate sporulation (σ^F and σ^G in the forespore; σ^E and σ^K in the mother cell) were targeted using the RiboCas system in order to generate the RNA-Seq libraries (Chapter 5). As these genes appear to be essential for sporulation/germination according to the TraDIS library (Chapter 4), the reversion of the genes to the original locus and further characterisation of these mutants seemed relevant. The role of these genes in the sporulation cascade has already been described (Chapters 1 and 5).

6.1.1.2 CLSPO_c31830: *spoIIAA*

The gene *spoIIAA* appears to have an important role in sporulation in *C. sporogenes* (Chapter 4). In *B. subtilis*, the gene *spoIIAA* is part of the operon *spoIIAA-spoIIAB-sigF* (also named *spoIIA* operon), which is positively regulated by Spo0A[13]. As previously mentioned (Chapter 5), *sigF* encodes the sigma factor σ^F, which regulates early sporulation in the forespore[34]. However, being regulated by Spo0A, σ^F is present

in the cells before the forespore compartment is formed and separated from the mother cell[135]. Consequently, this sigma factor should be held inactive until the completion of asymmetric division and septum formation to ensure the temporal- and spatial-specific regulation of the σ^F regulon[171].

σ^F inactivation is mediated by the anti-sigma factor SpoIIAB[135], which binds to σ^F forming the complex SpoIIAB$_2$-σ^F (2 molecules of SpoIIAB per molecule of σ^F)[172]. While in this complex, the core RNA polymerase binding site of σ^F is blocked by SpoIIAB, inhibiting transcription[172] (Figure 6.1). Indeed, it has been shown that the transcription of genes regulated by σ^F is inhibited by the presence of SpoIIAB[171].

In addition to its role as an anti-sigma factor, SpoIIAB also functions as a serine kinase that maintains SpoIIAA in the phosphorylated state in the pre-divisional cell[135]. While phosphorylated, the anti-anti-sigma factor SpoIIAA is incapable of binding to the SpoIIAB$_2$-σ^F complex. Following asymmetric division, SpoIIE is responsible for the dephosphorylation of SpoIIAA allowing its binding to SpoIIAB and the release of σ^F [136, 173] (Figure 6.1). The activation of σ^F is, therefore, dependent on the rate of SpoIIAA dephosphorylation[174] and, after a certain threshold, is essentially irreversible[173]. Furthermore, the complex formed by SpoIIAA and SpoIIAB is stable and "long-lived", preventing SpoIIAB from binding to σ^F again[173].

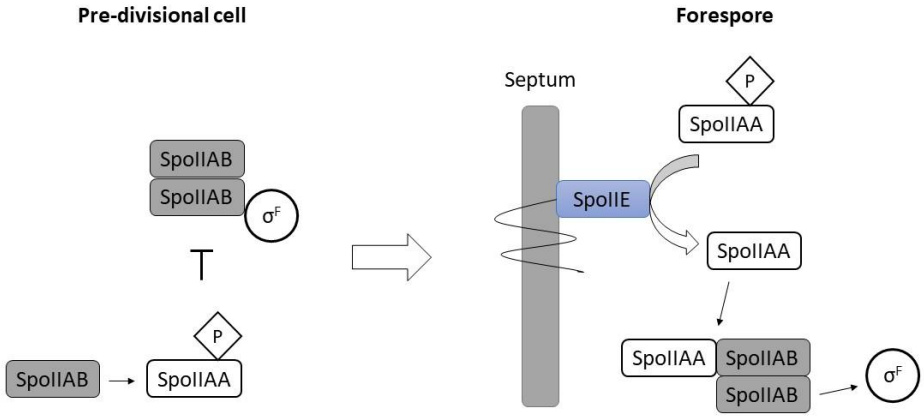


Figure 6.1: Schematic representation of the σ^F activation mechanism. In the pre-divisional cell and mother cell, σ^F activity is inhibited by SpoIIAB, which also keeps SpoIIAA phosphorylated. The complex is formed by two molecules of SpoIIAB per molecule of σ^F, but the latter only interacts directly with one of the SpoIIAB molecules[172]. After septum formation, SpoIIE dephosphorylates SpoIIAA, which then interacts with SpoIIAB, releasing σ^F.

6.1.1.3 CLSPO_c37040: *spoIIE*

Similar to *spoIIAA*, *spoIIE* appears to have a relevant role in the sporulation cascade of *C. sporogenes* (Chapter 4) due to its involvement in the activation of σ^F. As

described above, SpoIIE is a serine phosphatase that is responsible for the dephosphorylation of SpoIIAA, which then binds to the $SpoIIAB_2$-σ^F complex and releases σ^F [136, 175].

The SpoIIE phosphatase is a membrane protein[176], and it is located in the septum that separates the forespore from the mother cell[177] (Figure 6.1). The transmembrane segments in the N-terminal of SpoIIE appear to be relevant for its localisation, while the C-terminal contains a PP2C-like domain that is responsible for the phosphatase activity[176, 178]. The central domain of SpoIIE appears to interact with FtsZ, a GTPase that mediates the localisation of SpoIIE to the poles of the cell[176, 179, 180]. This is relevant for the localisation of SpoIIE in the septum.

The activation of σ^F exclusively in the forespore requires SpoIIE activity to be mainly confined to this compartment of the developing spore. Some studies propose that SpoIIE has a structural role in the septum formation and accumulates mainly on the forespore side, which would explain the forespore-specific activity of σ^F, as well as its dependency on septum formation[175, 181]. However, other studies have shown that SpoIIE accumulates in both the forespore and the mother cell, indicating that other mechanisms must be responsible for the compartmentalisation of σ^F activity[182].

Arigoni *et al.*[176] hypothesise that a SpoIIE-inhibitor ensures that the phosphatase activity occurs only in the forespore after septum formation, explaining the temporal and spatial-specific activation of σ^F. The authors have shown that *spoIIE* mutants lacking the transmembrane domain (and therefore located in the cytoplasm of both forespore and mother cell) can still promote compartment-specific activity of σ^F and produce spores (albeit in only 40% of the cells). The presence of an inhibitor would explain why SpoIIE doesn't activate σ^F in the mother cell[176]. This hypothesis is supported by another study reporting that SpoIIE contains a site in the central domain that regulates its phosphatase activity. This regulatory site also appears to be responsible for the important coupling of σ^F activation to septum formation[183]. Moreover, following septum formation, the forespore contains only one third of the chromosome (the origin-proximal part). The presence of a SpoIIE inhibitor near the terminus of the chromosome would repress SpoIIE activity in the mother cell (but not in the forespore) immediately after septum formation[38].

Finally, another explanation involving the transient genetic asymmetry has been proposed. The *spoIIA* operon is located at the terminus of the chromosome, and consequently *spoIIAA*, *spoIIAB* and *sigF* are not expressed in the forespore during a

short period following septum formation. This short period appears to be enough to facilitate σ^F activation due to the degradation of SpoIIAB by the protease complex ClpCP[34, 184].

Several different mechanisms have been proposed for the compartment specific activity of σ^F and its coupling to asymmetric division. It is likely that some of these mechanisms overlap to ensure the correct compartment- and time-specific activation of the sigma factor, which appears to be essential for proper spore formation[34].

6.1.1.4 CLSPO_c01510: *spoIIR*

Similar to what is observed with σ^F, σ^E is present in the pre-divisional cell and is not immediately active following transcription. While σ^F is sequestered in the $SpoIIAB_2$-σ^F complex[135], σ^E contains a pro-sequence (27 amino acids) that inhibits its activity (the precursor is called pro-σ^E)[185]. The post-translational cleavage of the pro-sequence is mediated by the aspartic protease SpoIIGA, which contains a trans-membrane domain at the N-terminal and a cytoplasmic domain at the C-terminal[137, 186] (Figure 6.2). The *spoIIGA* gene forms an operon with *sigE*[164], which means that σ^E should become active upon transcription (in the pre-divisional cell). Indeed, Stragier *et al.*[137] showed that *spoIID* (regulated by σ^E) is expressed in vegetative cells following induction of the *spoIIGA-sigE* operon, although its expression levels are significantly reduced when comparing to sporulating cells.

Nevertheless, it appears that SpoIIGA alone is not able to activate σ^E during sporulation[137]. This is because the processing of pro-σ^E also requires SpoIIR, a protein regulated by σ^F and therefore expressed in the forespore[60, 187]. Indeed, the gene encoding SpoIIR seems to be relevant for sporulation in *C. sporogenes* (Chapter 4).

After being expressed, SpoIIR appears to be secreted to the periplasm[186] due to the presence of a signal peptide[60, 187]. It has been hypothesised that once SpoIIR reaches the periplasm, it interacts with the membrane-associated domain of SpoIIGA and induces a conformational change at the C-terminal that activates its protease activity[186, 188]. Moreover, it seems that not only SpoIIGA but also pro-σ^E are associated to the membrane[57, 186]. It has been suggested that pro-σ^E is cleaved at the membrane by the activated SpoIIGA dimer, and that σ^E is released to the cytoplasm of the mother cell[186] (Figure 6.2). The premature activation of σ^E appears to have a negative impact on sporulation efficiency[57], which highlights the importance of SpoIIR in the sporulation process. Furthermore, the location of *spoIIR* in the chromosome is also relevant for the

correct timing of σ^E activation. The gene is located near the origin of the chromosome, ensuring its expression in the forespore immediately after septum formation. If *spoIIR* was located at the terminus, its expression would be delayed until SpoIIIE transferred the rest of the chromosome to the forespore[189].

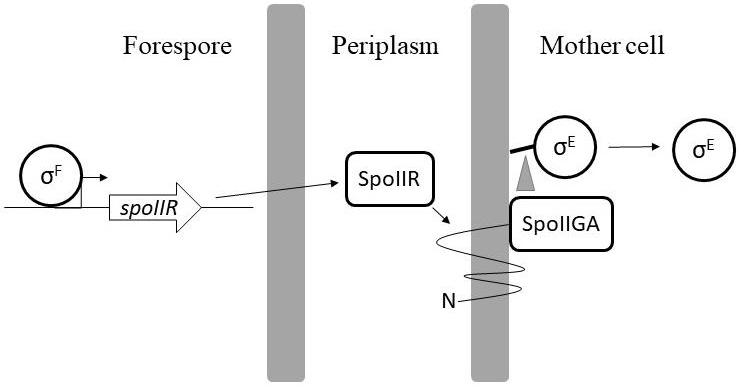


Figure 6.2: Schematic representation of the σ^E activation mechanism. The expression of *spoIIR* is regulated by σ^F in the forespore. SpoIIR is then secreted to the periplasm, where it interacts with the membrane-associated domain of SpoIIGA. This interaction activates the cytoplasmic protease domain of SpoIIGA, which is able to cleave the also membrane-bound pro-σ^E. The active σ^E is finally released to the mother cell cytoplasm. Adapted from Imamura *et al.*[186].

The SpoIIR-SpoIIGA interaction explains the coupling of σ^E activation to σ^F activity and septum formation. However, it does not elucidate the mother cell specific activity of the former. Several hypotheses have been proposed for the compartmentalisation of σ^E activity:

i) SpoIIGA is more abundant in the mother cell membrane than in the forespore, which means that SpoIIR (present in low amounts) is more likely to activate SpoIIGA from the mother cell once the latter is anchored in the septum[190, 191];

ii) σ^E appears to be inhibited in the forespore by CsfB and by another factor produced by σ^E itself in the mother cell[57, 190];

iii) Transcription of the *spoIIGA-sigE* operon is higher in the mother cell than in the forespore following septum formation due to increased Spo0A activity[57].

These mechanisms might co-exist to ensure the activation of σ^E only in the mother cell and/or the maintenance of its compartment-specific activity. Nevertheless, the activation of σ^E in the forespore appears not to have a big impact on the sporulation

ability of *B. subtilis*, possibly because the σ^E regulon is not fully functional in the forespore[190].

6.1.1.5 CLSPO_c01480: *gerKA4*

The gene *gerKA4* appears to encode a germinant receptor in *C. sporogenes* that has a relevant role in germination (Chapter 4). The spore is able to protect the cells under adverse conditions. However, when the conditions are favourable again, the spore is able to germinate into a functional vegetative cell. Several receptors in the inner membrane of *B. subtilis* spores are able to recognise extracellular signals and trigger germination. The nutrient receptors encoded by the tricistronic operons *gerA, gerB* and *gerK* recognise L-alanine (GerA) and a mixture of asparagine, glucose, fructose and K^+ known as AGFK (GerB and GerK)[5]. The *gerK* operon appears to contain three genes: *gerKA*, *gerKC* and *gerKB*[192]. The GerKA protein presents some homology with GerAA and GerBA, namely in a putative C-terminal transmembrane domain[192]. Although some *ger* operons have been identified in clostridia, the structures of the operons are variable and different from *Bacillus*[65].

6.1.1.6 CLSPO_c22010: M24 family metallopeptidase (APP)

A gene encoding an M24 family metallopeptidase (CLSPO_c22010) appears to be essential for germination in *C. sporogenes* (Chapter 4). The protein encoded by the CLSPO_c22010 gene is small (71 aa), and an analysis of the conserved domains indicates that it is a member of the APP_MetAP super family (Interval: 2-60; e-value: $8.25e^{-14}$), which comprises aminopeptidase P, aminopeptidase M, and prolidase[102]. A BLASTn search revealed that the gene is homologous to a Xaa-Pro aminopeptidase from *C. botulinum* strain F160 (95.27% identity, e-value: $4e^{-68}$)[193]. It is also known that the M24 family includes methionyl aminopeptidase, X-Pro dipeptidase and aminopeptidase P[194].

Altogether, these findings indicate that this protein functions as an aminopeptidase P (APP), a cytosolic exopeptidase that is responsible for the removal of the amino acid at the N-terminal of a peptide (long or short) if the following amino acid is a proline[195, 196]. Aminopeptidase P is a homotetramer that requires the presence of a metal ion in the active site to function[194]. Different metals have been linked to the activation of APP, namely zinc, manganese and cobalt[196]. Aminopeptidases are often involved in: the degradation of proteins (exogenous or intracellular) for nutrition purposes;

degradation of abnormal proteins; and maturation of certain proteins by removing the N-terminal methionine (methionine aminopeptidase)[195]. Other roles have been described, such as their involvement in plasmid stability (PepA), antibiotic uptake and peptidoglycan synthesis or degradation (D-aminopeptidase)[195].

The presence of aminopeptidases in spore formers has been previously reported. Cheggour *et al.*[197] have studied a gene, *dppA*, encoding a D-aminopeptidase in *B. subtilis*. Despite the fact that the *dppA* gene is expressed during early sporulation and regulated by Spo0A[198], the authors suggest that the protein is involved in adaptation to nutrient limitation rather than sporulation. This idea is supported by the fact that cells with null mutations in the *dciA* operon (containing the *dppA* gene) can still sporulate[199]. Aminopeptidases have also been detected in the mother cell and forespore of *Bacillus megaterium*[200]. The levels of aminopeptidase were constant between the two compartments and at different times during sporulation. However, the authors did not look into the possible role of the aminopeptidases during the sporulation process.

Despite the lack of studies reporting the role of aminopeptidase P in spore germination, some hypotheses can be suggested. The protein might be involved in:

i) release of amino acids that can be used for the synthesis of new proteins during germination and outgrowth;
ii) degradation of small acid-soluble proteins (SASPs);
iii) post-translational modification of peptides or proteins, provided they contain a proline next to the N-terminal amino acid.

6.1.1.7 CLSPO_c33910: Hypothetical protein

The gene CLSPO_c33910, which appears to be essential for sporulation (Chapter 4), encodes a small protein (89 aa) with unknown function and no known conserved domains[102]. A BLASTn search indicates that the gene shows 100% homology with a putative membrane protein from *C. sporogenes* strain DSM 795 (100% identity; e-value: $2e^{-137}$) and *C. botulinum* Prevot_594 (100% identity; e-value: $2e^{-137}$)[193].

6.2 Chapter overview

As mentioned above, data obtained from high-throughput approaches such as TraDIS and RNA-Seq require validation. Therefore, some genes identified in the previous chapters have been selected and targeted with the RiboCas system. The resulting

mutants have been briefly characterised: the ability to sporulate/germinate was evaluated, as well as the growth profile. Moreover, the spores and vegetative cells have been stained using the Schaeffer-Fulton method and observed under the microscope. The aim of this work was to confirm the involvement of these genes in the processes of sporulation and/or germination.

6.3 Results

6.3.1 Creation of the desired mutants

6.3.1.1 Deletion of selected genes using RiboCas

The selected genes were targeted with RiboCas[87], a system in which Cas9 is under the control of a riboswitch and that has been described in Chapter 5. However, an innovation was introduced to allow reversion of the gene to the original locus: the addition of a bookmark between the two start codons and the three stop codons of each gene (Figure 6.3). The bookmark consisted of a sgRNA target sequence from *C. botulinum* (GTTAACAGAGATTTACACAG) that showed good knock-out efficiencies (Daphne Groothuis, personal communication) and that was not present in the *C. sporogenes* genome.

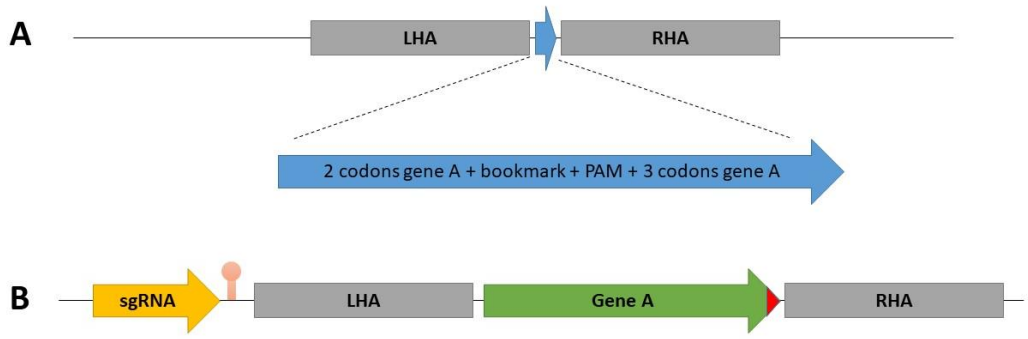


Figure 6.3: Schematic representation of the strategy used for reversion of the gene to the original locus. A: part of the genome of the knock-out strain; a bookmark (20 nucleotides) was introduced between the 2 first and 3 last codons of the gene that was deleted. A PAM was also inserted to ensure that the bookmark was recognised by Cas9. An additional nucleotide was introduced after the PAM, otherwise there would be a frame-shift that would possibly remove the stop codon. **B**: part of the vector used to revert the previously deleted gene; the sgRNA targets the bookmark and has a transcriptional terminator downstream; LHA and RHA are the left and right homology arms, respectively; gene A is the gene being reverted and contains a silent mutation in the stop codon (red).

The reversion of each gene to the original locus involved the generation of plasmids containing the entire gene and the homology arms previously used for the deletion of the gene. The plasmid also included a sgRNA targeting the sequence from *C. botulinum* described above (Figure 6.3). Once the plasmid was introduced into the cell, the homology arms allowed recombination between the plasmid and the genome, removing the bookmark from the chromosome and introducing the gene back at its

native locus. In the cells where recombination did not happen, Cas9 targeted the bookmark still present in the genome.

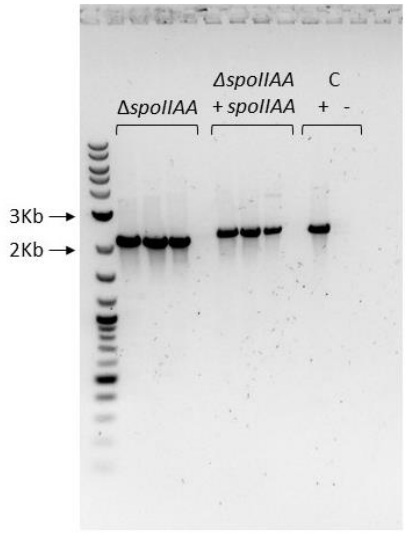


Figure 6.4: Gel electrophoresis picture of colony PCR samples demonstrating the deletion and reversion of the target genes. The *spoIIAA* deletion/reversion were screened using spoIIAA_scr_F and spoIIAA_scr_R primers. The gene was also screened in wild type (C+) and a negative control (water as template) was included (C-).

This method had one disadvantage: the reverted strains were indistinguishable from wild type. In order to address this issue, a silent mutation, or 'watermark', was introduced in the stop codon of the gene being complemented (Figure 6.3). For instance, the stop codon of the *sigE* gene (TAA) was changed to TAG in the complemented strain. This strategy allowed the reversion of the previously deleted genes at their original locus, ensuring wild-type expression levels.

The strains *C. sporogenes* Δ*sigF*, *C. sporogenes* Δ*sigE*, *C. sporogenes* Δ*sigG* and *C. sporogenes* Δ*sigK* were created as described in Chapter 5. The vectors pRECas1-*spoIIAA*_BM, pRECas1-APP_BM and pRECas1-c33910_BM were used to knock-out *spoIIAA*, APP and CLSPO_c33910, respectively, generating the strains *C. sporogenes* Δ*spoIIAA*, *C. sporogenes* ΔAPP and *C. sporogenes* ΔCLSPO_c33910. Later, the vectors pRECas1-*sigF*_rev, pRECas1-*sigE*_rev, pRECas1-*sigG*_rev, pRECas1-*sigK*_rev, pRECas1-*spoIIAA*_rev, pRECas1-APP_rev and pRECas1-c33910_rev were used to revert each gene in the respective mutant strain. Three independent biological replicates were obtained per strain, and the gene deletions/insertions were confirmed by colony PCR (Figure 6.4 shows one example) and by sequencing the region of the genome that includes the gene of interest, including the 'watermark'.

6.3.1.2 Insertion of ClosTron in selected genes

While most genes were knocked-out using the RiboCas system, others were targeted using ClosTron[85] as described in Chapter 3. Plasmids pMTL007::*Csp-spoIIE*-2368s, pMTL007::*Csp-spoIIR*-335s and pMTL007::*Csp- gerKA4*-775s were used to generate strains *C. sporogenes-spoIIE*2368s::CT, *C. sporogenes-spoIIR*335s::CT and *C. sporogenes-gerKA4*775s::CT, respectively (three independent biological replicates per strain). The insertion of the ClosTron in each target gene was confirmed by colony PCR (Figure 6.5 - A shows one example) and by sequencing the region of the genome containing the target gene. Single insertions of the ClosTron were confirmed by Southern blot (Figure 6.5 - B).

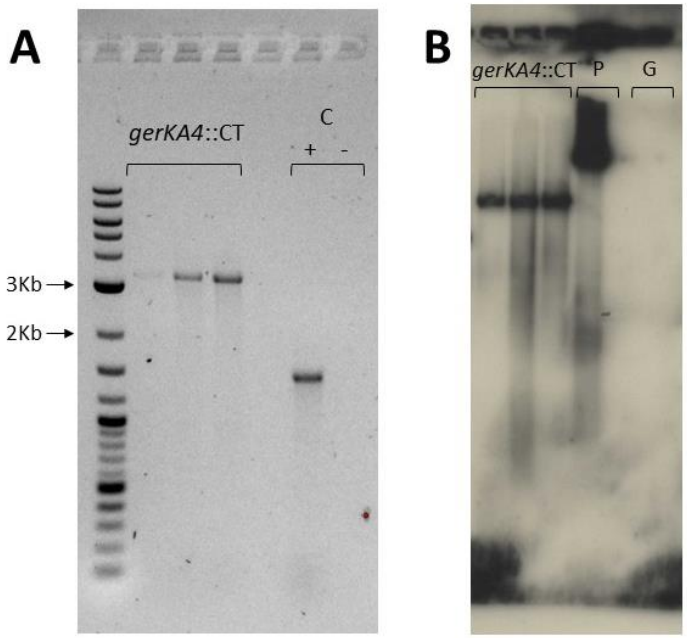


Figure 6.5: Evaluation of the ClosTron mutants. A: Gel electrophoresis of colony PCR samples demonstrating the insertion of the ClosTron in the target genes. The insertion into *gerKA4* was screened using gerKA4_scr_F and gerKA4_scr_R primers. The gene was also screened in wild type (C+); a negative control (water as template) was included for each reaction (C-).**B**: Southern blot demonstrating single ClosTron insertions in the *C. sporogenes-gerKA4*775s::CT strain. The membrane was developed for 15 minutes. P: control containing plasmid DNA, which should contain the ClosTron probe (positive control); G: control containing WT genomic DNA, which should not contain the ClosTron probe (negative control).

6.3.2 Characterisation of the sigma factor mutants

As described in Chapter 1, some sigma factors are well conserved across spore formers. Two examples are σ^F and σ^E, which are involved in the regulation of early sporulation in the forespore and mother cell, respectively. The results obtained in Chapter 5 indicate that σ^F might play a relevant role in the activation of the other sigma

factors by playing a part in the regulation of: i) *spoIIR*, involved in the activation of σ^E; ii) *sigG*, which encodes σ^G; and iii) *spoIVB*, involved in the activation of σ^K. The mother-cell specific σ^E, on the other hand, appears to regulate proteins involved in engulfment and the initial stages of the formation of the spore coat, and it also plays a role in the transcription and activation of σ^G and σ^K. In the absence of σ^F or σ^E, the expression of the sporulation cascade is stopped at a very early stage. Accordingly, *sigF* and *sigE* mutants are often arrested at the asymmetric division stage[10].

It is, therefore, not surprising that these sigma factors appear to be required for sporulation and germination according to the TraDIS results (Chapter 4) and that the *sigF* and *sigE* knock-outs are unable to form heat-resistant spores (Figure 6.6 and Figure 6.7). This is in accordance with previous studies with *C. difficile*[157] and *C. botulinum*[154]. The sporulation ability of the mutants was restored to wild type levels following reversion of the *sigF* and *sigE* genes, confirming that the asporogenous phenotype was indeed caused by the deletion of those genes. The growth profile of these mutants was not affected and is similar to wild type (Figure 6.10).

Although *C. perfringens* can still produce a few viable spores in the absence of σ^K [145], that appears not to be the case for *C. sporogenes*. The *sigG* and *sigK* mutants did not produce heat-resistant spores, and the reversion of the genes restored sporulation to wild type levels (Figure 6.8). The late stage sigma factors regulate the expression of proteins that appear to be required for sporulation/germination (Chapter 4), such as SASPs, SpoIVB (σ^G-dependent), YtaF and CotS (σ^K-dependent) (Table 7.1). Furthermore, the σ^G regulon includes proteins that are responsible for DNA protection at the spore core and the σ^K regulon appears to be involved in coat formation and spore maturation. Previous studies show that *sigG* and *sigK* mutants fail to produce spore coat or cortex in *B. subtilis*, and the same is observed in *sigK* mutants from *C. acetobutylicum* and *C. perfringens*[10]. On the other hand, *sigG* mutants from *C. difficile*, *C. acetobutylicum* and *C. perfringes* are able to form coat and/or cortex, but these appear to be defective[10]. These observations indicate that, despite being involved in the late stages of sporulation, σ^G and σ^K might be essential to the formation of mature, heat-resistant spores. Accordingly, the *C. sporogenes* Δ*sigG* and *C. sporogenes* Δ*sigK* strains appear to form immature, non-heat resistant spores (Figure 6.8) that do not stain green using the Schaeffer-Fulton method[122] (Figure 6.9). Nevertheless, the growth profile of the mutants was similar to wild type (Figure 6.10).

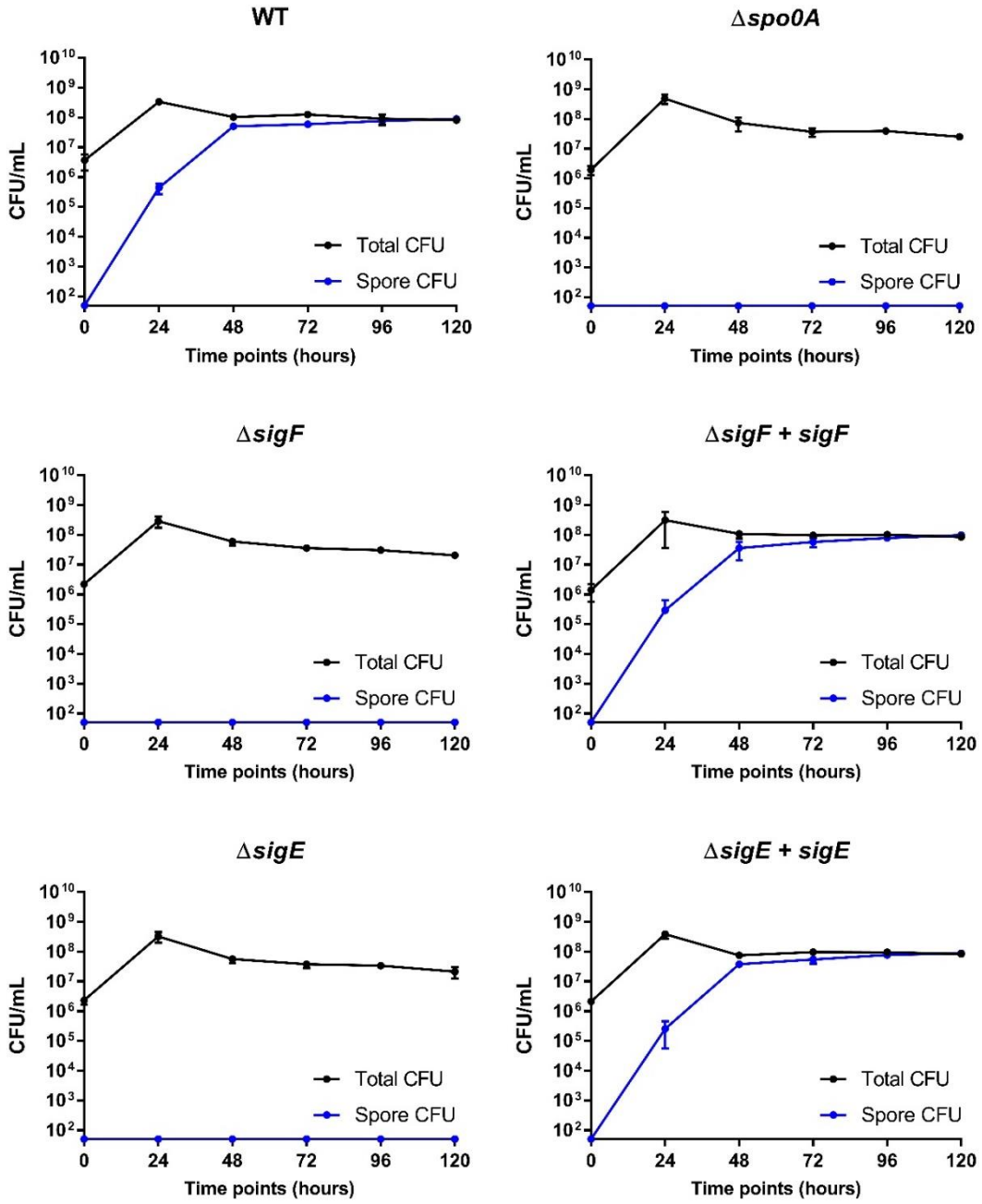


Figure 6.6: Sporulation profile of the *C. sporogenes* Δ*sigF*, *C. sporogenes* Δ*sigF* + *sigF*, *C. sporogenes* Δ*sigE* and *C. sporogenes* Δ*sigE* + *sigE* strains. *C. sporogenes* WT and *C. sporogenes* Δ*spo0A* were included as positive and negative controls, respectively. All the cultures were synchronised at the beginning of the assay, and samples were collected at 24 hour intervals during a total period of 120 hours. For the total CFU count, the cultures were diluted and spotted into TYG agar. For spore CFU count, the cultures were heat-treated at 80°C for 20 minutes prior to the dilutions. Data show the mean and error bars represent the standard deviation of three biological and three technical replicates. The Y axis starts at the limit of detection (50 spores/mL).

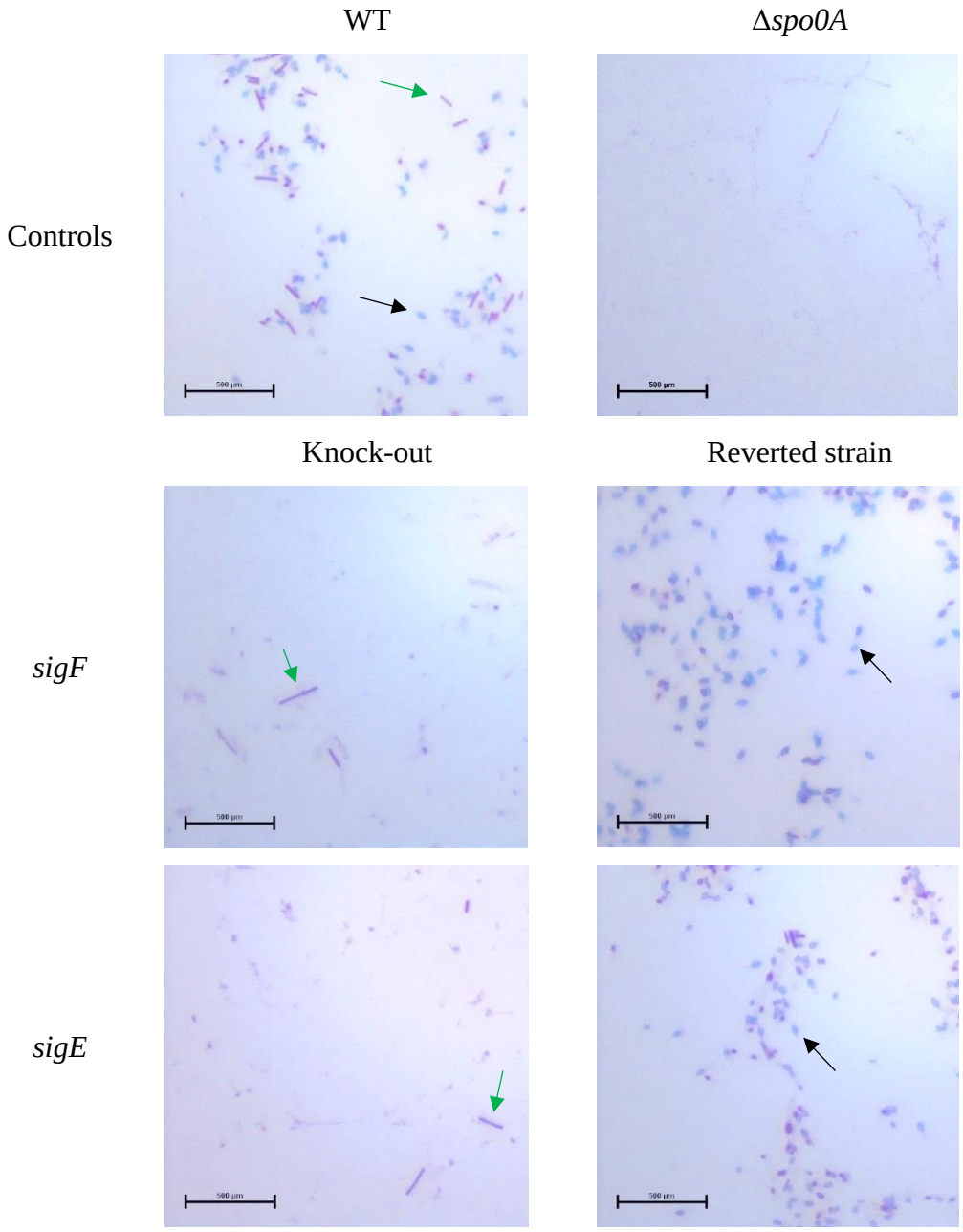


Figure 6.7: *C. sporogenes* strains stained using the Schaeffer-Fulton method. *C sporogenes* WT, *C sporogenes* Δ*spo0A*, *C sporogenes* Δ*sigF*, *C sporogenes* Δ*sigF*+*sigF*, *C sporogenes* Δ*sigE*, *C sporogenes* Δ*sigE*+*sigE* were grown for 120 hours and stained as described in Schaeffer *et al.*[122]. The scale bar represents 500 µm. Black arrows indicate green stained spores and green arrows indicate red stained vegetative cells. No spores were observed in the mutant strains.

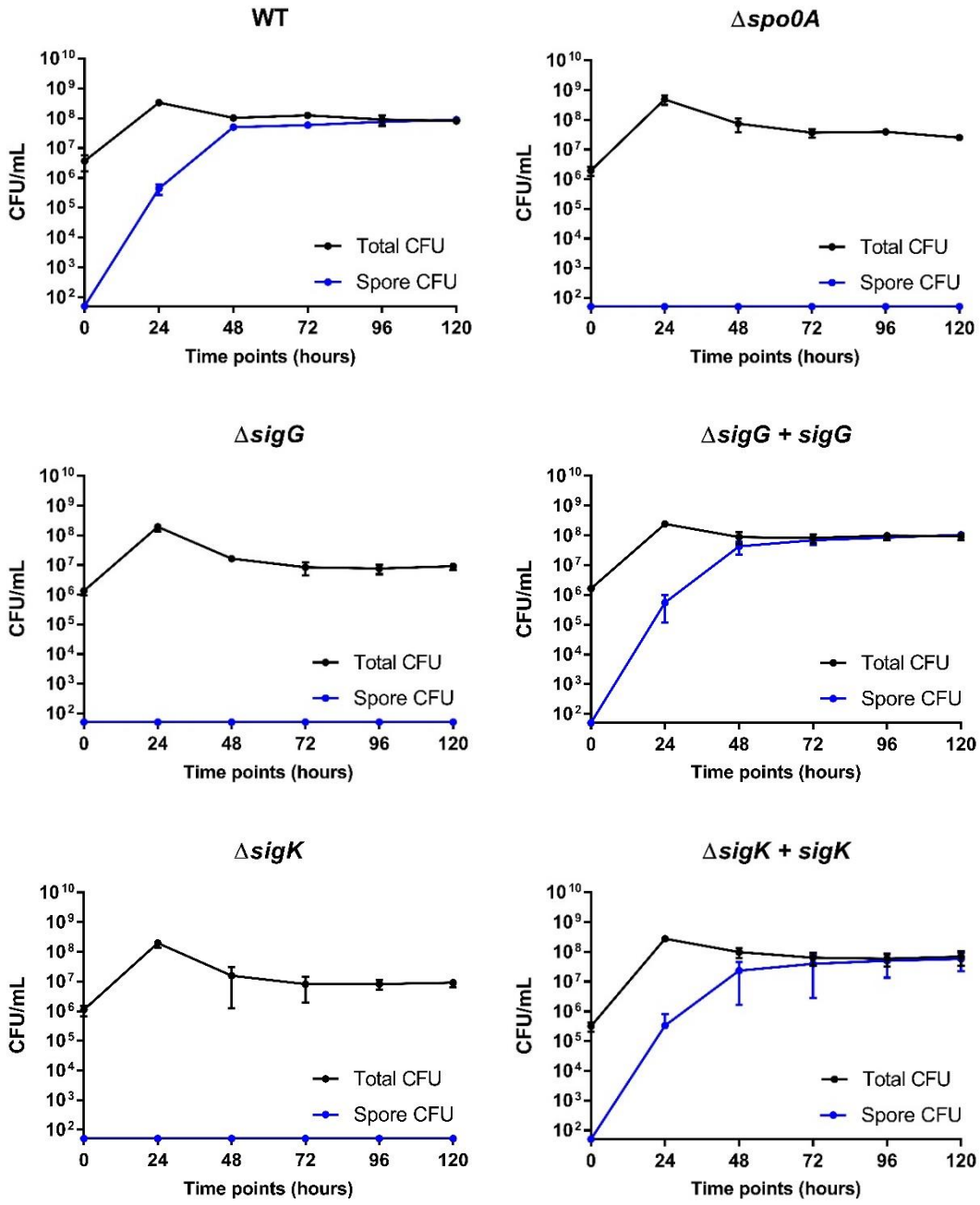


Figure 6.8: Sporulation profile of the *C. sporogenes* Δ*sigG*, *C. sporogenes* Δ*sigG* + *sigG*, *C. sporogenes* Δ*sigK* and *C. sporogenes* Δ*sigK* + *sigK* strains. *C. sporogenes* WT and *C. sporogenes* Δ*spo0A* were included as positive and negative controls, respectively. All the cultures were synchronised at the beginning of the assay, and samples were collected at 24 hour intervals during a total period of 120 hours. For the total CFU count, the cultures were diluted and spotted into TYG agar. For spore CFU count, the cultures were heat-treated at 80°C for 20 minutes prior to the dilutions. Data shows the mean and error bars represent the standard deviation of three biological and three technical replicates. The Y axis starts at the limit of detection (50 spores/mL).

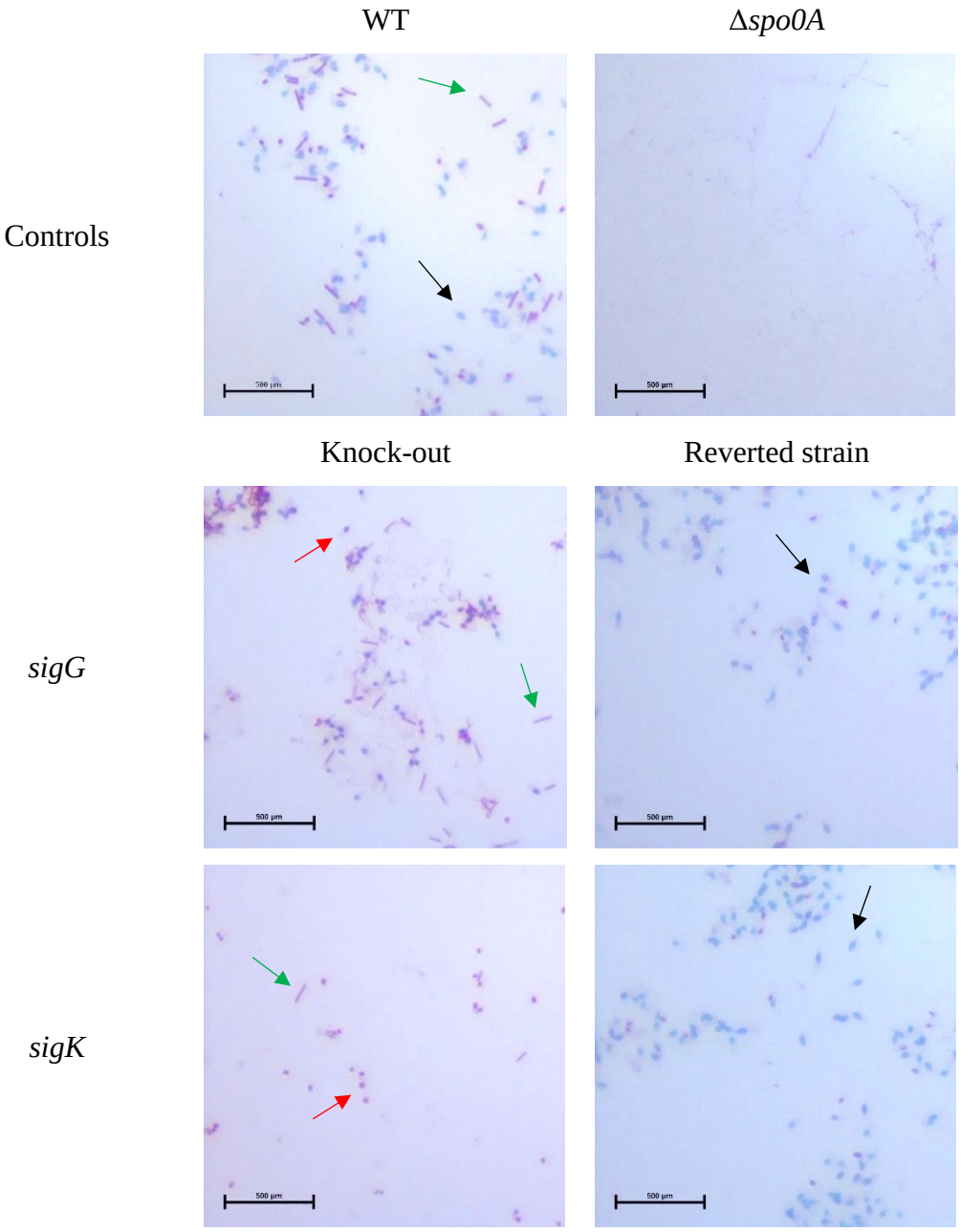


Figure 6.9: *C. sporogenes* strains stained using the Schaeffer-Fulton method. *C sporogenes* WT, *C sporogenes* Δ*spo0A*, *C sporogenes* Δ*sigG*, *C sporogenes* Δ*sigG*+*sigG*, *C sporogenes* Δ*sigK*, *C sporogenes* Δ*sigK*+*sigK* were grown for 120 hours and stained as described in Schaeffer *et al.*[122]. The scale bar represents 500 µm. Black arrows indicate green stained spores, green arrows indicate red stained vegetative cells and red arrows represent red stained underdeveloped spores. No spores were observed in the mutant strains.

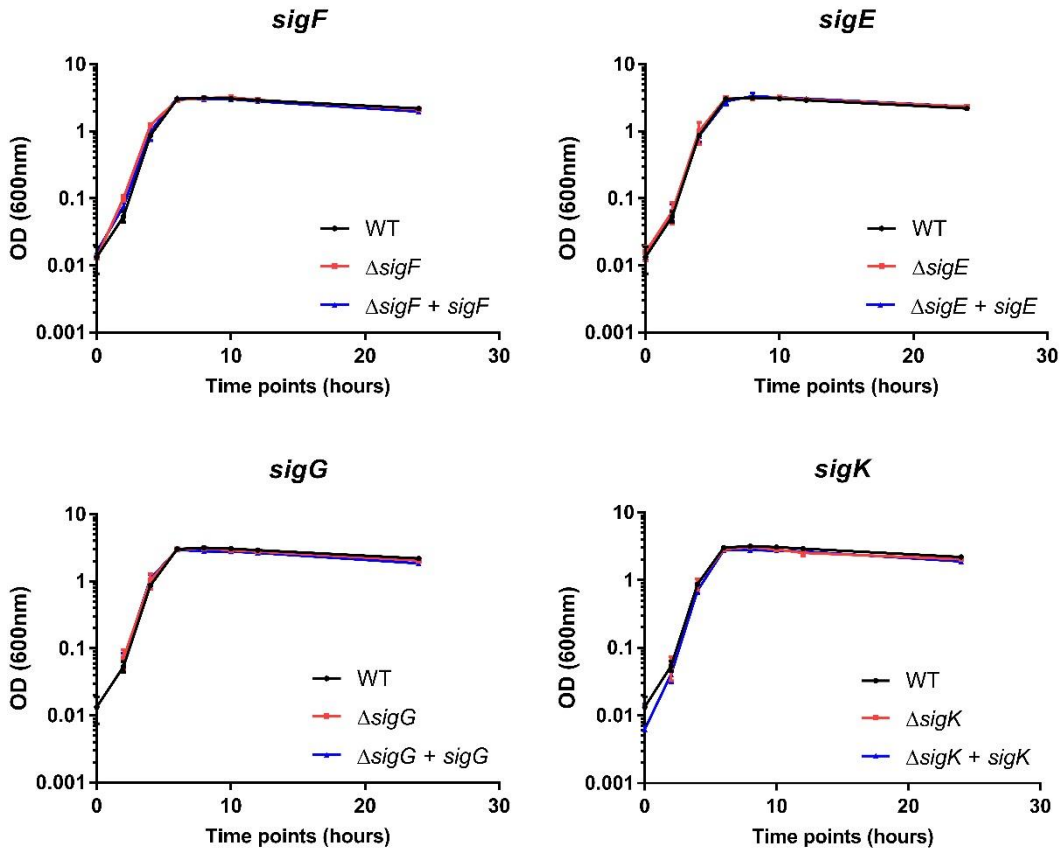


Figure 6.10: Growth profiles of *C. sporogenes* WT, *C. sporogenes* Δ*sigF*, *C. sporogenes* Δ*sigF* + *sigF*, *C. sporogenes* Δ*sigE*, *C. sporogenes* Δ*sigE* + *sigE*, *C. sporogenes* Δ*sigG*, *C. sporogenes* Δ*sigG* + *sigG*, *C. sporogenes* Δ*sigK* and *C. sporogenes* Δ*sigK* + *sigK*. The OD_{600} of each sample was measured every 2 hours during the first 12 hours of growth, and then again 24 hours after inoculation. Data shows the mean and error bars represent the standard deviation of three biological and three technical replicates.

6.3.3 Characterisation of the *spoIIAA* mutant

As described above, SpoIIAA is involved in the activation of σ^F, the sigma factor responsible for the regulation of early-sporulation in the forespore. It has been shown in this chapter that σ^F is essential for sporulation in *C. sporogenes*, as *sigF* mutants are unable to sporulate. Thus, it is not surprising that SpoIIAA is also a crucial part of this process. In the absence of SpoIIAA, σ^F remains trapped in the SpoIIAB$_2$-σ^F complex and the transcription of the forespore-specific genes is inhibited[136, 171].

The relevant role of SpoIIAA in sporulation in *C. sporogenes* was suggested by the TraDIS results (Chapter 4) and confirmed by a sporulation assay (Figure 6.11 - A). Indeed, mutants lacking the *spoIIAA* gene were unable to sporulate, while the reversion of the gene restored sporulation to wild type levels. Furthermore, when the mutant strains were stained using the Schaeffer-Fulton method[122], no spores were visible (only vegetative cell debris) (Figure 6.12).

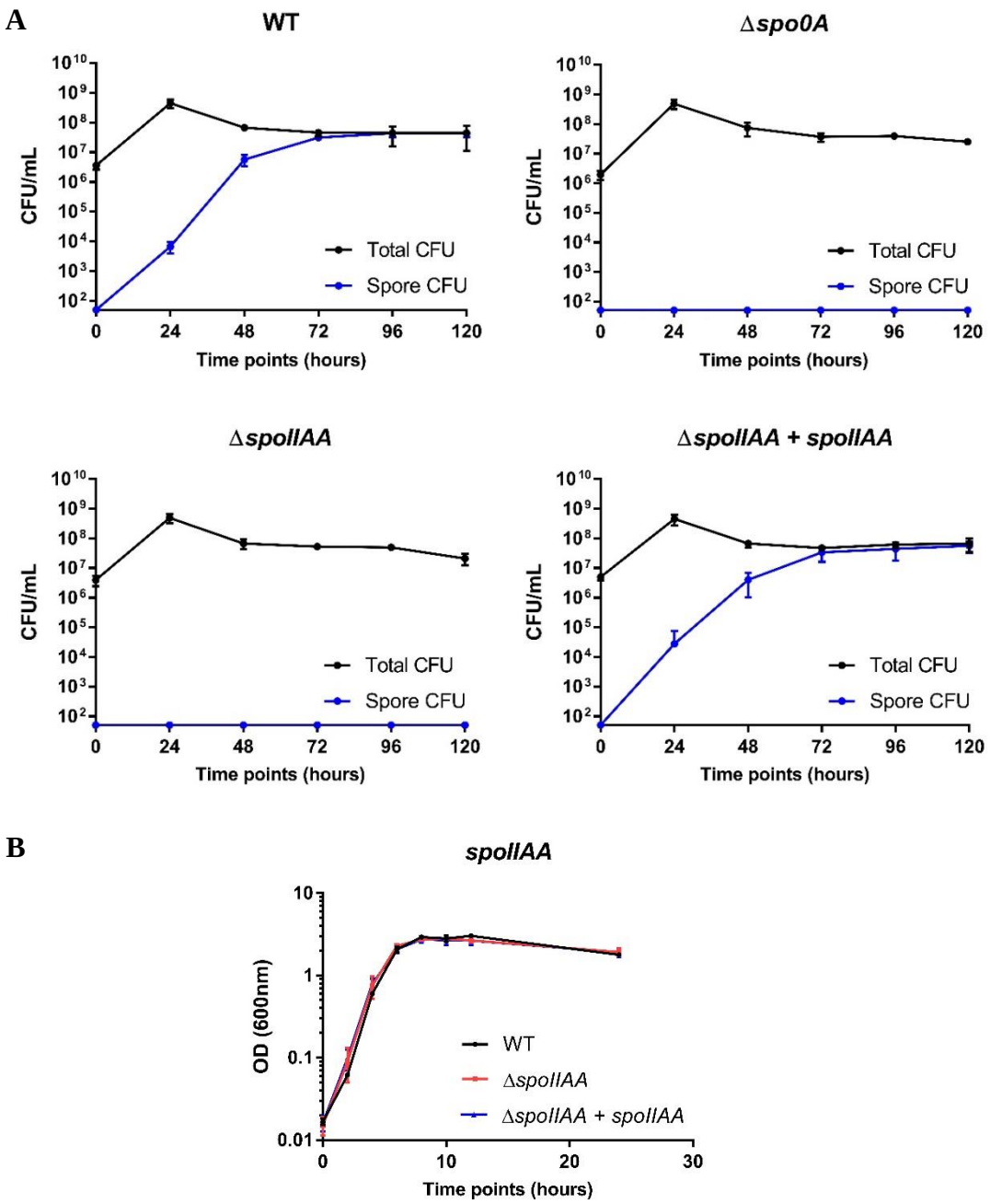


Figure 6.11: Characterisation of *C. sporogenes* Δ*spoIIAA* and *C. sporogenes* Δ*spoIIAA* + *spoIIAA* strains. A: Sporulation profile of the *C. sporogenes* Δ*spoIIAA* and *C. sporogenes* Δ*spoIIAA* + *spoIIAA* strains. *C. sporogenes* WT and *C. sporogenes* Δ*spo0A* were included as positive and negative controls, respectively. All the cultures were synchronised at the beginning of the assay, and samples were collected at 24 hour intervals during a total period of 120 hours. For the total CFU count, the cultures were diluted and spotted into TYG agar. For spore CFU count, the cultures were heat-treated at 80°C for 20 minutes prior to the dilutions. Data shows the mean and error bars represent the standard deviation of three biological and three technical replicates. The Y axis starts at the limit of detection (50 spores/mL). **B**: Growth profiles of *C. sporogenes* WT, *C. sporogenes* Δ*spoIIAA* and *C. sporogenes* Δ*spoIIAA* + *spoIIAA*. The OD_{600} of each sample was measured every 2 hours during the first 12 hours of growth, and then again 24 hours after inoculation. Data shows the mean and error bars represent the standard deviation of three biological and three technical replicates.

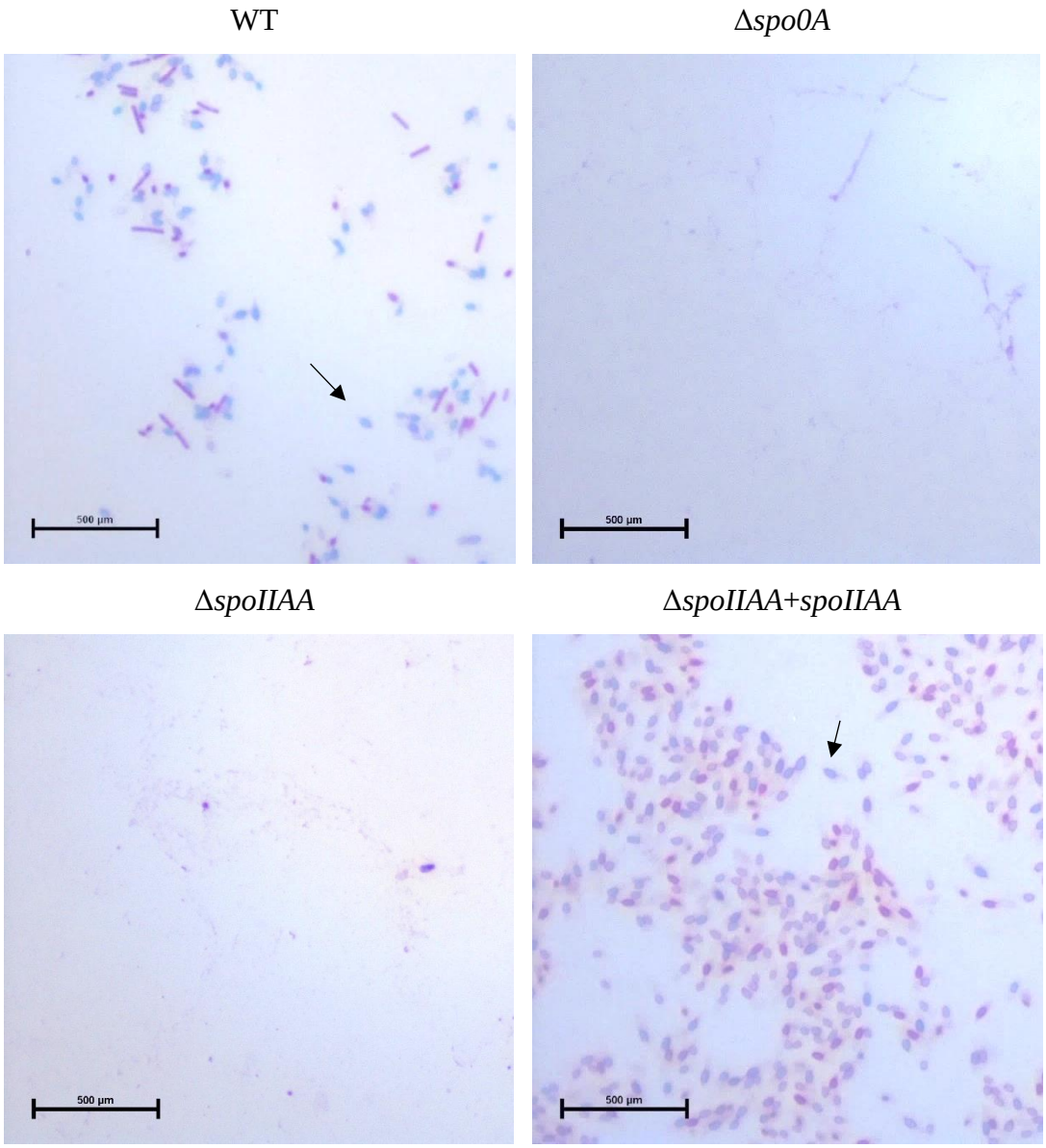


Figure 6.12: *C. sporogenes* strains stained using the Schaeffer-Fulton method. *C sporogenes* WT, *C sporogenes* Δ*spo0A*, *C sporogenes* Δ*spoIIAA* and *C sporogenes* Δ*spoIIAA*+*spoIIAA* were grown for 120 hours and stained as described in Schaeffer *et al.*[122]. The scale bar represents 500 µm. Black arrows indicate green stained spores. No spores were observed in the *C sporogenes* Δ*spoIIAA* strain.

On the other hand, several spores were observed in the reverted strains (Figure 6.12). The growth profile of the *spoIIAA* mutants and reverted strains resembled that of wild type (Figure 6.11 - B). The RNA-Seq results (Chapter 5) indicate that *spoIIAA* is not regulated by any of the sigma factors. This is in line with what was observed in *B. subtilis*, where the expression of the *spoIIAA-spoIIAB-sigF* operon is regulated by Spo0A and σ^{H} [13, 34].

6.3.4 Characterisation of the APP mutant

A gene encoding a M24 family metallopeptidase contains several insertions in the vegetative cell and sporulation libraries, but very few insertions in the germination library (Chapter 4). These results suggest that this gene might be essential for the germination of *C. sporogenes* spores. In addition, the RNA-Seq data indicate that this gene is regulated by σ^F and σ^E (Chapter 5). The involvement of sigma factors in the transcriptional regulation of aminopeptidases has been previously reported[195].

The deletion of the APP gene had no effect on the sporulation profile of *C. sporogenes* (Figure 6.13 - A). This is further evidenced by the observation of spores in the *C. sporogenes* ΔAPP strains (Figure 6.14). The ability of the APP mutant to sporulate might be explained by the presence of other aminopeptidases with similar functions in the *C. sporogenes* genome (CLSPO_c03740, CLSPO_c22000). In order to establish a proper link between these and the process of germination, a mutant should have been created where all the aminopeptidases would be knocked-out. Nevertheless, a germination assay might show differences between the mutant and wild type, as germination is the process expected to be affected by the mutation.

The results presented in Figure 6.13 - B show that the APP mutant has a slight growth disadvantage relative to wild type. An unpaired t-test (performed using GraphPad 7) revealed that the difference in growth was statistically significant after 12 hours (p-value = 0.0003) and 24 hours (p-value = 0.0028) of growth. The difference in growth was also observed in the reverted strain (statistically significant after 10, 12 and 24 hours of growth), suggesting that there might be a SNP in another gene (other than APP) causing this phenotype.

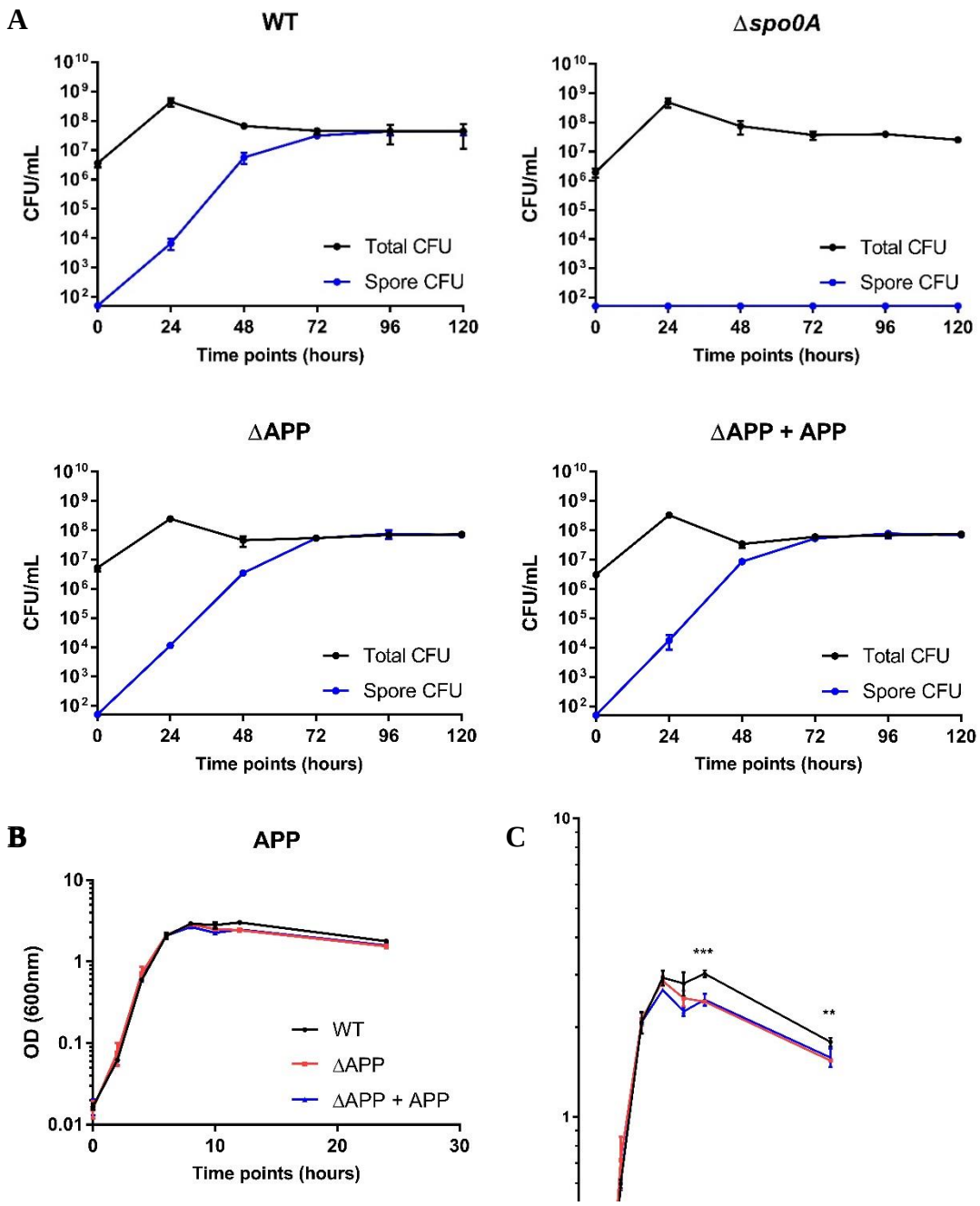


Figure 6.13: Characterisation of *C. sporogenes* ΔAPP and *C. sporogenes* ΔAPP + APP strains. **A**: Sporulation profile of the *C. sporogenes* ΔAPP and *C. sporogenes* ΔAPP + APP strains. *C. sporogenes* WT and *C. sporogenes* Δ*spo0A* were included as positive and negative controls, respectively. All the cultures were synchronised at the beginning of the assay, and samples were collected at 24 hour intervals during a total period of 120 hours. For the total CFU count, the cultures were diluted and spotted into TYG agar. For spore CFU count, the cultures were heat-treated at 80°C for 20 minutes prior to the dilutions. Data shows the mean and error bars represent the standard deviation of three biological and three technical replicates. The Y axis starts at the limit of detection (50 spores/mL). **B**: Growth profiles of *C. sporogenes* WT, *C. sporogenes* ΔAPP and *C. sporogenes* ΔAPP + APP. The OD_{600} of each sample was measured every 2 hours during the first 12 hours of growth, and then again 24 hours after inoculation. Error bars represent the standard deviation of three biological and three technical replicates. **C**: amplified image of the growth profile graphic (B) showing statistically significant differences in growth between *C. sporogenes* WT and *C. sporogenes* ΔAPP (unpaired t-test).

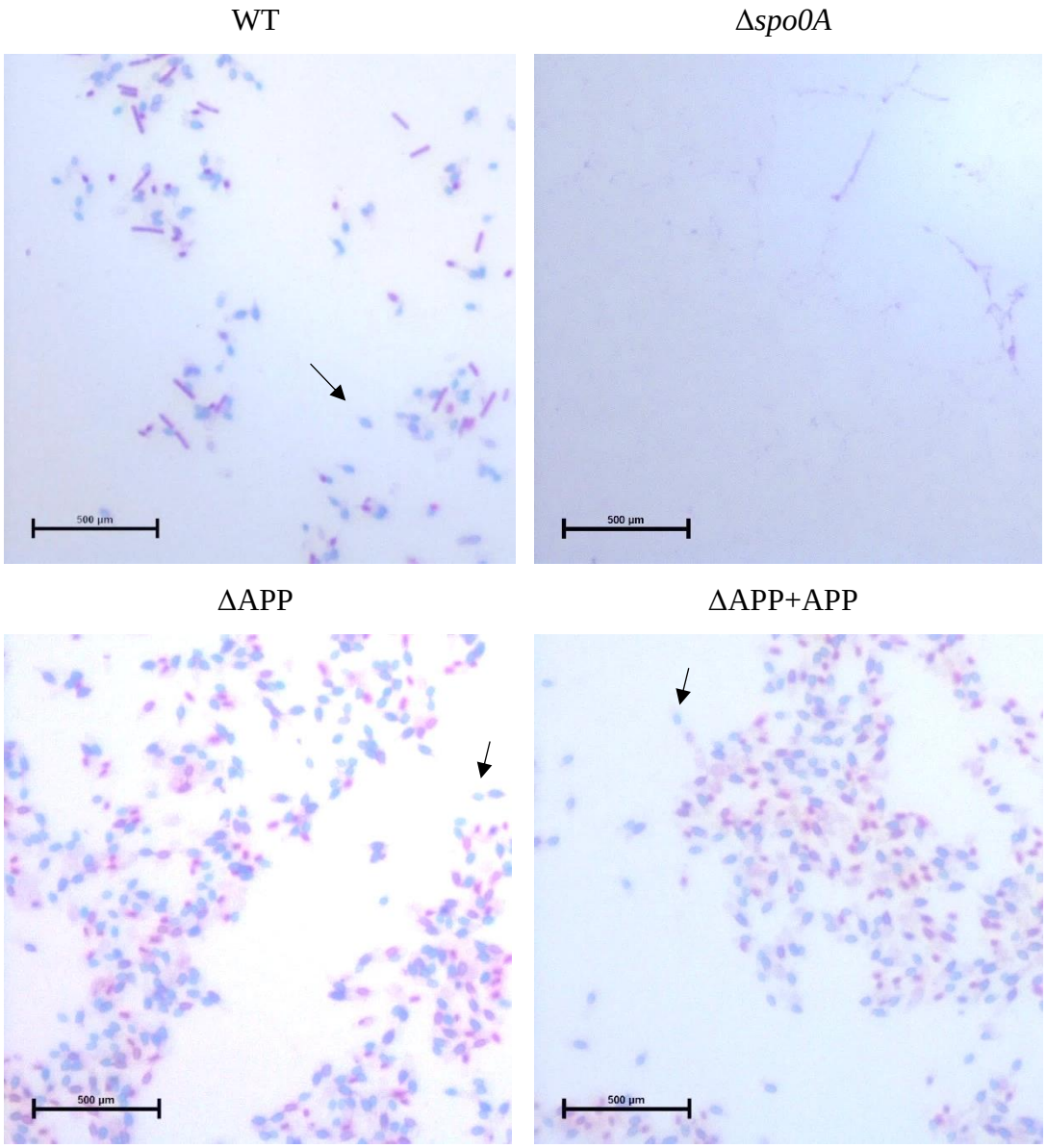


Figure 6.14: *C. sporogenes* strains stained using the Schaeffer-Fulton method. *C sporogenes* WT, *C sporogenes* Δ*spo0A*, *C sporogenes* ΔAPP and *C sporogenes* ΔAPP+APP were grown for 120 hours and stained as described in Schaeffer *et al.*[122]. The scale bar represents 500 µm. Black arrows indicate green stained spores.

6.3.5 Characterisation of the CLSPO_c33910 mutant

The gene with the locus tag CLSPO_c33910 had insertions in the vegetative cell library but none in the sporulation and germination libraries, indicating an essential role in these mechanisms in *C. sporogenes* (Chapter 4). Despite this apparent role, this gene does not appear to be regulated by any of the four main sigma factors that regulate sporulation (Chapter 5).

The sporulation profile of the *C. sporogenes* Δc33910 strain was intriguing, with two of the biological replicates yielding spores to wild type levels and a third replicate showing a lower ability to sporulate (Figure 6.15 - A). The sporulation profiles were

in line with the microscope observations, with a lower number of spores visible in biological replicate 3 than in replicates 1 and 2 (Figure 6.16). With the aim of identifying the cause of the different phenotypes, the whole genome of each biological replicate was sequenced. The reads were mapped to the *C. sporogenes* NCIMB 10696 genome using the CLC Genomics Workbench 12 and a variant analysis was performed. The analysis revealed that the only difference between biological replicate 3 and the other two biological replicates was an insertion of an adenosine nucleotide in the gene *resE4*, encoding a sensor histidine kinase ResE. The gene *resE4* appears not to be regulated by any of the sigma factors analysed in the RNA-Seq experiments, and it does not seem to be required for sporulation or germination according to the TraDIS data. Nevertheless, this gene appears to be a part of the *agrBD* quorum sensing system described by Cooksley *et al.*[201]. The system has been hypothesised to have a role in the regulation of sporulation, as the sporulation ability of *agrD* mutants (*C. botulinum*) and strains with modulated *agrB* activity (*C. sporogenes*) was significantly reduced[201]. The strain used by Cooksley *et al.*[201] was *C. sporogenes* NCIMB 10696, the same strain that was the focus of this thesis. Furthermore, the gene *resE4* that had an insertion in biological replicate 3 was the same as CSP0340[201], a gene that encodes an orphan kinase and is located immediately downstream of *agrD2*. Thus, it is likely that the reduced sporulation levels observed with *C. sporogenes* Δc33910 – biological replicate 3 were due to the mutation of the *resE4* gene rather than the deletion of CLSPO_c33910.

Similar to what was observed with the APP mutant, *C. sporogenes* Δc33910 has a slight growth disadvantage compared to wild type (Figure 6.15 - B) and the difference is statistically significant after 8 hours (p-value = 0.0046), 12 hours (p-value = 0.0114) and 24 hours (p-value = 0.0325) of growth (unpaired t-test performed with GraphPad 7). The difference in growth between *C. sporogenes* Δc33910 + c33910 and wild type is statistically significant only after 12 hours (p-value = 0.0185). It is possible that the difference in growth caused the misidentification of CLSPO_c33910 as essential for sporulation, but further investigations are required to confirm this.

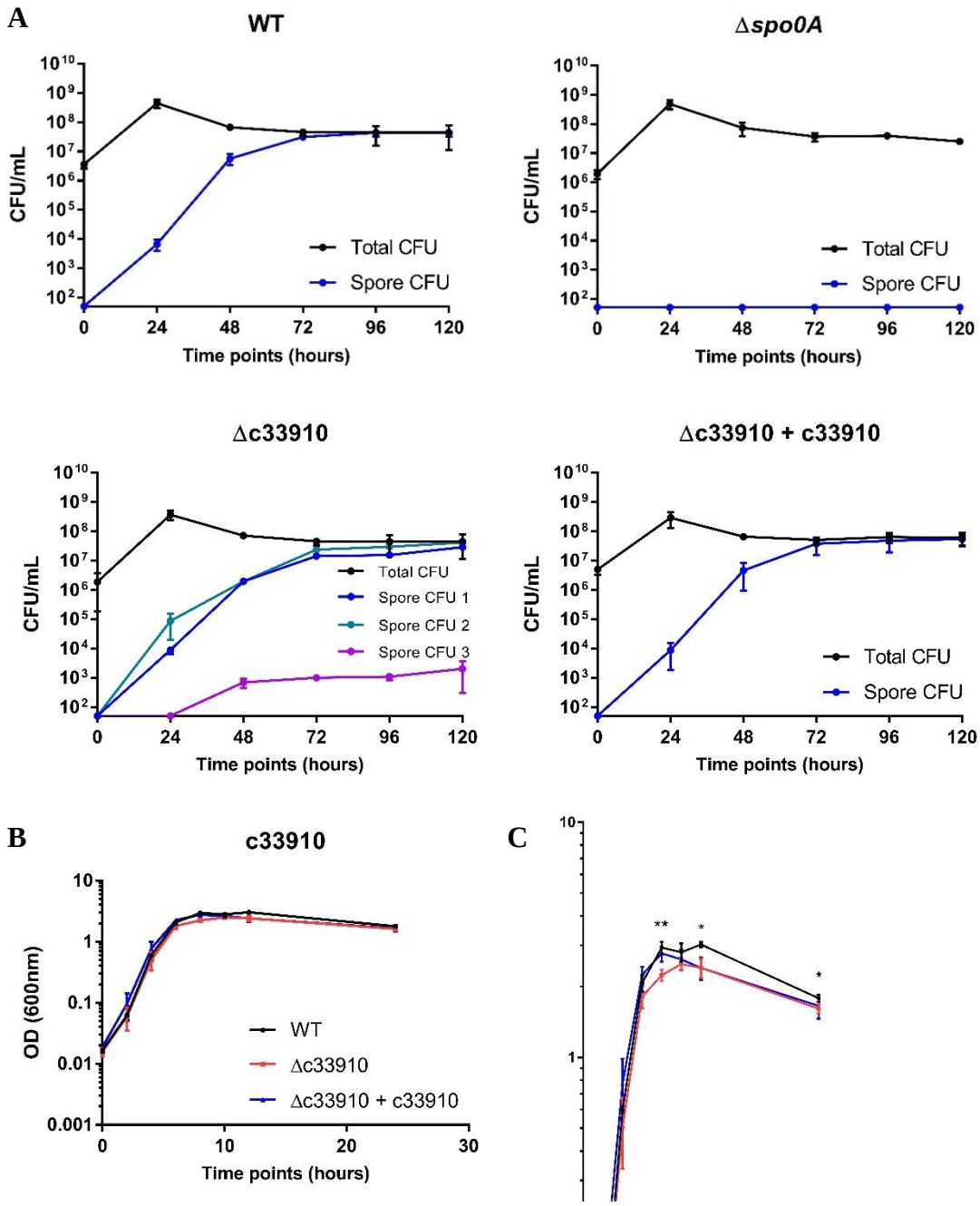


Figure 6.15: Characterisation of *C. sporogenes* Δc33910 and *C. sporogenes* Δc33910 + c33910 strains. A: Sporulation profile of the *C. sporogenes* Δc33910 and *C. sporogenes* Δc33910 + c33910 strains. *C. sporogenes* WT and *C. sporogenes* Δ*spo0A* were included as positive and negative controls, respectively. All the cultures were synchronised at the beginning of the assay, and samples were collected at 24 hour intervals during a total period of 120 hours. For the total CFU count, the cultures were diluted and spotted into TYG agar. For spore CFU count, the cultures were heat-treated at 80°C for 20 minutes prior to the dilutions. Data show the mean and error bars represent the standard deviation of three biological and three technical replicates. The Y axis starts at the limit of detection (50 spores/mL). **B**: Growth profiles of *C. sporogenes* WT, *C. sporogenes* Δc33910 and *C. sporogenes* Δc33910 + c33910. The OD_{600} of each sample was measured every 2 hours during the first 12 hours of growth, and then again 24 hours after inoculation. Data show the mean and error bars represent the standard deviation of three biological and three technical replicates. **C**: amplified image of the growth profile graphic (B) showing statistically significant differences in growth between *C. sporogenes* WT and *C. sporogenes* Δc33910 (unpaired t-test).

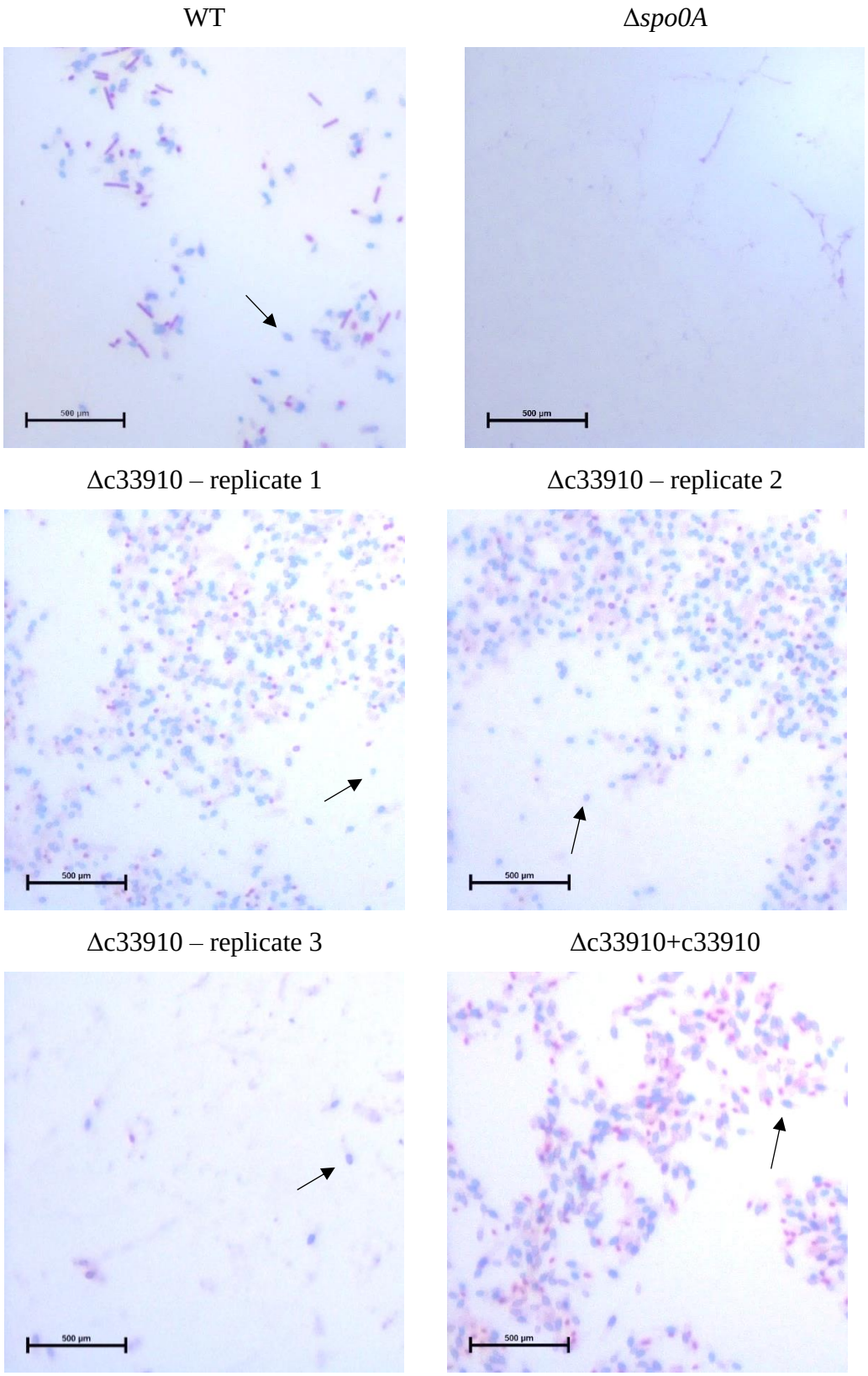


Figure 6.16: *C. sporogenes* strains stained using the Schaeffer-Fulton method. *C. sporogenes* WT, *C. sporogenes* Δ*spo0A*, *C. sporogenes* Δc33910 and *C. sporogenes* Δc33910+c33910 were grown for 120 hours and stained as described in Schaeffer *et al.*[122]. The scale bar represents 500 µm. Black arrows indicate green stained spores.

6.3.6 Sporulation ability of other selected genes

At the beginning of the PhD project, some genes involved in sporulation/germination were selected and targeted with ClosTron with the purpose of learning the technique. Hence, the strains *C. sporogenes-spoIIE*2368s::CT, *C. sporogenes-spoIIR*335s::CT and *C. sporogenes-gerKA4*775s::CT were created. The ability of these mutants to sporulate was evaluated. This work was done in collaboration with Daphne Groothuis.

6.3.6.1 Characterisation of the *spoIIE* mutant

Similar to SpoIIAA, SpoIIE is involved in the activation of σ^F in the forespore. The lack of activation of σ^F in the *spoIIE* mutant might explain why this gene was deemed important for sporulation in *C. sporogenes* (Chapter 4) and why the mutant failed to produce viable spores (Figure 6.17).

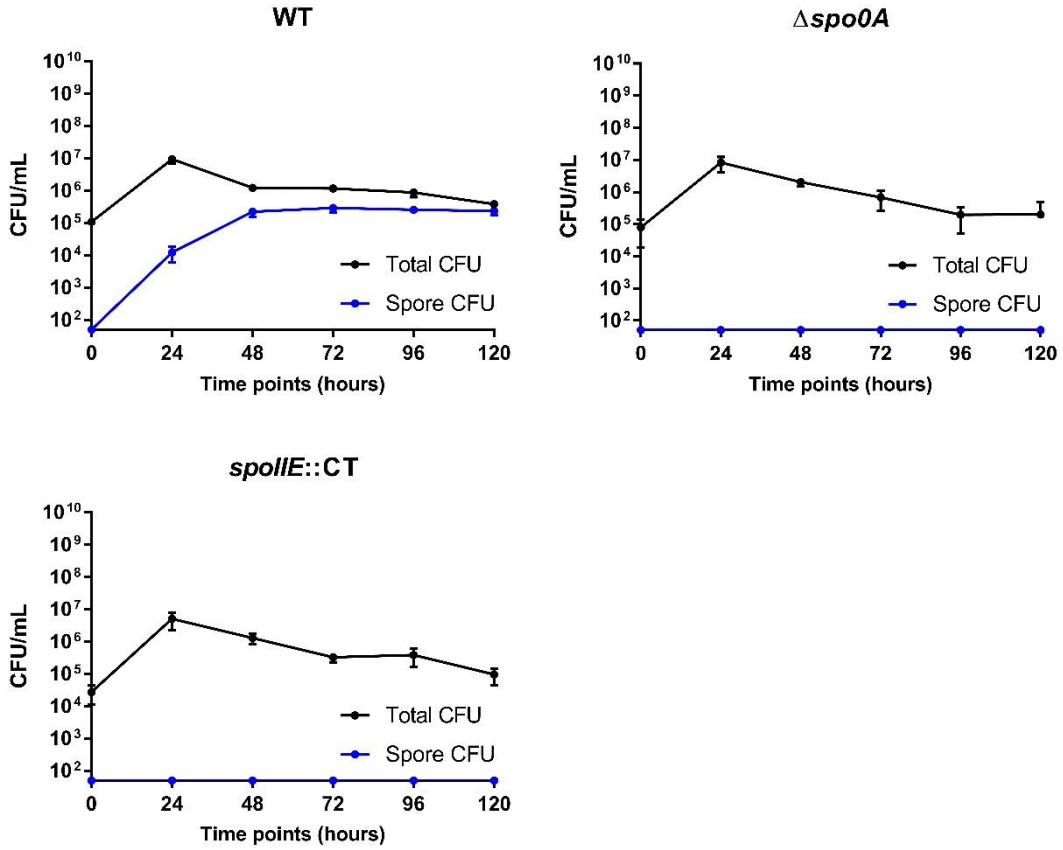


Figure 6.17: Sporulation profile of the *C. sporogenes-spoIIE*2368s::CT strain. *C. sporogenes* WT and *C. sporogenes* Δ*spo0A* were included as positive and negative controls, respectively. All the cultures were synchronised at the beginning of the assay, and samples were collected at 24 hour intervals during a total period of 120 hours. For the total CFU count, the cultures were diluted and spotted into TYG agar. For spore CFU count, the cultures were heat-treated at 80°C for 20 minutes prior to the dilutions. Data show the mean and error bars represent the standard deviation of three biological and three technical replicates. The Y axis starts at the limit of detection (50 spores/mL).

The sporulation-defective phenotype of *spoIIE* null mutants has been previously reported in *B. subtilis* [178]. The *spoIIE* gene is not regulated by any of the sigma factors (Chapter 5) and is probably regulated by Spo0A, similar to what was reported for *B. subtilis*[13].

6.3.6.2 Characterisation of the *spoIIR* mutant

SpoIIR is involved in the activation of σ^{E} [60], which has been shown to be essential for the sporulation process in *C. sporogenes* (see above). In *B. subtilis*, *spoIIR* is regulated by σ^{F} [60]. The same is observed in *C. sporogenes*, according to the RNA-Seq data (Chapter 5). However, contrary to what was observed in *B. subtilis*[60], *spoIIR* also appears to be regulated by σ^{G} in *C. sporogenes*.

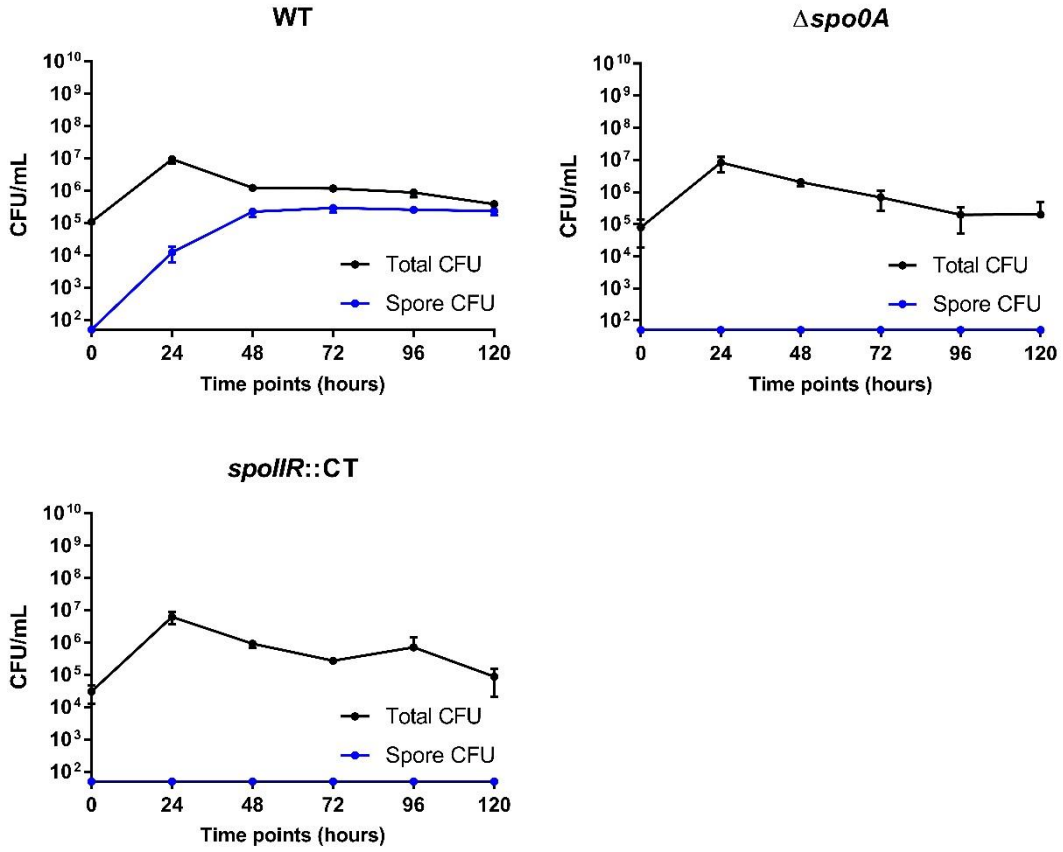


Figure 6.18: Sporulation profile of the *C. sporogenes-spoIIR*335s::CT strain. *C. sporogenes* WT and *C. sporogenes* Δ*spo0A* were included as positive and negative controls, respectively. All the cultures were synchronised at the beginning of the assay, and samples were collected at 24 hour intervals during a total period of 120 hours. For the total CFU count, the cultures were diluted and spotted into TYG agar. For spore CFU count, the cultures were heat-treated at 80°C for 20 minutes prior to the dilutions. Data show the mean and error bars represent the standard deviation of three biological and three technical replicates. The Y axis starts at the limit of detection (50 spores/mL).

A study by de Hoon *et al.*[9] reports that SpoIIR is conserved in all 24 spore formers that were analysed, suggesting that this protein might be essential for sporulation. That hypothesis is supported by the present study, as the *spoIIR* gene was identified as relevant for sporulation in *C. sporogenes* (Chapter 4) and the *C. sporogenes-spoIIR*::CT strain failed to produce viable spores (Figure 6.18). The absence of heat resistant spores in a *spoIIR* mutants has been reported in *B. subtilis*[187, 202] and in *C. difficile*[147].

6.3.6.3 Characterisation of the *gerKA4* mutant

In *B. subtilis*, the *gerK* operon usually consists of three genes and encodes nutrient receptors with the ability to trigger sporulation. In *C. sporogenes*, the gene annotated as *gerKA4* (CLSPO_c01480) appears to be in an operon with CLSPO_c01490 (spore germination protein) and CLSPO_c01500 (germination protein, GerC family) The operon appears to be regulated by σ^G (Chapter 5). The *gerK* operon is also regulated by σ^G in *B. subtilis*[58].

All three genes of the *gerK* operon have been shown to be essential for germination induced by AGFK in *B. subtilis*[192]. However, the *C. sporogenes* spores with a ClosTron insertion in the *gerKA4* gene were able to germinate (Figure 6.19) and the gene does not appear to be essential for germination in the TraDIS library (Chapter 4). Nevertheless, the *gerKA4* gene could tolerate fewer transposon insertions in the germination library than in the sporulation library ($log_2FC = -4.3$), indicating that the product of this gene has a relevant role in germination.

One possible explanation for ability of *C. sporogenes-gerKA4*775s::CT spores to germinate is the fact that other *ger* operons are present in the *C. sporogenes* genome:

- *gerKA1*-CLSPO_c02980-*ger(X)C*;
- *gerKA2-gerAB-gerKA5*-CLSPO19800- *ger(X)C* - CLSPO19820;
- *gerKA3*-CLSPO_c28710-*gerAC*;
- CLSPO_c29430-*gerKA6*-CLSPO-c29450.

The expression of all these operons also appears to be regulated by σ^G (Chapter 5), and all the unidentified proteins are annotated as “spore germination protein”. The fact that some of these operons are not tricistronic and that CLSPO_c29430 is oriented in the opposite direction relative to *gerKA6* and CLSPO-c29450 is in line with a previous report that *ger* operons show unusual structures in Clostridia[65]. The presence of other

potential *gerK* operons in *C. sporogenes* NCIMB 10696 suggests that other receptors might be able to recognise AGFK in the *gerKA4* mutant.

Another explanation is that other factors/nutrients might have triggered germination in the *C. sporogenes-gerKA4*775s::CT strain. This hypothesis is supported by the fact that germination was triggered by plating the spores onto rich, non-defined medium (TYG) and by the presence of other potential receptors.

It is surprising to find *gerK* operons in *C. sporogenes* NCIMB 10696, since *C. sporogenes* NCIMB 701792 and *C. sporogenes* strain 79-3 were not able to germinate in response to AGFK[67]. As mentioned above, the medium used in this study was non-defined and further studies would be necessary to evaluate how *C. sporogenes* NCIMB 10696 responds to AGFK and/or other nutrients.

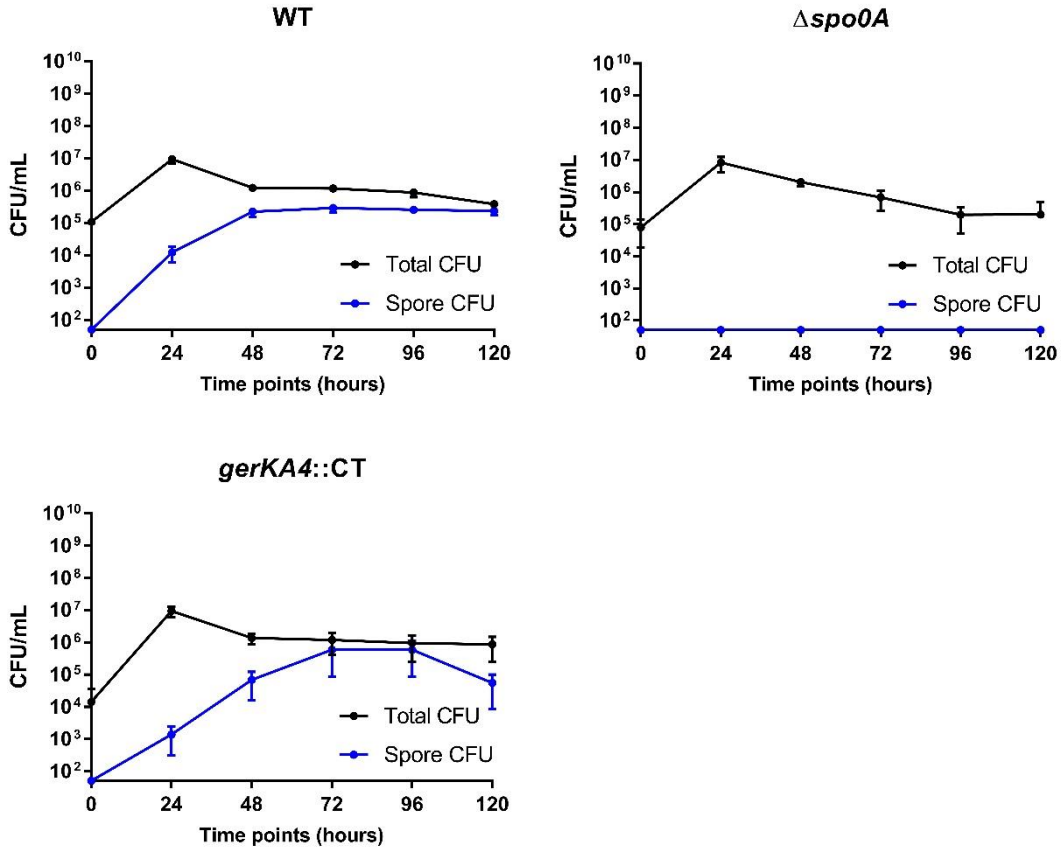


Figure 6.19: Sporulation profile of the *C. sporogenes-gerKA4*775s::CT strain. *C. sporogenes* WT and *C. sporogenes* Δ*spo0A* were included as positive and negative controls, respectively. All the cultures were synchronised at the beginning of the assay, and samples were collected at 24 hour intervals during a total period of 120 hours. For the total CFU count, the cultures were diluted and spotted into TYG agar. For spore CFU count, the cultures were heat-treated at 80°C for 20 minutes prior to the dilutions. Data show the mean and error bars represent the standard deviation of three biological and three technical replicates. The Y axis starts at the limit of detection (50 spores/mL).

6.4 Discussion

The results obtained in this chapter reinforce the importance of the sigma factors σ^F, σ^E, σ^G and σ^K for the process of sporulation. This is not surprising, as the sigma factors are highly conserved across spore formers and are known to regulate the transcription of the genes that mediate spore formation. The genes encoding the sigma factors were found to be essential for sporulation. The same was observed for genes involved in the activation of σ^F and σ^E. The early-stage sigma factors are present in the pre-divisional cell and must be activated only in the respective compartment: forespore for σ^F and mother cell for σ^E. SpoIIAA, SpoIIAB and SpoIIE are responsible for the activation of σ^F in the forespore upon completion of septum formation; SpoIIGA and SpoIIR, on the other hand, ensure activation of σ^E in the mother cell. The results obtained in this chapter suggest that the mechanism of activation of sigma factors in *C. sporogenes* is similar to that of *B. subtilis*. The importance of these regulation mechanisms is highlighted by the sporulation-defective phenotype of *spoIIAA*, *spoIIE* and *spoIIR* mutants in *C. sporogenes.*

Most of the characterised genes have been identified as relevant for sporulation in the TraDIS library. However, some of the in-frame deletion mutants were still able to produce spores to wild-type levels. This is possibly a result of polar effects of the transposon insertion. In the case of the M24 metallopeptidase (APP), the presence of other genes with similar functions could explain why the deletion of one copy had no effect on sporulation. The products of the other genes were possibly still able to carry out the same function, which is still not clear. The role of the protein encoded by the CLSPO_c33910 gene is unknown, but the protein might be localised in the membrane of the cell. Both *C. sporogenes* ΔAPP and *C. sporogenes* Δc33910 showed a slight growth disadvantage in comparison to wild type. Hence, it is possible that these transposon mutants were out-competed by the other strains during the 10-day long incubation of the cells. This would explain the lack of transposon insertions in these genes (Chapter 4). However, since the growth profile of the reverted strains was also affected, this phenotype might not be related to those genes. Further studies would be required to determine why these genes were wrongly identified as essential for sporulation or germination.

The lack of correlation between some of TraDIS data and the sporulation assay results highlight the importance of validation in high-throughput studies. Fitness disadvantage

or polar effects could lead to false positives in TraDIS data, which is one of the weaknesses of the approach. Nevertheless, the sporulation profile of most of the characterised genes (*sigE*, *sigF*, *sigG*, *sigK*, *spoIIAA*, *spoIIE*, *spoIIR* and *gerKA4*) was in line with the TraDIS results, reinforcing the usefulness of this approach in the identification of genes essential for sporulation/germination.

Chapter 7: General conclusions

7 Conclusions and future perspectives

The results presented in this thesis have provided unique insight into a range of molecular mechanisms that operate in *C. sporogenes* NCIMB 10696, with an emphasis on sporulation and germination. To aid these studies, initial work focussed on optimising DNA transfer frequencies and investigating the role any restriction modification (RM) systems may have in limiting their efficiency. Two RM systems were found in the genome of *C. sporogenes* NCIMB 10696, and different strategies were developed to circumvent any barrier to gene transfer they posed. Thus, it was demonstrated that targeting the type IV restriction enzyme in the recipient strain or using donors that were deficient in Dcm methylation represented two alternative approaches that could be used to improve the efficiency of conjugative plasmid transfer. The results appear to confirm the role of the type IV enzyme in the cleavage of cytosine-methylated DNA.

No improvements were made towards countering any possible effects of the type I system. As the recognition sequence of this system is unknown, it is not known whether the vector used to determine the efficiency of gene transfer carries a target sequence. In fact, a similar study showed that the recognition sites 1 and 2 tested in Chapter 3 are recognised by an equivalent type I system in the related *C. botulinum* strain ATCC 19397 and that the ClosTron disruption of its *hsdR* gene leads to increased conjugation efficiencies (Daphne Groothuis, personal communication). Two strategies were developed to evade the type I system of *C. sporogenes* NCIMB 10696 and can be tested once the recognition sequence is identified.

Other approaches may also be explored in the future to improve conjugation and transformation efficiencies in *C. sporogenes* NCIMB 10696. For instance, it has been demonstrated that heat treatment prior to conjugation increases the frequency of DNA transfer in *C. difficile*[203]. Likewise, *Corynebacterium glutamicum* shows higher transformation efficiencies when the competent cells are grown at a sub-optimal temperature and heat-shocked after transformation[204].

The major advance made in this study was towards the understanding of the processes of sporulation and germination in *C. sporogenes* NCIMB 10696. Thus, several genes that are essential for sporulation/germination were identified using TraDIS (Chapter 4) and the regulons of each sigma factor were identified using RNA-Seq (Chapter 5).

Moreover, some of these genes were further characterised and their role in sporulation was confirmed (Chapter 6). Genes that are essential for sporulation and regulated by sigma factors are of particular interest and are shown in Table 7.1. These are very likely to have a relevant role in sporulation, as they were linked to this process using two independent approaches.

Many of the identified genes are conserved in other spore formers and their involvement in sporulation has been demonstrated. Nevertheless, the role of other genes is less clear, and the work described here paves the way to future studies that might shed further light on this process. This work might also allow the annotation of some hypothetical proteins – in fact, a stage II sporulation protein (CLSPO_c30560) has been hypothesised to correspond to SpoIIP based on the results obtained in Chapters 4 and 5.

The sigma factors σ^F, σ^E, σ^G and σ^K appear to be essential for sporulation (Chapter 6) and the results obtained in Chapter 5 indicate that the regulons of each sigma factor have similar functions to those described in other species. The activity of each sigma factor appears to be partially independent from the others, which deviates from the *B. subtilis* model[34] but is in line with *C. difficile*[147, 157]. Examination of the sigma factor mutants using transmission electron microscopy (TEM) will allow the establishment of a link between the functions of the regulons and the morphological changes of the developing spore.

Moreover, some genes involved in the activation of the sigma factors in *C. sporogenes* were identified, but these mechanisms remain to be fully elucidated. It appears that the activation of σ^F is similar to other species and involves the proteins SpoIIAA, SpoIIAB and SpoIIE[34, 147]. In fact, the genes *spoIIAA* and *spoIIE* were targeted with CRISPR/Cas9 and ClosTron (respectively), and the resulting mutants were unable to grow after heat treatment. Moreover, although *spoIIAB* tolerated some transposon insertions in the sporulation library, the insertion index was lower than in the vegetative cells library ($log_2FC = -1.51$).

Regarding σ^E, its activation appears to be mediated by SpoIIGA and SpoIIR: both genes were identified as essential for sporulation/germination in Chapter 4 and the *spoIIR* mutant failed to sporulate. However, SpoIIR appears not to be absolutely required for σ^E activation because this sigma factor appears to be partially independent from σ^F. Alternatively, it is possible that *spoIIR* expression is not exclusively mediated by σ^F but also by Spo0A, as appears to be the case in *C. difficile*[147].

The transcription of *sigG* appears to be mediated by σ^E and not σ^F, which is not in line with *B. subtilis* and suggests that σ^E might be active in the forespore compartment. Alternatively, σ^E might indirectly regulate the transcription of *sigG*. However, these surprising results require validation by RT-qPCR. The products of the *spoIIIA* operon might mediate the activation of σ^G, and this hypothesis is supported by the fact that these genes seem to be essential for sporulation. The other protein that is involved in the activation of σ^G in *B. subtilis* is SpoIIIJ, but no orthologues were found in the genome of *C. sporogenes*. The protein SpoIIIJ is a putative lipoprotein, and two lipoproteins were identified as essential for sporulation/germination in the TraDIS data (CLSPO_c15060 and CLSPO_c26720). Further investigation of these proteins might clarify whether they are involved in the activation of σ^G.

Finally, the gene *sigK* appears to be transcriptionally regulated by σ^E and not to contain a *skin* element. In contrast with *C. difficile*, the activation of σ^K seems to involve the processing of a pro-sequence, and the proteins BofA and SpoIVB might have a role in this process. In fact, the gene *spoIVB* appears to be regulated by σ^G and essential for sporulation. On the other hand, BofA does not appear to be essential for spore formation. The potential involvement of other proteins in this process remains to be elucidated.

Based on the results discussed above, a mechanism can be proposed for the regulation and activation of sigma factors in *C. sporogenes* (Figure 7.1). A further elucidation of this mechanism can be achieved by confirming transcriptional regulation of each sigma factor by RT-qPCR. Moreover, analysis of the sigma factors by Western blot will determine whether they are present and/or processed in the different sigma factor mutants and in mutants of the proteins potentially involved in their activation.

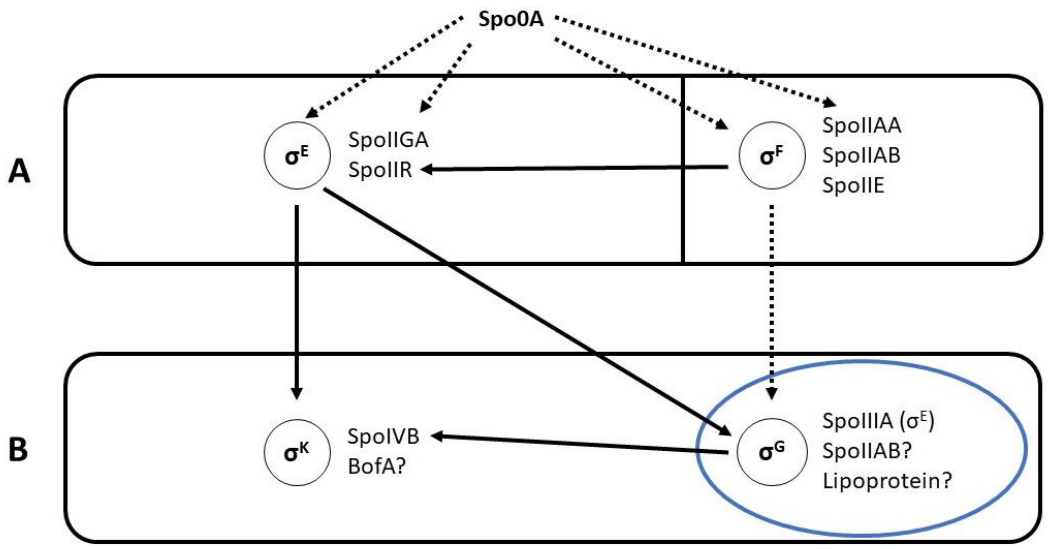


Figure 7.1: Schematic representation of the regulation of sigma factors in *C. sporogenes*. Dotted arrows represent transcriptional regulation that requires confirmation by RT-qPCR, while solid arrows represent transcriptional regulation confirmed by RNA-Seq. Next to each sigma factor is a list of proteins involved in their activation. Note that the proteins are next to the sigma factor they activate, but not necessarily in the compartment where they are expressed/active. **A:** The mother cell and forespore compartment appear separated by the septum, represented as a black line. σ^F is activated in the forespore upon asymmetric division completion. Then, σ^E is activated in the mother cell in response to the expression of SpoIIR in the forespore. The transcription of *spoIIR* is regulated by σ^F and possibly by Spo0A. Both σ^F and σ^E are transcriptionally regulated by Spo0A. **B:** The forespore is represented as a blue smaller cell inside the bigger mother cell. σ^E (and possibly σ^F) is responsible for the transcription of the gene encoding σ^G in the forespore. Following engulfment, σ^G is activated due to a signal from SpoIIIA (synthesised in the mother cell and regulated by σ^E). Likewise, the gene encoding σ^K is transcribed by σ^E in the mother cell and activated due to a signal from the forespore (SpoIVB).

Intriguingly, some genes that are usually involved in the translation and activation of sigma factors had insertion indexes above the essentiality cut-off in the sporulation library (Chapter 4), although the genes could tolerate fewer insertions in spores than in vegetative cells. Some of these genes were targeted with directed mutagenesis approaches in Chapter 6, namely *sigF*, *spoIIAA*, *spoIIE* and *spoIIR*. It was demonstrated that *sigF* and *spoIIAA* are in fact essential for sporulation, as observation under the microscope indicated that the mutant strains were unable to form spores. Regarding *spoIIE* and *spoIIR*, the sporulation assays resulted in no growth following heat treatment. As these genes are involved in the activation of σ^F and σ^E (respectively) and these sigma factors have been shown to be essential for sporulation (Chapter 6), it can be hypothesised that the *spoIIE* and *spoIIR* mutants are also unable to form spores. However, microscope observations are still required to confirm this hypothesis. These observations indicate that some genes with $\log_2FC < -2$ (spores relative to vegetative

cells) are essential for sporulation, despite tolerating some transposon insertions. Others, such as *gerKA4*, are not essential (Chapter 6). The TraDIS approach allowed the identification of many genes that play an important role in sporulation and germination, but each gene must be further analysed to confirm whether it is involved or essential for sporulation or germination.

It should be noted that some genes analysed in Chapter 6 did not confirm the TraDIS results, namely APP and CLSPO_c33910. It is unclear why these genes appeared to be essential for sporulation, but polar effects or slight growth disadvantage of the mutants are possible explanations. Moreover, the genes encoding the modification and restriction subunits of the type I RM system appear to be essential for sporulation and germination, respectively. The *hsdR* gene was targeted with ClosTron and the mutant obtained was found to be unaffected in sporulation, which is not in accordance with the TraDIS results. It is likely that these genes were identified as essential due to the fact that the *hsdR* mutant buries into the agar and its DNA was probably not obtained for subsequent sequence analysis.

Overall, the main goal of this thesis was achieved and several genes herein identified can be used to generate a conditionally sporulating strain. The requirement of some of these genes in sporulation was confirmed in Chapter 6, so the genes *sigF*, *sigE*, *sigG*, *sigK*, *spoIIAA*, *spoIIE* and *spoIIR* are potential candidates. In fact, the use of *spoIIR* as the target gene in a conditionally sporulating strain has already been tested, but the gene was shown to be expressed even in the absence of inducer, and wild type levels of expression were not achieved after induction[78]. This highlights the need for several targets, as sporulation appears to be sensitive to gene expression levels[78] and the ideal candidate remains to be found. In addition to the genes investigated in Chapter 6, more targets can be selected from the TraDIS results and further analysed.

The work presented in this thesis shed some light on the sporulation and germination mechanisms in *C. sporogenes*, but it also paves the way for future work to further elucidate these processes.

Table 7.1: List of genes essential for sporulation/germination identified with TraDIS and RNA-Seq. Essential genes have an insertion index below the essentiality threshold in at least one biological replicate, or $Log_2FC < -2$ and q-value < 0.05. Genes regulated by a sigma factor have $Log_2FC < -1$ and adjusted p-value < 0.05.

Locus tag	Gene name	Gene function	Essential for	Regulated by
CLSPO_c00800	CLSPO_c00800	inhibitor of sigma-G Gin	Sporulation	σ^K
CLSPO_c01080	CLSPO_c01080	hypothetical protein	Sporulation	σ^E
CLSPO_c01090	CLSPO_c01090	hypothetical protein	Sporulation	σ^E
CLSPO_c01280	*manC*	mannose-1-phosphate guanylyltransferase ManC	Sporulation	σ^E
CLSPO_c01340	CLSPO_c01340	putative mannosyltransferase	Germination	σ^E
CLSPO_c01350	CLSPO_c01350	CotS family spore coat protein	Sporulation	σ^K
CLSPO_c01360	CLSPO_c01360	hypothetical protein	Sporulation	σ^E
CLSPO_c01370	CLSPO_c01370	hypothetical protein	Germination	σ^E
CLSPO_c01390	CLSPO_c01390	glycosyl transferase, group 1 family protein	Germination	σ^K
CLSPO_c01400	CLSPO_c01400	putative spore coat protein	Sporulation	σ^E
CLSPO_c01430	CLSPO_c01430	LysM domain-containing protein	Sporulation	σ^E
CLSPO_c01480	*gerKA4*	spore germination protein KA	Germination	σ^G
CLSPO_c01490	CLSPO_c01490	spore germination protein	Germination	σ^G
CLSPO_c01500	CLSPO_c01500	germination protein, GerC family	Germination	σ^G
CLSPO_c01510	*spoIIR*	stage II sporulation protein R	Sporulation	σ^F/σ^G
CLSPO_c01880	CLSPO_c01880	hypothetical protein DUF1779	Sporulation	σ^K
CLSPO_c01890	*murA2*	UDP-N-acetylglucosamine 1-carboxyvinyltransferase MurA	Sporulation	σ^K
CLSPO_c01900	*spoIID*	stage II sporulation protein D	Sporulation	σ^E
CLSPO_c01920	*spoIIID*	stage III sporulation protein D	Sporulation	σ^E
CLSPO_c02270	CLSPO_c02270	hypothetical protein DUF4363	Germination	σ^G

Table 7.1: List of genes essential for sporulation/germination identified with TraDIS and RNA-Seq (continuation).

Locus tag	Gene name	Gene function	Essential for	Regulated by
CLSPO_c02280	CLSPO_c02280	hypothetical protein	Germination	σ^G
CLSPO_c02320	CLSPO_c02320	CBS domain-containing protein	Germination	σ^K
CLSPO_c08250	CLSPO_c08250	hypothetical protein	Sporulation	σ^G
CLSPO_c13320	CLSPO_c13320	hypothetical protein	Germination	σ^G
CLSPO_c15070	CLSPO_c15070	pseudouridine synthase	Sporulation	σ^G
CLSPO_c16610	*ytaF*	putative sporulation protein YtaF	Sporulation	σ^K
CLSPO_c17560	CLSPO_c17560	hypothetical protein	Sporulation	σ^K
CLSPO_c17600	*safA*	spore coat assembly protein SafA	Germination	σ^G
CLSPO_c17810	CLSPO_c17810	small, acid-soluble spore protein beta	Germination	σ^G
CLSPO_c17880	*miaA*	tRNA dimethylallyltransferase MiaA	Sporulation	σ^E
CLSPO_c18470	*dacB1*	D-alanyl-D-alanine carboxypeptidase DacB	Germination	σ^G
CLSPO_c18600	*spoIIM*	stage II sporulation protein M	Sporulation	σ^E/σ^G
CLSPO_c18660	*spoIVB*	SpoIVB peptidase	Sporulation	σ^G
CLSPO_c18810	*spoIIIAH*	stage III sporulation protein AH	Sporulation	σ^E
CLSPO_c18820	*spoIIIAG*	stage III sporulation protein AG	Sporulation	σ^E
CLSPO_c18830	*spoIIIAF*	stage III sporulation protein AF	Sporulation	σ^E
CLSPO_c18840	*spoIIIAE*	stage III sporulation protein AE	Sporulation	σ^E
CLSPO_c18850	*spoIIIAD*	stage III sporulation protein AD	Sporulation	σ^E
CLSPO_c18860	*spoIIIAC*	stage III sporulation protein AC	Sporulation	σ^E
CLSPO_c18870	*spoIIIAB*	stage III sporulation protein AB	Sporulation	σ^E
CLSPO_c18880	*spoIIIAA*	stage III sporulation protein AA	Sporulation	σ^E

Table 7.1: List of genes essential for sporulation/germination identified with TraDIS and RNA-Seq (continuation).

Locus tag	Gene name	Gene function	Essential for	Regulated by
CLSPO_c19770	*gerKA2*	spore germination protein KA	Germination	σ^G
CLSPO_c19800	CLSPO_c19800	spore germination protein	Germination	σ^E
CLSPO_c19850	CLSPO_c19850	hypothetical protein	Germination	σ^K
CLSPO_c20600	*codA*	cytosine deaminase CodA	Germination	σ^F
CLSPO_c21240	*eutN*	ethanolamine utilization protein EutN	Germination	σ^F
CLSPO_c21250	CLSPO_c21250	BMC domain-containing protein	Sporulation	σ^F
CLSPO_c21320	*adhE3*	aldehyde-alcohol dehydrogenase AdhE	Germination	σ^F
CLSPO_c22000	CLSPO_c22000	M24 family metallopeptidase	Germination	σ^E
CLSPO_c22010	CLSPO_c22010	M24 family metallopeptidase	Germination	σ^E
CLSPO_c22240	*lcdC*	activator of lactoyl-CoA dehydratase LcdC	Sporulation	σ^E
CLSPO_c22260	*etfA1*	electron transfer flavoprotein subunit alpha	Sporulation	σ^E
CLSPO_c22270	*etfB1*	electron transfer flavoprotein subunit beta	Sporulation	σ^E
CLSPO_c22290	CLSPO_c22290	benzoyl-CoA reductase/2-hydroxyglutaryl-CoA dehydratase subunit, BcrC/BadD/HgdB	Sporulation	σ^E
CLSPO_c22300	CLSPO_c22300	benzoyl-CoA reductase/2-hydroxyglutaryl-CoA dehydratase subunit, BcrC/BadD/HgdB	Sporulation	σ^E
CLSPO_c22310	CLSPO_c22310	putative acyl-CoA transferases/carnitine dehydratase	Sporulation	σ^E
CLSPO_c22950	CLSPO_c22950	amino acid permease family protein	Germination	σ^F
CLSPO_c24960	CLSPO_c24960	hypothetical protein	Sporulation	σ^E
CLSPO_c25100	CLSPO_c25100	PRC-barrel domain-containing protein	Sporulation	σ^E
CLSPO_c25860	CLSPO_c25860	nucleoside recognition domain-containing protein	Germination	σ^E

Table 7.1: List of genes essential for sporulation/germination identified with TraDIS and RNA-Seq (continuation).

Locus tag	Gene name	Gene function	Essential for	Regulated by
CLSPO_c26120	*spoIVA*	stage IV sporulation protein A	Sporulation	σ^E
CLSPO_c26260	CLSPO_c26260	PRC-barrel domain-containing protein	Sporulation	σ^E
CLSPO_c26270	*sigG*	RNA polymerase sigma-G factor SigG	Sporulation	σ^E
CLSPO_c26360	*sigK*	RNA polymerase sigma-28 factor SigK	Sporulation	σ^E
CLSPO_c26720	CLSPO_c26720	putative lipoprotein	Germination	σ^E
CLSPO_c27120	CLSPO_c27120	putative calcium-translocating P-type ATPase, PMCA-type	Sporulation	σ^E
CLSPO_c27140	CLSPO_c27140	hypothetical protein	Sporulation	σ^E
CLSPO_c28700	CLSPO_c28700	spore germination protein, GerAC	Germination	σ^E
CLSPO_c28720	*gerKA3*	spore germination protein KA	Germination	σ^E
CLSPO_c30310	CLSPO_c30310	spore coat protein-like protein	Sporulation	σ^E
CLSPO_c30420	*yqfC*	sporulation protein YqfC	Sporulation	σ^E
CLSPO_c30560	CLSPO_c30560	stage II sporulation protein	Sporulation	σ^E
CLSPO_c30570	*gpr*	germination protease Gpr	Germination	σ^G
CLSPO_c30850	CLSPO_c30850	putative metalloprotease	Germination	σ^E
CLSPO_c30980	CLSPO_c30980	hypothetical protein	Sporulation	σ^E
CLSPO_c31780	CLSPO_c31780	SpoVA family protein	Germination	σ^G
CLSPO_c31790	*spoVAD*	stage V sporulation protein AD	Germination	σ^G
CLSPO_c31800	*spoVAC*	stage V sporulation protein AC	Germination	σ^G
CLSPO_c31860	CLSPO_c31860	polysaccharide deacetylase family protein	Germination	σ^G
CLSPO_c32100	*pgcA1*	phosphoglucomutase PgcA	Germination	σ^K
CLSPO_c32220	CLSPO_c32220	hypothetical protein	Sporulation	σ^K

Table 7.1: List of genes essential for sporulation/germination identified with TraDIS and RNA-Seq (continuation).

Locus tag	Gene name	Gene function	Essential for	Regulated by
CLSPO_c32230	CLSPO_c32230	hypothetical protein	Sporulation	σ^E
CLSPO_c32250	CLSPO_c32250	beta 1,4 glucosyltransferase	Sporulation	σ^K
CLSPO_c33260	CLSPO_c33260	hypothetical protein	Sporulation	σ^K
CLSPO_c33270	CLSPO_c33270	hypothetical protein	Sporulation	σ^K
CLSPO_c33280	CLSPO_c33280	hypothetical protein	Sporulation	σ^K
CLSPO_c33840	CLSPO_c33840	hypothetical protein	Sporulation	σ^E
CLSPO_c34630	CLSPO_c34630	hypothetical protein	Germination	σ^K
CLSPO_c36050	CLSPO_c36050	N-acetylmuramoyl-L-alanine amidase	Germination	σ^E
CLSPO_c37140	*yabQ*	spore cortex biosynthesis protein YabQ	Sporulation	σ^E
CLSPO_c37150	*yabP*	sporulation protein YabP	Germination	σ^E
CLSPO_c37190	*spoVB2*	stage V sporulation protein B	Germination	σ^E
CLSPO_c37200	*spoVT*	stage V sporulation protein T	Germination	σ^G
CLSPO_c37390	CLSPO_c37390	putative phage head-tail adaptor	Sporulation	σ^E

Bibliography

1. Setlow, P., *I will survive: DNA protection in bacterial spores.* Trends in Microbiology, 2007. **15**(4): p. 172-180.
2. Nicholson, W.L., et al., *Resistance of Bacillus endospores to extreme terrestrial and extraterrestrial environments.* Microbiology and Molecular Biology Reviews, 2000. **64**(3): p. 548-572.
3. Kennedy, M.J., S.L. Reader, and L.M. Swierczynski, *Preservation records of micro-organisms: evidence of the tenacity of life.* Microbiology, 1994. **140**(10): p. 2513-2529.
4. Cano, R. and M. Borucki, *Revival and identification of bacterial spores in 25- to 40-million-year-old Dominican amber.* Science, 1995. **268**(5213): p. 1060-1064.
5. Setlow, P., *Spore germination.* Current Opinion in Microbiology, 2003. **6**(6): p. 550-556.
6. Galperin, M.Y., *Genomic determinants of sporulation in Bacilli and Clostridia: towards the minimal set of sporulation-specific genes.* Environ. Microbiol., 2012. **14**: p. 2870-2890.
7. Collins, M.D., et al., *The phylogeny of the genus Clostridium: proposal of five new genera and eleven new species combinations.* International Journal of Systematic and Evolutionary Microbiology, 1994. **44**(4): p. 812-826.
8. Ludwig, W., K. Schleifer, and W.B. Whitman, *Bacilli class. nov*, in *Bergey's Manual of Systematics of Archaea and Bacteria*. 2015. p. 1-1.
9. de Hoon, M.J., P. Eichenberger, and D. Vitkup, *Hierarchical evolution of the bacterial sporulation network.* Curr. Biol., 2010. **20**: p. R735-R745.
10. Fimlaid, K.A. and A. Shen, *Diverse mechanisms regulate sporulation sigma factor activity in the Firmicutes.* Current Opinion in Microbiology, 2015. **24**: p. 88-95.
11. Britton, R.A., et al., *Genome-wide analysis of the stationary-phase sigma factor (sigma-H) regulon of Bacillus subtilis.* Journal of Bacteriology, 2002. **184**(17): p. 4881-4890.
12. Eichenberger, P., et al., *The program of gene transcription for a single differentiating cell type during sporulation in Bacillus subtilis.* PLoS Biology, 2004. **2**(10): p. 1664-1683.
13. Molle, V., et al., *The Spo0A regulon of Bacillus subtilis.* Molecular Microbiology, 2003. **50**(5): p. 1683-1701.
14. Paredes-Sabja, D., A. Shen, and J.A. Sorg, *Clostridium difficile spore biology: sporulation, germination, and spore structural proteins.* Trends in Microbiology, 2014. **22**(7): p. 406-416.
15. Paredes, C.J., K.V. Alsaker, and E.T. Papoutsakis, *A comparative genomic view of clostridial sporulation and physiology.* Nature Reviews Microbiology, 2005. **3**(12): p. 969-978.
16. De Spain Smith, L. and H. Sugiyama, *Botulism: the organism, its toxins, the disease*. 1988: Charles C. Thomas Publisher, Limited.
17. Wells, C.L. and T.D. Wilkins, *Clostridia: sporeforming anaerobic bacilli*, in *Medical Microbiology, 4th Edition*, S. Baron, Editor. 1996, Galveston (TX).
18. Joyce, A.M. and D.L. Burns, *Recurrent Clostridium difficile colitis.* Comprehensive Therapy, 2004. **30**(3): p. 160-163.

19. Lamed, R. and J.G. Zeikus, *Ethanol production by thermophilic bacteria: relationship between fermentation product yields of and catabolic enzyme activities in Clostridium thermocellum and Thermoanaerobium brockii*. Journal of Bacteriology, 1980. **144**(2): p. 569-578.
20. Napoli, F., et al., *Butanol production by Clostridium acetobutylicum in a continuous packed bed reactor.* Journal of Industrial Microbiology & Biotechnology, 2010. **37**(6): p. 603-608.
21. Tran, H.G., et al., *Biocatalytic production of novel glycolipids with cellodextrin phosphorylase.* Bioresource Technology, 2012. **115**: p. 84-87.
22. Desvaux, M., E. Guedon, and H. Petitdemange, *Cellulose catabolism by Clostridium cellulolyticum growing in batch culture on defined medium.* Applied and Environmental Microbiology, 2000. **66**(6): p. 2461-2470.
23. Kubiak, A., *The use of Clostridium sporogenes in directed enzyme pro-drug therapy*. 2013, University of Nottingham.
24. Kubiak, A.M., et al., *Complete genome sequence of the nonpathogenic soil-dwelling bacterium Clostridium sporogenes strain NCIMB 10696.* Genome Announcements, 2015. **3**(4).
25. Brown, K.L., *Control of bacterial spores.* British Medical Bulletin, 2000. **56**(1): p. 158-171.
26. Brown, J.L., N. Tran-Dinh, and B. Chapman, *Clostridium sporogenes PA 3679 and its uses in the derivation of thermal processing schedules for low-acid shelf-stable foods and as a research model for proteolytic Clostridium botulinum.* Journal of Food Protection, 2012. **75**(4): p. 779-792.
27. Kubiak, A.M. and N.P. Minton, *The potential of clostridial spores as therapeutic delivery vehicles in tumour therapy.* Research in Microbiology, 2015. **166**(4): p. 244-254.
28. Karnofsky, D.A., *Mechanisms of action of anticancer drugs at a cellular level.* Ca: A Cancer Journal For Clinicians, 1968. **18**(4): p. 232-234.
29. Rautio, J., et al., *Prodrugs: design and clinical applications.* Nat Rev Drug Discov, 2008. **7**(3): p. 255-270.
30. Heap, J.T., et al., *Spores of Clostridium engineered for clinical efficacy and safety cause regression and cure of tumors in vivo.* Oncotarget, 2014. **5**(7): p. 1761-1769.
31. Weinmann, M., C. Belka, and L. Plasswilm, *Tumour hypoxia: impact on biology, prognosis and treatment of solid malignant tumours.* Oncology Research and Treatment, 2004. **27**(1): p. 83-90.
32. Möse, J.R. and G. Möse, *Oncolysis by clostridia. I. Activity of Clostridium butyricum (M-55) and other nonpathogenic clostridia against the Ehrlich carcinoma.* Cancer Research, 1964. **24**(2 Part 1): p. 212-216.
33. Thiele, E.H., R.N. Arison, and G.E. Boxer, *Oncolysis by clostridia. III. Effects of clostridia and chemotherapeutic agents on rodent tumors.* Cancer Research, 1964. **24**(2 Part 1): p. 222-233.
34. Hilbert, D.W. and P.J. Piggot, *Compartmentalization of gene expression during Bacillus subtilis spore formation.* Microbiology and Molecular Biology Reviews, 2004. **68**(2): p. 234-262.
35. Errington, J., *Regulation of endospore formation in Bacillus subtilis.* Nature Reviews Microbiology, 2003. **1**(2): p. 117-126.
36. Levin, P.A. and R. Losick, *Transcription factor Spo0A switches the localization of the cell division protein FtsZ from a medial to a bipolar pattern in Bacillus subtilis.* Genes & Development, 1996. **10**(4): p. 478-488.

37. Piggot, P.J. and D.W. Hilbert, *Sporulation of Bacillus subtilis.* Current Opinion in Microbiology, 2004. **7**(6): p. 579-586.
38. Frandsen, N., et al., *Transient gene asymmetry during sporulation and establishment of cell specificity in Bacillus subtilis.* Genes & development, 1999. **13**(4): p. 394-399.
39. Camp, A.H. and R. Losick, *A feeding tube model for activation of a cell-specific transcription factor during sporulation in Bacillus subtilis.* Genes & Development, 2009. **23**(8): p. 1014-1024.
40. Tipper, D.J. and P.E. Linnett, *Distribution of peptidoglycan synthetase activities between sporangia and forespores in sporulating cells of Bacillus sphaericus.* Journal of Bacteriology, 1976. **126**(1): p. 213-221.
41. Popham, D.L., et al., *Analysis of the peptidoglycan structure of Bacillus subtilis endospores.* Journal of Bacteriology, 1996. **178**(22): p. 6451-8.
42. Beaman, T.C. and P. Gerhardt, *Heat resistance of bacterial spores correlated with protoplast dehydration, mineralization, and thermal adaptation.* Applied and Environmental Microbiology, 1986. **52**(6): p. 1242-1246.
43. Driks, A., *Bacillus subtilis spore coat.* Microbiol. Mol. Biol. Rev., 1999. **63**: p. 1-20.
44. Stewart, G.C., *The exosporium layer of bacterial spores: a connection to the environment and the infected host.* Microbiology and molecular biology reviews : MMBR, 2015. **79**(4): p. 437-457.
45. Meaney, C.A., et al., *The role of small acid-soluble proteins (SASPs) in protection of spores of Clostridium botulinum against nitrous acid.* International Journal of Food Microbiology, 2016. **216**: p. 25-30.
46. Sonenshein, A.L., *Control of sporulation initiation in Bacillus subtilis.* Current Opinion in Microbiology, 2000. **3**(6): p. 561-566.
47. Pettit, L.J., et al., *Functional genomics reveals that Clostridium difficile Spo0A coordinates sporulation, virulence and metabolism.* BMC Genomics, 2014. **15**(1): p. 1-15.
48. Harris, L.M., N.E. Welker, and E.T. Papoutsakis, *Northern, morphological, and fermentation analysis of spo0A inactivation and overexpression in Clostridium acetobutylicum ATCC 824.* Journal of Bacteriology, 2002. **184**(13): p. 3586-3597.
49. Huang, I.H., et al., *Disruption of the gene (spo0A) encoding sporulation transcription factor blocks endospore formation and enterotoxin production in enterotoxigenic Clostridium perfringens type A.* FEMS Microbiology Letters, 2004. **233**(2): p. 233-240.
50. Wilkinson, S.R., et al., *Molecular genetics and the initiation of solventogenesis in Clostridium beijerinckii (formerly Clostridium acetobutylicum) NCIMB 8052.* FEMS Microbiology Reviews, 1995. **17**(3): p. 275-285.
51. Nölling, J., et al., *Genome sequence and comparative analysis of the solvent-producing bacterium Clostridium acetobutylicum.* Journal of Bacteriology, 2001. **183**(16): p. 4823-4838.
52. Dürre, P. and C. Hollergschwandner, *Initiation of endospore formation in Clostridium acetobutylicum.* Anaerobe, 2004. **10**(2): p. 69-74.
53. Alsaker, K.V. and E.T. Papoutsakis, *Transcriptional program of early sporulation and stationary-phase events in Clostridium acetobutylicum.* Journal of Bacteriology, 2005. **187**(20): p. 7103-7118.

54. Lukat, G.S., et al., *Phosphorylation of bacterial response regulator proteins by low molecular weight phospho-donors.* Proceedings of the National Academy of Sciences of the United States of America, 1992. **89**(2): p. 718-722.
55. Zhao , Y., et al., *Intracellular butyryl phosphate and acetyl phosphate concentrations in Clostridium acetobutylicum and their implications for solvent formation.* Applied and Environmental Microbiology, 2005. **71**(1): p. 530-537.
56. Underwood, S., et al., *Characterization of the sporulation initiation pathway of Clostridium difficile and its role in toxin production.* Journal of bacteriology, 2009. **191**(23): p. 7296-7305.
57. Fujita, M. and R. Losick, *An investigation into the compartmentalization of the sporulation transcription factor σ^E in Bacillus subtilis.* Molecular Microbiology, 2002. **43**(1): p. 27-38.
58. Steil, L., et al., *Genome-wide analysis of temporally regulated and compartment-specific gene expression in sporulating cells of Bacillus subtilis.* Microbiology, 2005. **151**(2): p. 399-420.
59. Li, Z. and P.J. Piggot, *Development of a two-part transcription probe to determine the completeness of temporal and spatial compartmentalization of gene expression during bacterial development.* Proceedings of the National Academy of Sciences, 2001. **98**(22): p. 12538-12543.
60. Karow, M.L., P. Glaser, and P.J. Piggot, *Identification of a gene, spoIIR, that links the activation of sigma E to the transcriptional activity of sigma F during sporulation in Bacillus subtilis.* Proceedings of the National Academy of Sciences, 1995. **92**(6): p. 2012-2016.
61. Moir, A., *How do spores germinate?* Journal of Applied Microbiology, 2006. **101**(3): p. 526-530.
62. Huang, I.-H., et al., *Clostridium perfringens : sporulation, spore resistance and germination.* Bangladesh Journal of Microbiology, 2008. **24**(1): p. 8.
63. Hudson, K.D., et al., *Localization of GerAA and GerAC germination proteins in the Bacillus subtilis spore.* Journal of Bacteriology, 2001. **183**(14): p. 4317-4322.
64. Behravan, J., et al., *Mutations in the gerP locus of Bacillus subtilis and Bacillus cereus affect access of germinants to their targets in spores.* Journal of Bacteriology, 2000. **182**(7): p. 1987-1994.
65. Xiao, Y., et al., *Clostridial spore germination versus bacilli: genome mining and current insights.* Food Microbiology, 2011. **28**(2): p. 266-274.
66. Moir, A., E. Lafferty, and D.A. Smith, *Genetic analysis of spore germination mutants of Bacillus subtilis 168: the correlation of phenotype with map location.* Microbiology, 1979. **111**(1): p. 165-180.
67. Broussolle, V., et al., *Molecular and physiological characterisation of spore germination in Clostridium botulinum and C. sporogenes.* Anaerobe, 2002. **8**(3): p. 89-100.
68. Thackray, P.D., et al., *GerN, an antiporter homologue important in germination of Bacillus cereus endospores.* Journal of Bacteriology, 2001. **183**(2): p. 476-482.
69. Paredes-Sabja, D., P. Setlow, and M.R. Sarker, *GerO, a putative Na+/H+-K+ antiporter, is essential for normal germination of spores of the pathogenic bacterium Clostridium perfringens.* Journal of Bacteriology, 2009. **191**(12): p. 3822-3831.

70. Shah, I.M., et al., *A eukaryotic-like Ser/Thr kinase signals bacteria to exit dormancy in response to peptidoglycan fragments.* Cell, 2008. **135**(3): p. 486-496.
71. Boland, F.M., et al., *Complete spore-cortex hydrolysis during germination of Bacillus subtilis 168 requires SleB and YpeB.* Microbiology, 2000. **146**(1): p. 57-64.
72. Meaney, C.A., et al., *Optimal spore germination in Clostridium botulinum ATCC 3502 requires the presence of functional copies of SleB and YpeB, but not CwlJ.* Anaerobe, 2015. **34**: p. 86-93.
73. Olguín-Araneda, V., et al., *Recent advances in germination of Clostridium spores.* Research in Microbiology, 2015. **166**(4): p. 236-243.
74. Burns, D.A., J.T. Heap, and N.P. Minton, *SleC is essential for germination of Clostridium difficile spores in nutrient-rich medium supplemented with the bile salt taurocholate.* Journal of Bacteriology, 2010. **192**(3): p. 657-664.
75. Paredes-Sabja, D., P. Setlow, and M.R. Sarker, *SleC is essential for cortex peptidoglycan hydrolysis during germination of spores of the pathogenic bacterium Clostridium perfringens.* Journal of Bacteriology, 2009. **191**(8): p. 2711-2720.
76. Shimamoto, S., et al., *Partial characterization of an enzyme fraction with protease activity which converts the spore peptidoglycan hydrolase (SleC) precursor to an active enzyme during germination of Clostridium perfringens S40 spores and analysis of a gene cluster involved in the activity.* Journal of Bacteriology, 2001. **183**(12): p. 3742-3751.
77. Dembek, M., et al., *Inducible expression of spo0A as a universal tool for studying sporulation in Clostridium difficile.* Frontiers in microbiology, 2017. **8**: p. 1793-1793.
78. Cañadas, I.C., *Novel genetic tools for Clostridia and a synthetic biology approach to control spore formation in process organisms*. 2018, University of Nottingham.
79. Purdy, D., et al., *Conjugative transfer of clostridial shuttle vectors from Escherichia coli to Clostridium difficile through circumvention of the restriction barrier.* Molecular Microbiology, 2002. **46**(2): p. 439-452.
80. Horton, R.M., *In vitro recombination and mutagenesis of DNA*, in *PCR Cloning Protocols: From Molecular Cloning to Genetic Engineering*, B.A. White, Editor. 1997, Humana Press: Totowa, NJ. p. 141-150.
81. Cartman, S.T. and N.P. Minton, *A mariner-based transposon system for in vivo random mutagenesis of Clostridium difficile.* Applied and Environmental Microbiology, 2010. **76**(4): p. 1103-1109.
82. Tellmann, G., *The E-Method: a highly accurate technique for gene-expression analysis.* Nature Methods, 2006. **3**(7): p. i-ii.
83. Naghili, H., et al., *Validation of drop plate technique for bacterial enumeration by parametric and nonparametric tests.* Veterinary Research Forum, 2013. **4**(3): p. 179-183.
84. Heap, J.T., et al., *The ClosTron: a universal gene knock-out system for the genus Clostridium.* Journal of Microbiological Methods, 2007. **70**(3): p. 452-464.
85. Kuehne, S.A., et al., *ClosTron-mediated engineering of Clostridium*, in *Strain Engineering: Methods and Protocols*, A.J. Williams, Editor. 2011, Humana Press: Totowa, NJ. p. 389-407.

86. Ingle, P., et al., *Generation of a fully erythromycin-sensitive strain of Clostridioides difficile using a novel CRISPR-Cas9 genome editing system.* Scientific Reports, 2019. **9**(1): p. 8123.
87. Cañadas, I.C., et al., *RiboCas: A universal CRISPR-based editing tool for Clostridium.* ACS Synthetic Biology, 2019.
88. Barquist, L., et al., *The TraDIS toolkit: sequencing and analysis for dense transposon mutant libraries.* Bioinformatics (Oxford, England), 2016. **32**(7): p. 1109-1111.
89. Jiang, H., et al., *Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads.* BMC bioinformatics, 2014. **15**: p. 182-182.
90. Kim, D., B. Langmead, and S.L. Salzberg, *HISAT: a fast spliced aligner with low memory requirements.* Nature Methods, 2015. **12**: p. 357.
91. Anders, S. and W. Huber, *Differential expression analysis for sequence count data.* Genome Biol, 2010. **11**.
92. Langridge, G.C., et al., *Simultaneous assay of every Salmonella Typhi gene using one million transposon mutants.* Genome Research, 2009. **19**(12): p. 2308-2316.
93. van Opijnen, T. and A. Camilli, *Transposon insertion sequencing: a new tool for systems-level analysis of microorganisms.* Nature Reviews Microbiology, 2013. **11**(7): p. 435-442.
94. Bickle, T.A. and D.H. Krüger, *Biology of DNA restriction.* Microbiological Reviews, 1993. **57**(2): p. 434-450.
95. Murray, N.E., *Type I restriction systems: sophisticated molecular machines (a legacy of Bertani and Weigle).* Microbiology and Molecular Biology Reviews, 2000. **64**(2): p. 412-434.
96. Janscak, P., U. Sandmeier, and T.A. Bickle, *Single amino acid substitutions in the HsdR subunit of the type IB restriction enzyme EcoAI uncouple the DNA translocation and DNA cleavage activities of the enzyme.* Nucleic Acids Research, 1999. **27**(13): p. 2638-2643.
97. Davies, G.P., et al., *EcoKI with an amino acid substitution in any one of seven DEAD-box motifs has impaired ATPase and endonuclease activities.* Nucleic Acids Research, 1998. **26**(21): p. 4828-4836.
98. Humbert, O., M.S. Dorer, and N.R. Salama, *Characterization of Helicobacter pylori factors that control transformation frequency and integration length during inter-strain DNA recombination.* Molecular Microbiology, 2011. **79**(2): p. 387-401.
99. Corvaglia, A.R., et al., *A type III-like restriction endonuclease functions as a major barrier to horizontal gene transfer in clinical Staphylococcus aureus strains.* Proceedings of the National Academy of Sciences, 2010. **107**(26): p. 11954-11958.
100. Lesiak, J.M., W. Liebl, and A. Ehrenreich, *Development of an in vivo methylation system for the solventogen Clostridium saccharobutylicum NCP 262 and analysis of two endonuclease mutants.* Journal of Biotechnology, 2014. **188**(Supplement C): p. 97-99.
101. Roberts, R.J., et al., *REBASE—a database for DNA restriction and modification: enzymes, genes and genomes.* Nucleic Acids Research, 2014. **43**(D1): p. D298-D299.
102. Marchler-Bauer, A., et al., *CDD: NCBI's conserved domain database.* Nucleic acids research, 2015. **43**(Database issue): p. D222-D226.

103. Xu, S.-y., et al., *A type IV modification-dependent restriction enzyme SauUSI from Staphylococcus aureus subsp. aureus USA300.* Nucleic Acids Research, 2011. **39**(13): p. 5597-5610.
104. Dong, H., et al., *Engineering Clostridium strain to accept unmethylated DNA.* PLOS ONE, 2010. **5**(2): p. e9038.
105. Zhang, G., et al., *A mimicking-of-DNA-methylation-patterns pipeline for overcoming the restriction barrier of bacteria.* PLOS Genetics, 2012. **8**(9): p. e1002987.
106. Kuehne, S.A. and N.P. Minton, *ClosTron-mediated engineering of Clostridium.* Bioengineered, 2012. **3**(4): p. 247-254.
107. Barrangou, R., et al., *CRISPR provides acquired resistance against viruses in prokaryotes.* Science, 2007. **315**(5819): p. 1709-1712.
108. Jansen, R., et al., *Identification of genes that are associated with DNA repeats in prokaryotes.* Molecular Microbiology, 2002. **43**(6): p. 1565-1575.
109. Doudna, J.A. and E. Charpentier, *The new frontier of genome engineering with CRISPR-Cas9.* Science, 2014. **346**(6213).
110. Jinek, M., et al., *A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity.* Science, 2012. **337**(6096): p. 816-821.
111. Wang, Y., et al., *Bacterial genome editing with CRISPR-Cas9: deletion, integration, single nucleotide modification, and desirable "clean" mutant selection in Clostridium beijerinckii as an example.* ACS Synthetic Biology, 2016. **5**(7): p. 721-732.
112. Woods, C., et al., *A novel conjugal donor strain for improved DNA transfer into Clostridium spp.* Anaerobe, 2019.
113. Heap, J.T., et al., *A modular system for Clostridium shuttle plasmids.* Journal of Microbiological Methods, 2009. **78**(1): p. 79-85.
114. Dembek, M., et al., *High-throughput analysis of gene essentiality and sporulation in Clostridium difficile.* mBio, 2015. **6**(2): p. e02383-14.
115. Chopra, I. and M. Roberts, *Tetracycline antibiotics: mode of action, applications, molecular biology, and epidemiology of bacterial resistance.* Microbiology and Molecular Biology Reviews, 2001. **65**(2): p. 232-260.
116. Eckert, B. and C.F. Beck, *Overproduction of transposon Tn10-encoded tetracycline resistance protein results in cell death and loss of membrane potential.* Journal of Bacteriology, 1989. **171**(6): p. 3557-3559.
117. Hillen, W. and C. Berens, *Mechanisms underlying expression of Tn10 encoded tetracycline resistance.* Annual Review of Microbiology, 1994. **48**(1): p. 345-369.
118. Corrigan, R.M. and T.J. Foster, *An improved tetracycline-inducible expression vector for Staphylococcus aureus.* Plasmid, 2009. **61**(2): p. 126-129.
119. Zhang, Y., A. Grosse-Honebrink, and N.P. Minton, *A universal mariner transposon system for forward genetic studies in the genus Clostridium.* PLoS ONE, 2015. **10**(4): p. e0122411.
120. Walker, D., *Development of novel molecular tools for the identification of essential genes of Clostridium difficile and a Clostridium tetracycline inducible promoter system*. 2011, University of Nottingham.
121. Hitchcock, D.S., et al., *Discovery of a bacterial 5-methylcytosine deaminase.* Biochemistry, 2014. **53**(47): p. 7426-7435.
122. Schaeffer, A.B. and M.D. Fulton, *A simplified method of staining endospores.* Science, 1933. **77**(1990): p. 194-194.

123. Roels, S., A. Driks, and R. Losick, *Characterization of spoIVA, a sporulation gene involved in coat morphogenesis in Bacillus subtilis.* J. Bacteriol., 1992. **174**: p. 575-585.
124. Abe, A., et al., *A Bacillus subtilis spore coat polypeptide gene, cotS.* Microbiology, 1995. **141**(6): p. 1433-1442.
125. Takamatsu, H., *A spore coat protein, CotS, of Bacillus subtilis is synthesized under the regulation of [sigma]K and GerE during development and is located in the inner coat layer of spores.* J. Bacteriol., 1998. **180**: p. 2968-2974.
126. Eichenberger, P., et al., *The σ^E regulon and the identification of additional sporulation genes in Bacillus subtilis.* Journal of Molecular Biology, 2003. **327**(5): p. 945-972.
127. Veiga, H. and M. G. Pinho, *Staphylococcus aureus requires at least one FtsK/SpoIIIE protein for correct chromosome segregation.* Molecular Microbiology, 2017. **103**(3): p. 504-517.
128. Asai, K., et al., *The Bacillus subtilis yabQ gene is essential for formation of the spore cortex.* Microbiology, 2001. **147**(4): p. 919-927.
129. Li, Y., et al., *Role of a SpoVA protein in dipicolinic acid uptake into developing spores of Bacillus subtilis.* Journal of Bacteriology, 2012. **194**(8): p. 1875-1884.
130. Wang, S.T., et al., *The forespore line of gene expression in Bacillus subtilis.* Journal of Molecular Biology, 2006. **358**(1): p. 16-37.
131. Buchanan, C.E. and M.L. Ling, *Isolation and sequence analysis of dacB, which encodes a sporulation-specific penicillin-binding protein in Bacillus subtilis.* Journal of Bacteriology, 1992. **174**(6): p. 1717-1725.
132. Fawcett, P., et al., *The transcriptional profile of early to middle sporulation in Bacillus subtilis.* Proceedings of the National Academy of Sciences, 2000. **97**(14): p. 8063-8068.
133. Moule, M.G., et al., *Genome-wide saturation mutagenesis of Burkholderia pseudomallei K96243 predicts essential genes and novel targets for antimicrobial development.* mBio, 2014. **5**(1): p. e00926-13.
134. Barquist, L., et al., *A comparison of dense transposon insertion libraries in the Salmonella serovars Typhi and Typhimurium.* Nucleic Acids Research, 2013. **41**(8): p. 4549-4564.
135. Min, K.-T., et al., *σF, the first compartment-specific transcription factor of B. subtilis, is regulated by an anti-σ factor that is also a protein kinase.* Cell, 1993. **74**(4): p. 735-742.
136. Duncan, L., et al., *Activation of cell-specific transcription by a serine phosphatase at the site of asymmetric division.* Science, 1995. **270**(5236): p. 641-644.
137. Stragier, P., C. Bonamy, and C. Karmazyn-Campelli, *Processing of a sporulation sigma factor in Bacillus subtilis: how morphological structure could control gene expression.* Cell, 1988. **52**(5): p. 697-704.
138. Sun, D.X., R.M. Cabrera-Martinez, and P. Setlow, *Control of transcription of the Bacillus subtilis spoIIIG gene, which codes for the forespore-specific transcription factor sigma G.* Journal of Bacteriology, 1991. **173**(9): p. 2977-2984.
139. Evans, L., A. Feucht, and J. Errington, *Genetic analysis of the Bacillus subtilis sigG promoter, which controls the sporulation-specific transcription factor σ^G.* Microbiology, 2004. **150**(7): p. 2277-2287.

140. Serrano, M., et al., *Expression of spoIIIJ in the prespore is sufficient for activation of σ^G and for sporulation in Bacillus subtilis.* Journal of Bacteriology, 2003. **185**(13): p. 3905-3917.
141. Kellner, E.M., A. Decatur, and J. Moran, Charles P., *Two-stage regulation of an anti-sigma factor determines developmental fate during bacterial endospore formation.* Molecular Microbiology, 1996. **21**(5): p. 913-924.
142. Kirchman, P.A., et al., *Forespore-specific disappearance of the sigma-factor antagonist SpollAB: implications for its role in determination of cell fate in Bacillus subtilis.* Molecular Microbiology, 1993. **8**(4): p. 663-671.
143. Camp, A.H. and R. Losick, *A novel pathway of intercellular signalling in Bacillus subtilis involves a protein with similarity to a component of type III secretion channels.* Molecular Microbiology, 2008. **69**(2): p. 402-417.
144. Chary, V.K., P. Xenopoulos, and P.J. Piggot, *Expression of the σ^F-directed csfB locus prevents premature appearance of σ^G activity during sporulation of Bacillus subtilis.* Journal of Bacteriology, 2007. **189**(23): p. 8754-8757.
145. Harry, K.H., et al., *Sporulation and enterotoxin (CPE) synthesis are controlled by the sporulation-specific sigma factors SigE and SigK in Clostridium perfringens.* Journal of bacteriology, 2009. **191**(8): p. 2728-2742.
146. Kirk, D.G., et al., *Involvement of Clostridium botulinum ATCC 3502 sigma factor K in early-stage sporulation.* Applied and Environmental Microbiology, 2012. **78**(13): p. 4590-4596.
147. Saujet, L., et al., *Genome-wide analysis of cell type-specific gene transcription during spore formation in Clostridium difficile.* PLOS Genetics, 2013. **9**(10): p. e1003756.
148. Wang, Z., M. Gerstein, and M. Snyder, *RNA-Seq: a revolutionary tool for transcriptomics.* Nature reviews. Genetics, 2009. **10**(1): p. 57-63.
149. Febrer, M., et al., *Advances in bacterial transcriptome and transposon insertion-site profiling using second-generation sequencing.* Trends in Biotechnology, 2011. **29**(11): p. 586-594.
150. Pinto, A.C., et al., *Application of RNA-seq to reveal the transcript profile in bacteria.* Genet. Mol. Res., 2011. **10**(3): p. 1707-1718.
151. Bachman, J., *Chapter Two - Reverse-Transcription PCR (RT-PCR),* in *Methods in Enzymology*, J. Lorsch, Editor. 2013, Academic Press. p. 67-74.
152. Heid, C.A., et al., *Real time quantitative PCR.* Genome Research, 1996. **6**(10): p. 986-994.
153. Bustin, S.A., et al., *The MIQE Guidelines: Minimum Information for Publication of Quantitative Real-Time PCR Experiments.* Clinical Chemistry, 2009. **55**(4): p. 611-622.
154. Kirk, D.G., et al., *Alternative sigma factors SigF, SigE, and SigG are essential for sporulation in Clostridium botulinum ATCC 3502.* Applied and Environmental Microbiology, 2014. **80**(16): p. 5141-5150.
155. Jones, S.W., et al., *Inactivation of σ^F in Clostridium acetobutylicum ATCC 824 blocks sporulation prior to asymmetric division and abolishes σ^E and σ^G protein expression but does not block solvent formation.* Journal of Bacteriology, 2011. **193**(10): p. 2429-2440.
156. Li, J. and B.A. McClane, *Evaluating the involvement of alternative sigma factors SigF and SigG in Clostridium perfringens sporulation and enterotoxin synthesis.* Infection and immunity, 2010. **78**(10): p. 4286-4293.
157. Fimlaid, K.A., et al., *Global analysis of the sporulation pathway of Clostridium difficile.* PLOS Genetics, 2013. **9**(8): p. e1003660.

158. Paradzik, T., et al., *Structure–function relationships of two paralogous single-stranded DNA-binding proteins from Streptomyces coelicolor: implication of SsbB in chromosome segregation during sporulation.* Nucleic Acids Research, 2013. **41**(6): p. 3659-3672.
159. Lang, D.R. and D.G. Lundgren, *Lipid composition of Bacillus cereus during growth and sporulation.* Journal of bacteriology, 1970. **101**(2): p. 483-489.
160. Nawrocki, K.L., et al., *Ethanolamine is a valuable nutrient source that impacts Clostridium difficile pathogenesis.* Environmental Microbiology, 2018. **20**(4): p. 1419-1435.
161. Deng, L., et al., *Manipulation of major membrane lipid synthesis and its effects on sporulation in Saccharomyces cerevisiae.* Bioscience, Biotechnology, and Biochemistry, 2008. **72**(9): p. 2362-2368.
162. Galperin, M.Y., et al., *Genomic determinants of sporulation in Bacilli and Clostridia: towards the minimal set of sporulation-specific genes.* Environmental Microbiology, 2012. **14**(11): p. 2870-2890.
163. Daniel, R.A., et al., *The Bacillus subtilis spoVD gene encodes a mother-cell-specific penicillin-binding protein required for spore morphogenesis.* Journal of Molecular Biology, 1994. **235**(1): p. 209-220.
164. Jonas, R.M., et al., *The Bacillus subtilis spoIIG operon encodes both sigma E and a gene necessary for sigma E activation.* Journal of Bacteriology, 1988. **170**(2): p. 507-511.
165. Gilmore, M.E., et al., *Production of muramic delta-lactam in Bacillus subtilis spore peptidoglycan.* Journal of bacteriology, 2004. **186**(1): p. 80-89.
166. Paredes-Sabja, D., et al., *Roles of DacB and Spm proteins in Clostridium perfringens spore resistance to moist heat, chemicals, and UV radiation.* Applied and Environmental Microbiology, 2008. **74**(12): p. 3730-3738.
167. Lai, E.-M., et al., *Proteomic analysis of the spore coats of Bacillus subtilis and Bacillus anthracis.* Journal of Bacteriology, 2003. **185**(4): p. 1443-1454.
168. Brunt, J., K.L. Cross, and M.W. Peck, *Apertures in the Clostridium sporogenes spore coat and exosporium align to facilitate emergence of the vegetative cell.* Food Microbiology, 2015. **51**: p. 45-50.
169. Nugroho, F.A., et al., *Characterization of a new Sigma-K-dependent peptidoglycan hydrolase gene that plays a role in Bacillus subtilis mother cell lysis.* Journal of Bacteriology, 1999. **181**(20): p. 6230-6237.
170. Haraldsen, J.D. and A.L. Sonenshein, *Efficient sporulation in Clostridium difficile requires disruption of the σ^K gene.* Molecular Microbiology, 2003. **48**(3): p. 811-821.
171. Duncan, L. and R. Losick, *SpoIIAB is an anti-sigma factor that binds to and inhibits transcription by regulatory protein sigma F from Bacillus subtilis.* Proceedings of the National Academy of Sciences, 1993. **90**(6): p. 2325-2329.
172. Campbell, E.A., et al., *Crystal structure of the Bacillus stearothermophilus anti-σ factor SpoIIAB with the sporulation σ factor σF.* Cell, 2002. **108**(6): p. 795-807.
173. Igoshin, O.A., C.W. Price, and M.A. Savageau, *Signalling network with a bistable hysteretic switch controls developmental activation of the σF transcription factor in Bacillus subtilis.* Molecular Microbiology, 2006. **61**(1): p. 165-184.
174. Clarkson, J., I.D. Campbell, and M.D. Yudkin, *Efficient regulation of σ^F, the first sporulation-specific sigma factor in B.subtilis.* Journal of Molecular Biology, 2004. **342**(4): p. 1187-1195.

175. Feucht, A., et al., *Bifunctional protein required for asymmetric cell division and cell-specific transcription in Bacillus subtilis.* Genes & Development, 1996. **10**(7): p. 794-803.
176. Arigoni, F., et al., *The SpoIIE phosphatase, the sporulation septum and the establishment of forespore-specific transcription in Bacillus subtilis: a reassessment.* Molecular Microbiology, 1999. **31**(5): p. 1407-1415.
177. Arigoni, F., et al., *Localization of protein implicated in establishment of cell type to sites of asymmetric division.* Science, 1995. **270**(5236): p. 637-640.
178. Adler, E., et al., *Structural relationship between a bacterial developmental protein and eukaryotic PP2C protein phosphatases.* Molecular Microbiology, 1997. **23**(1): p. 57-62.
179. Levin, P.A., et al., *Localization of the sporulation protein SpoIIE in Bacillus subtilis is dependent upon the cell division protein FtsZ.* Molecular Microbiology, 1997. **25**(5): p. 839-846.
180. Lucet, I., et al., *Direct interaction between the cell division protein FtsZ and the cell differentiation protein SpoIIE.* The EMBO Journal, 2000. **19**(7): p. 1467-1475.
181. Wu, L.J., A. Feucht, and J. Errington, *Prespore-specific gene expression in Bacillus subtilis is driven by sequestration of SpoIIE phosphatase to the prespore side of the asymmetric septum.* Genes & development, 1998. **12**(9): p. 1371-1380.
182. King, N., et al., *Septation, dephosphorylation, and the activation of sigma F during sporulation in Bacillus subtilis.* Genes & development, 1999. **13**(9): p. 1156-1167.
183. Feucht, A., L. Abbotts, and J. Errington, *The cell differentiation protein SpoIIE contains a regulatory site that controls its phosphatase activity in response to asymmetric septation.* Molecular Microbiology, 2002. **45**(4): p. 1119-1130.
184. Pan, Q., D.A. Garsin, and R. Losick, *Self-reinforcing activation of a cell-specific transcription factor by proteolysis of an anti-σ factor in B. subtilis.* Molecular Cell, 2001. **8**(4): p. 873-883.
185. LaBell, T.L., J.E. Trempy, and W.G. Haldenwang, *Sporulation-specific sigma factor sigma 29 of Bacillus subtilis is synthesized from a precursor protein, P31.* Proceedings of the National Academy of Sciences, 1987. **84**(7): p. 1784-1788.
186. Imamura, D., et al., *Evidence that the Bacillus subtilis SpoIIGA protein is a novel type of signal-transducing aspartic protease.* Journal of Biological Chemistry, 2008. **283**(22): p. 15287-15299.
187. Londoño-Vallejo, J.A. and P. Stragier, *Cell-cell signaling pathway activating a developmental transcription factor in Bacillus subtilis.* Genes & Development, 1995. **9**(4): p. 503-508.
188. Hofmeister, A.E.M., et al., *Extracellular signal protein triggering the proteolytic activation of a developmental transcription factor in B. subtilis.* Cell, 1995. **83**(2): p. 219-226.
189. Khvorova, A., et al., *The chromosomal location of the Bacillus subtilis sporulation gene spoIIR is important for its function.* Journal of Bacteriology, 2000. **182**(16): p. 4425-4429.
190. Chary, V.K., et al., *Loss of compartmentalization of σ^E activity need not prevent formation of spores by Bacillus subtilis.* Journal of Bacteriology, 2010. **192**(21): p. 5616-5624.

191. Fawcett, P., A. Melnikov, and P. Youngman, *The Bacillus SpoIIGA protein is targeted to sites of spore septum formation in a SpoIIE-independent manner.* Molecular Microbiology, 1998. **28**(5): p. 931-943.
192. Irie, R., Y. Fujita, and M. Kobayashi, *Nucleotide sequence and gene organization of the gerK spore germination locus of Bacillus subtilis 168.* The Journal of General and Applied Microbiology, 1996. **42**(2): p. 141-153.
193. Zhang, Z., et al., *A greedy algorithm for aligning DNA sequences.* Journal of Computational Biology, 2000. **7**(1-2): p. 203-214.
194. Rawlings, N.D. and A.J. Barrett, *Evolutionary families of metallopeptidases*, in *Methods in Enzymology*. 1995, Academic Press. p. 183-228.
195. Gonzales, T. and J. Robert-Baudouy, *Bacterial aminopeptidases: Properties and functions.* FEMS Microbiology Reviews, 1996. **18**(4): p. 319-344.
196. Cunningham, D.F. and B. O'Connor, *Proline specific peptidases.* Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology, 1997. **1343**(2): p. 160-186.
197. Cheggour, A., et al., *The dppA gene of Bacillus subtilis encodes a new D-aminopeptidase.* Molecular Microbiology, 2000. **38**(3): p. 504-513.
198. Mathiopoulos, C. and A.L. Sonenshein, *Identification of Bacillus subtilis genes expressed early during sporulation.* Molecular Microbiology, 1989. **3**(8): p. 1071-1081.
199. Mathiopoulos, C., et al., *A Bacillus subtilis dipeptide transport system expressed early during sporulation.* Molecular Microbiology, 1991. **5**(8): p. 1903-1913.
200. Singh, R.P., B. Setlow, and P. Setlow, *Levels of small molecules and enzymes in the mother cell compartment and the forespore of sporulating Bacillus megaterium.* Journal of Bacteriology, 1977. **130**(3): p. 1130-1138.
201. Cooksley, C.M., et al., *Regulation of neurotoxin production and sporulation by a putative agrBD signaling system in proteolytic Clostridium botulinum.* Applied and Environmental Microbiology, 2010. **76**(13): p. 4448-4460.
202. Eldar, A., *Partial penetrance facilitates developmental evolution in bacteria.* Nature, 2009. **460**: p. 510-514.
203. Kirk, J.A. and R.P. Fagan, *Heat shock increases conjugation efficiency in Clostridium difficile.* Anaerobe, 2016. **42**(Supplement C): p. 1-5.
204. van der Rest, M.E., C. Lange, and D. Molenaar, *A heat shock following electroporation induces highly efficient transformation of Corynebacterium glutamicum with xenogeneic plasmid DNA.* Applied Microbiology and Biotechnology, 1999. **52**(4): p. 541-545.

Appendix

Chapter 2: Supporting information

List of bacterial strains

Table S1: List of bacterial strains used in this study.

Strain	Relevant features	Source
	General use	
E. coli TOP10	Cloning and plasmid storage strain	Life Technologies
E. coli CA434	Conjugative strain (*E. coli* HB101 carrying R702)	Purdy *et al.*[79]
E. coli sExpress	Conjugative strain (NEB Express carrying R702)	Woods *et al.*[112]
C. sporogenes NCIMB 10696[c]	Wild type isolate	SBRC Laboratory collection
C. sporogenes Δ*spo0A*	*spo0A* mutant (in-frame deletion)	Dr. Inés Cañadas Blasco, SBRC[78]
	Chapter 3	
C. sporogenes-CLSPO_c06750-1240a::CT	ClosTron mutant (type IV restriction system)	This study
C. sporogenes ΔCLSPO_c06750	CRISPR mutant (type IV restriction system)	This study
*C. sporogenes-hsdR*1537a::CT	ClosTron mutant (type I restriction system)	This study
	Chapter 5	
C. sporogenes Δ*sigF*	CRISPR mutant (sigma factor σ^F)	This study
C. sporogenes Δ*sigE*	CRISPR mutant (sigma factor σ^E)	This study
C. sporogenes Δ*sigG*	CRISPR mutant (sigma factor σ^G)	This study
C. sporogenes Δ*sigK*	CRISPR mutant (sigma factor σ^K)	This study

Table S1: List of bacterial strains used in this study (continuation).

Strain	Relevant features	Source
	Chapter 6	
C. sporogenes Δ*sigF* + *sigF*	Reverted strain (sigma factor σ^F)	This study
C. sporogenes Δ*sigE* + *sigE*	Reverted strain (sigma factor σ^E)	This study
C. sporogenes Δ*sigG* + *sigG*	Reverted strain (sigma factor σ^G)	This study
C. sporogenes Δ*sigK* + *sigK*	Reverted strain (sigma factor σ^K)	This study
C. sporogenes Δ*spoIIAA*	CRISPR mutant (anti-anti sigma factor SpoIIAA)	This study
C. sporogenes Δ*spoIIAA* + *spoIIAA*	Reverted strain (anti-anti sigma factor SpoIIAA)	This study
C. sporogenes ΔAPP	CRISPR mutant (Aminopeptidase P)	This study
C. sporogenes ΔAPP + APP	Reverted strain (Aminopeptidase P)	This study
C. sporogenes Δc33910	CRISPR mutant (Gene of unknown function)	This study
C. sporogenes Δc33910 + c33910	Reverted strain (Gene of unknown function)	This study
*C. sporogenes-spoIIE*2368s::CT	ClosTron mutant (Stage II sporulation gene)	This study
*C. sporogenes-spoIIR*335s::CT	ClosTron mutant (Stage II sporulation gene)	This study
*C. sporogenes-gerKA4*775s::CT	ClosTron mutant (Germinant receptor)	This study

List of vectors

Table S2: List of vectors used in this study.

Vector	Feature/use	Source
	Modular vectors	
pMTL83151	Modular plasmid; pCB102; ColE1+*tra*; Cm^R/Tm^R	Heap *et al.*[113]
pMTL80221	Modular plasmid; no Gram + replicon; p15+*tra*; Erm^R	Dr. Robert Mansfield, SBRC
pRECas1-empty	RiboCas vector, Cm^R, P_{fdx}*-E-*cas9*	Cañadas *et al.*[87]
	Chapter 3	
pMTL43151-CAETHG_*pyrE*	Template to generate the CRISPR/Cas9 vectors below	Dr. Peter Rowe, SBRC
pMTL007::*Csp*-CLSPO_c06750-1240a	ClosTron vector targeting CLSPO_c06750	DNA2.0
pMTL43151-CLSPO-c06750CT	CRISPR/Cas9 vector deleting CLSPO_c06750	This study
pMTL007::*Csp-hsdR*-1537a	ClosTron vector targeting *hsdR*	DNA2.0
pMTL43121-CLSPO-*hsdR*	CRISPR/Cas9 vector deleting *hsdR*	This study
pMTL80221-P_{neg}-RFP	Erm^R; p15a; no promoter driving RFP	This study
pMTL80221-P_{hsd}-RFP	Erm^R; p15a; P_{hsd}-RFP	This study
pMTL80221-P_{BAD}-RFP	Erm^R; p15a; P_{BAD}-RFP	This study
pMTL80221-P_{BAD}-*hsdMS*	Erm^R; p15a; P_{BAD}-*hsdMS*	This study
pMTL83151_RS1/2	pMTL83151 containing type I recognition sites 1 and 2	This study
pMTL83151_RS3	pMTL83151 containing type I recognition site 3	This study

Table S2: List of vectors used in this study (continuation).

Vector	Feature/use	Source
	Chapter 4	
pMTL-tet3nO	Erm^R; $P_{fdx/tetO}$- *catP*; P_{thl}- *tetR*; 2 RBS upstream of *catP*	SBRC Laboratory collection
pMTL-tet3nO_RBS	Erm^R; $P_{fdx/tetO}$- *catP*; P_{thl}- *tetR*; 1 RBS upstream of *catP*	This study
pMTL-tet3nO_RFtetR	Erm^R; $P_{xylA/tetO}$- *catP*; $P_{tetR/tetO}$- *tetR*; 1 RBS upstream of *catP*	SBRC Laboratory collection
pMTL82254_P_{neg}	Erm^R; no promoter driving *catP*	SBRC Laboratory collection
pMTL82254_P_{fdx}	Erm^R; P_{fdx}-*catP*	SBRC Laboratory collection
pMTL82254_Lac	Erm^R; lactose inducible system driving *catP*	SBRC Laboratory collection
pMTL-CW21	Erm^R; pCB102; ColE1+*tra*; *catP* Tn; $P_{xylA/tetO}$- *Himar1*; $P_{tetR/tetO}$-*tetR*; I-*Sce*I sites; *codA*	Dr. Craig Woods, SBRC
	Chapter 5	
pRECas1-*sigF*_BM	RiboCas vector deleting *sigF*; introduces bookmark (BM)	This study
pRECas1-*sigE*_BM	RiboCas vector deleting *sigE*; introduces BM	This study
pRECas1-*sigG*_BM	RiboCas vector deleting *sigG*; introduces BM	This study
pRECas1-*sigK*_BM	RiboCas vector deleting *sigK*; introduces BM	This study
	Chapter 6	
pRECas1-*spoIIAA*_BM	RiboCas vector deleting *spoIIAA*; introduces BM	This study
pRECas1-APP_BM	RiboCas vector deleting APP; introduces BM	This study

Table S2: List of vectors used in this study (continuation).

Vector	Feature/use	Source
	Chapter 6	
pRECas1-c33910_BM	RiboCas vector deleting CLSPO_c33910; introduces BM	This study
pRECas1-*sigF*_rev	RiboCas vector targeting BM and introducing *sigF* at original locus; introduces watermark	This study
pRECas1-*sigE*_rev	RiboCas vector targeting BM and introducing *sigE* at original locus; introduces watermark	This study
pRECas1-*sigG*_rev	RiboCas vector targeting BM and introducing *sigG* at original locus; introduces watermark	This study
pRECas1-*sigK*_rev	RiboCas vector targeting BM and introducing *sigK* at original locus; introduces watermark	This study
pRECas1-*spoIIAA*_rev	RiboCas vector targeting BM and introducing *spoIIAA* at original locus; introduces watermark	This study
pRECas1-APP_rev	RiboCas vector targeting BM and introducing APP at original locus; introduces watermark	This study
pRECas1-c33910_rev	RiboCas vector targeting BM and introducing CLSPO_c33910 at original locus; introduces watermark	This study
pMTL007::*Csp-spoIIE*-2368s	ClosTron vector targeting *spoIIE*	DNA2.0
pMTL007::*Csp-spoIIR*-335s	ClosTron vector targeting *spoIIR*	DNA2.0
pMTL007::*Csp- gerKA4*-775s	ClosTron vector targeting *gerKA4*	DNA2.0

List of primers

Table S3: List of primers used in this study.

Primer name	Sequence (5' → 3')	Use/features
	General primers	
M13_F	TGTAAAACGACGGCCAGT	Modular primer
M13_R	CAGGAAACAGCTATGAC	Modular primer
pCD6_R1	GACTTTAAGCCTACGAATACC	Modular primer
pCB102_R1	CTGTTATGCCTTTTGACTATC	Modular primer
RAM_F	ACGCGTTATATTGATAAAAATAATAATAGTGGG	Screening (ClosTron)
RAM_R	ACGCGTGCGACTCATAGAATTATTTCCTCCCG	Screening (ClosTron)
EBS Universal	CGAAATTAGAAACTTGCGTTCAGTAAAC	Screening (ClosTron)
erm_inv_F	GAAGCACCAGCATCTCTTG	Screening (Inverse PCR)
erm_inv_R	CACTCAAGTCTCGATTCAGCA	Screening (Inverse PCR)
erm_RAM_seq	GGAAATGGCAATGATAGCGA	Sequencing (Inverse PCR)
EBS2	TGAACGCAAGTTTCTAATTTCGGTTAATATCCGATAGAGGA	Southern blot probe
SalI-R1	ATTACTGTGACTGGTTTGCACCACCCTCTTCG	Southern blot probe
sgRNA_F	ATATATTCTAGATTTATATTTAGTCCCTTGCC	Screening (CRISPR)
sgRNA_AatII_R	ATATATGACGTCATAAAAATAAGAAGCCTGCAAATGCAGGCTT CTTATTTTTATAAAAAAAGCACCGACTCGGTGCCACTTTTTCAA GTTG	Obtain sgRNA fragment; *Aat*II site

Table S3: List of primers used in this study (continuation).

Chapter 3		
c06750_CTscr_F	CCTAAATTTCTTTTAGGTATGACAGC	Screening/sequencing (ClosTron)
c06750_CTscr_R	GCTTTACCCTTTCCTCTGC	Screening/sequencing (ClosTron)
sgRNA_AsiSI_R	CGCGCGCGGCGATCGCATAAAAATAAGAAGCCTGCAAATGCA GGCTTCTTATTTTTATAAAAAAAGCACCGACTCGGTGCCACTTT TTCAAGTTG	Obtain sgRNA fragment; *Asi*SI site
CT_sgRNA_F	TCGTCGACTGGATATTCACCGAACACTAGTTTTAGAGCTAGAA ATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAA GTGGCACCGAGTCGGTGCTTTTTTTATAAAAATAAGAAGC	Obtain sgRNA fragment; *Sal*I site; targets ClosTron insertion
c06750_LHA_F	TATATAGCGATCGCGGAACAGATAATTTAAAGAAAGCCTTAG	Obtain LHA; *Asi*SI site
c06750_SOE_R	TCTATGATATATCCATTTTTACCTCTTACTTGAACTAAATTC	Obtain LHA; SOEing PCR
c06750_SOE_F	AAAAATGGATATATCATAGAGAAGATAACAAAATAAAAATAG GG	Obtain RHA; SOEing PCR
c06750_RHA_R	GGCGCGCCCTTCTGCCTTCCCATTTTTACAAC	Obtain RHA; *Asc*I site
c06750_scr_F	GGGATATGCTGAAAGTAAAATGG	Screening
c06750_scr_R	GAGTTACAGTAAAAAAATAGGCAGG	Screening
c06750_fl_F	CATAGCTGCTGGAATTTTATATGG	Sequencing
c06750_fl_R	GAGTTACAGTAAAAAAATAGGCAGG	Sequencing
hsdR_CTscr_F	GGATAATGGAAAGATAAAGCATAGACC	Screening/sequencing
hsdR_CTscr_R	GCTTTTCCTCCTACTTTCCCC	Screening/sequencing

Table S3: List of primers used in this study (continuation).

Chapter 3

hsdR_sgRNA_F	CGTCGACACTTCCTACATCCCAAAGGGGTTTTAGAGCTAGAAA TAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAG TGGCACCGAGTCGGTGCTTTTTTTATAAAAATAA	Obtain sgRNA fragment; *Sal*I site
hsdR_LHA_F	TATATAGCGATCGCATGAAAAATATTTATATTATTA	Obtain LHA; *Asi*SI site
hsdR_SOE_R	TTTATATTGTTATCATTACAATTCGCTCTCCTTTATATCT	Obtain LHA; SOEing PCR
hsdR_SOE_F	TGTAATGATAACAATATAAAATTTGAATAAGGTAATGTTTTG	Obtain RHA; SOEing PCR
hsdR_RHA_R1	TATATAGGCGCGCCTAACTATGTCTTTAGTTATA	Obtain RHA; *Asc*I site; 1Kb HA
hsdR_RHA_R2	TATATAGGCGCGCCGCTTTATTCCATTTTCTATTTTTTCC	Obtain RHA; *Asc*I site; 500bp HA
hsdMSR_LHA_F	TATAGCGATCGCATTAGGACTAATAATATGATATTTAATATG	Obtain LHA; *Asi*SI site
hsdMSR_SOE_R	CAAATTTTATATTGTATCCATGTATTTTTCCTCCAATATA	Obtain LHA; SOEing PCR
hsdMSR_SOE_F	ATACATGGATACAATATAAAATTTGAATAAGGTAATGTTTTG	Obtain RHA; SOEing PCR
hsdR_scr_F	GGATGAAACTCCACGAACTG	Screening
hsdR_scr_R	ATTTGAGGCACTAGCCTTCC	Screening
hsdR_fl_F	GAAGATGAAATAATTCATAATGGG	Sequencing
hsdR_fl_R	GCTAGTAATCCTGCAAAACATTAC	Sequencing
RFP_NotI_F	TATAGCGGCCGCGGTCTCAAATACATGGCTTCCTCCGAAGACG TTATCAAAGAGTTCAT	Clone RFP constructs
RFP_NheI_R	TATAGCTAGCTGTAAAACGACGGCCAGTTTAAGCACCGGTGGA GTGACGACCTTCAGCACGTT	Clone RFP constructs

Table S3: List of primers used in this study (continuation).

	Chapter 3	
Phsd_NotI_F	TATAGCGGCCGCACTTTGATGGGTTTAGTATAACTGT	Clone RFP constructs
Phsd_BsaI_R	TATAGGTCTCGTATTTTTCCTCCAATATATAAAATATATATTATAAA	Clone RFP constructs
araC_NotI_F	GACCGCGGCCGCGTGTAGTA	Clone RFP constructs
PBAD_BsaI_R	TATAGGTCTCATGTATATCTCCTTCTTAAACCCAAAAAAACG	Clone RFP constructs
hsdM_BsaI_F	TATAGGTCTCATACATATGGATAATCAAATAGTAAACTTTATATGG	Clone pMTL80221-P_{BAD}-*hsdSM*
hsdS_NheI_R	TATAGCTAGCTTACAATTCGCTCTCCTTTATATCTGC	Clone pMTL80221-P_{BAD}-*hsdSM*
T1_RS1/2_NdeI_F	CATATGCTAAATACGATGAAAATGATAAACTTATTCG	Generate pMTL83151_RS1/2
T1_RS1/2_NotI_R	GCGGCCGCTTCATAATGACTAGCTAAATATAAAAAATTCAC	Generate pMTL83151_RS1/2
T1_RS3_F	TATGGCATTATATAACTGC	Generate pMTL83151_RS3
T1_RS3_R	GGCCGCAGTTATATAATGCCA	Generate pMTL83151_RS3
	Chapter 4	
tet3nO_SalI_F	GTCGACTGATCGGATCCAATTTATAC	Delete RBS from pMTLtet3nO
tet3nO_1RBS_R	CATATGTGTAACACACCTCCTTAAAAATTACAC	Delete RBS from pMTLtet3nO

Table S3: List of primers used in this study (continuation).

Chapter 5

sigF_sgRNA_F	CGTCGACTAAAATAAGTGAAAATACAGGTTTTAGAGCTAGAA ATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAA GTGGCACCGAGTCGGTGCTTTTTTTATAAAAATAA	Obtain sgRNA fragment; *Sal*I site
sigF_LHA_F	ATATATGACGTCAATCAAATAACATTATAACAATGTGTTCTAG	Obtain LHA; AatII site
sigF_SOE_R	TCCTCTGTGTAAATCTCTGTTAACACTCATATTATTCACCTAAC	Obtain LHA; bookmark
sigF_SOE_F	GTTAACAGAGATTTACACAGAGGAACTGTATAAAAACCATATG CA	Obtain RHA; bookmark
sigF_RHA_R	ATATATGGCGCGCCAAAGGTATATCTAAATCCCTTGCTG	Obtain RHA; *Asc*I site
sigF_scr_F	CTCCTTTCATAACTAAAATTTAAAA	Screening
sigF_scr_R	GTAGAACAAGCTCCATATAATCC	Screening
sigF_fl_F	GGTTTTCAAATAATTAAGTTAGGTG	Sequencing
sigE_sgRNA_F	CGTCGACAAGGAAGTAGCAGATATGCTGTTTTAGAGCTAGAAA TAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAG TGGCACCGAGTCGGTGCTTTTTTTATAAAAATAA	Obtain sgRNA fragment; *Sal*I site
sigE_LHA_F	ATATATGACGTCACTGTTTAAATATACTTTTTAATATATTATTT TTG	Obtain LHA; *Aat*II site
sigE_SOE_R	CATTCCTCTGTGTAAATCTCTGTTAACTATCATAAAAGTCCCTC CTATAT	Obtain LHA; bookmark

Table S3: List of primers used in this study (continuation).

Chapter 5

sigE_SOE_F	ATAGTTAACAGAGATTTACACAGAGGAATGGTTTAGTTTGTCCTTAGTAT	Obtain RHA; bookmark
sigE_RHA_R	TATATAGGCGCGCCTATACCCCCTTAAACAGTTTTTATATTATAAT	Obtain RHA; *Asc*I site
sigE_scr_F	CAGCAAAACATCCTTTCCTA	Screening
sigE_scr_R	GGAATATAACTCTACTTGTTCCATG	Screening
sigE_fl_F	CCGAGGAATTATATAGGAGG	Sequencing
sigG_sgRNA_F	CGTCGACTTACTGGTAAGGATTAGGAAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTTATAAAAATAA	Obtain sgRNA fragment; *Sal*I site
sigG_LHA_F	ATATATGACGTCTTAGACACTTATGATAAATTTTATATACCTTATAGGG	Obtain LHA; *Aat*II site
sigG_SOE_R	TCCTCTGTGTAAATCTCTGTTAACAACCATAAAGTCATCAGTCC	Obtain LHA; bookmark
sigG_SOE_F	GTTAACAGAGATTTACACAGAGGATATGTTTAATCAAAAATGTT	Obtain RHA; bookmark
sigG_RHA_R	TATATAGGCGCGCCATTATATCTAATTCTTCTCCATTAGATTTTTAATCTC	Obtain RHA; *Asc*I site
sigG_scr_F	GGTAGAAAAGGATGTTTTTAGTAAT	Screening
sigG_scr_R	CCTAATACTATTACCATAATAGCATT	Screening

Table S3: List of primers used in this study (continuation).

Chapter 5

sigG_fl_F	CATTCTGAAGGGACTGATGA	Sequencing
sigK_sgRNA_F	CGTCGACAGAACTCAAAGAGAAATAGCGTTTTAGAGCTAGAA ATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAA AGTGGCACCGAGTCGGTGCTTTTTTTATAAAAATAA	Obtain sgRNA fragment; *Sal*I site
sigK_LHA_F	ATATATGACGTCAGAAGCAGAGTCAGGGAAAATAC	Obtain LHA; *Aat*II site
sigK_SOE_R	TCCTCTGTGTAAATCTCTGTTAACGAACACAAGTTGCTCCTCC T	Obtain LHA; bookmark
sigK_SOE_F	GTTAACAGAGATTTACACAGAGGAAATATTTAAACAAAAAAT CT	Obtain RHA; bookmark
sigK_RHA_R	TATATAGGCGCGCCTACCCCTACGGATAAATGTAAATCTG	Obtain RHA; *Asc*I site
sigK_scr_F	GGATTTTAACCAAGTAGGTG	Screening
sigK_scr_R	GACCATTTATTCTTATAGTTGG	Screening
sigK_fl_F	GCACTATTAGGAGGAGCAAC	Sequencing
rrfA_qPCR_F	AGCGGTGAAATGCGTAGAGA	RT-qPCR
rrfA_qPCR_R	GGCACAGGGGGAGTTGATAC	RT-qPCR
gyrA_qPCR_F	GAAAGGGTATTCAAGCCACATC	RT-qPCR
gyrA_qPCR_R	CTATCGGTGTGGCTATGGGGGCT	RT-qPCR
spoIID_qPCR_F	GCAGATGTATGTGATAGTGTTCAG	RT-PCR/RT-qPCR
spoIID_qPCR_R	CTAGTAGCAAAATAATATGGAGCC	RT-PCR/RT-qPCR

Table S3: List of primers used in this study (continuation).

Chapter 5		
spoIIM_qPCR_F	ACATACAAACCCCTACTCTTGCT	RT-PCR/RT-qPCR
spoIIM_qPCR_R	TGCATCAGTGTTAGCAATGG	RT-PCR/RT-qPCR
spoIIR_qPCR_F	TGATGCCAAAATTAGATAAAAGT	RT-PCR/RT-qPCR
spoIIR_qPCR_R	CCAATTCTGACCCTGACCAT	RT-PCR/RT-qPCR
gpr_qPCR_F	CCCCAAAACCCCTGGTGATATAG	RT-PCR/RT-qPCR
gpr_qPCR_R	ACTGCAGATGCTCTAGGACC	RT-PCR/RT-qPCR
spoIVB_qPCR_F	TCCAGCATCCTTTAGCATCCA	RT-PCR/RT-qPCR
spoIVB_qPCR_R	CAAGGCATGAGTGGAAGTCC	RT-PCR/RT-qPCR
sspH1_qPCR_F	GAGGAGCCATTATAGTTGTACCG	RT-PCR/RT-qPCR
sspH1_qPCR_R	TGGACGCAAGTAGAGCTTCTC	RT-PCR/RT-qPCR
cotA_qPCR_F	TTCCACTCCTCCGTTGAACC	RT-PCR/RT-qPCR
cotA_qPCR_R	TGAGATGTGAACCTGTGCCC	RT-PCR/RT-qPCR
cotI_qPCR_F	TCTAGCGTATATAGCATAGCGG	RT-PCR/RT-qPCR
cotI_qPCR_R	CCTCATCATCCCAGTCTTTTCG	RT-PCR/RT-qPCR
Chapter 6		
BM_sgRNA_F	CGTCGACGTTAACAGAGATTTACACAGGTTTTAGAGCTAGAA ATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAA AGTGGCACCGAGTCGGTGCTTTTTTTATAAAAATAA	Obtain sgRNA fragment; *Sal*I site; targets the bookmark
sigF_SNP_F	CTTACTGTATAGAAACCATATGC	Introduces SNP in stop codon

Table S3: List of primers used in this study (continuation).

Chapter 6

sigF_SNP_R	GCATATGGTTTCTATACAGTAAG	Introduces SNP in stop codon
sigE_SNP_F	GGTTTAATTTGTCCTTAGTATAAAATTAAGTTCC	Introduces SNP in stop codon
sigE_SNP_R	GGAACTTAATTTTATACTAAGGACAAATTAAACC	Introduces SNP in stop codon
sigG_SNP_F	GAGAAAGTATGTTTAGTCAAAAATGTTC	Introduces SNP in stop codon
sigG_SNP_R	GAACATTTTTGACTAAACATACTTTCTC	Introduces SNP in stop codon
sigK_SNP_F	GAATTAAATTATAATAAAAATATTTAGACAAAAAATCTAAAG	Introduces SNP in stop codon
sigK_SNP_R	CTTTAGATTTTTTGTCTAAATATTTTTATTATAATTTAATTC	Introduces SNP in stop codon
spoIIAA_sgRNA_F	GTTTTCGTCGACAAGAAATTATCATTAAGGGAGTTTTAGAGCT AGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTG AAAAAGTGGCACCGAGTCGGTGCTTTTTTTATAAA	Obtain sgRNA fragment; *Sal*I site
spoIIAA_LHA_F	ATATATGACGTCAGTATCTGAAAGATTAGATACAGTTGG	Obtain LHA; *Aat*II site
spoIIAA_SOE_R	TCCTCTGTGTAAATCTCTGTTAACATACATACTATTCCCCCCA ATTTC	Obtain LHA; bookmark
spoIIAA_SOE_F	GTTAACAGAGATTTACACAGAGGAAACTTATAATGGGGGGAT TTATAAGATG	Obtain RHA; bookmark
spoIIAA_RHA_R	TATATAGGCGCGCCTTTATCAATAAGCAAAACTGGTGC	Obtain RHA; *Asc*I site
spoIIAA_SNP_F	CGTAGAAAACTTATAGTGGGGGGATTTATAAG	Introduces SNP in stop codon
spoIIAA_SNP_R	CTTATAAATCCCCCCACTATAAGTTTTCTACG	Introduces SNP in stop codon
spoIIAA_scr_F	AAAGAGGCTTTGTGTAAAGC	Screening

Table S3: List of primers used in this study (continuation).

Chapter 6

spoIIAA_scr_R	TTTAGAGCCTCTTTTAATGC	Screening
spoIIAA_fl_F	AAAGAGATATAAAAGAAAAAGAAGG	Sequencing
APP_sgRNA_F	GTTTTCGTCGACAATGAAATTCTTGTATGCAAGTTTTAGAGCT AGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTG AAAAAGTGGCACCGAGTCGGTGCTTTTTTTATAAA	Obtain sgRNA fragment; *Sal*I site
APP_LHA_F	ATATATGACGTCGTAATTTGCAAAAAAGTTATCTCCAACC	Obtain LHA; *Aat*II site
APP_SOE_R	TCCTCTGTGTAAATCTCTGTTAACGTTCATCAAATATCCAACA CCGTGG	Obtain LHA; bookmark
APP_SOE_F	GTTAACAGAGATTTACACAGAGGAGGCCTATAGATCTATACG CAATAATTTC	Obtain RHA; bookmark
APP_RHA_R	TATATAGGCGCGCCGCATTGTAAATCCACCGATGG	Obtain RHA; *Asc*I site
APP_SNP_F	CAATTAGTTATGGCCTATAAATCTATACGC	Introduces SNP in stop codon
APP_SNP_R	GCGTATAGATTTATAGGCCATAACTAATTG	Introduces SNP in stop codon
APP_scr_F	GTATGGGATGACCTGCTGTG	Screening
APP_scr_R	CAAGTTTCAACCATACATTGGC	Screening
APP_fl_F	GAACAAGGTAGATTTATTCGTCC	Sequencing
c33910_sgRNA_F	GTTTTCGTCGACAGATGTTAGATTTTTCAAAGGTTTTAGAGCT AGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTG AAAAAGTGGCACCGAGTCGGTGCTTTTTTTATAAA	Obtain sgRNA fragment; *Sal*I site

Table S3: List of primers used in this study (continuation).

Chapter 6

c33910_LHA_F	ATATATGACGTCGAAAAAATGTAAAATTGGGTGC	Obtain LHA; *Aat*II site
c33910_SOE_R	TCCTCTGTGTAAATCTCTGTTAACTTTCATTTTTATCCTCTCCA CATGC	Obtain LHA; bookmark
c33910_SOE_F	GTTAACAGAGATTTACACAGAGGAGTATATTAATAGTGATTTT TCTATTAAGAG	Obtain RHA; bookmark
c33910_RHA_R	TATATAGGCGCGCCTGAAAATGTCTTTAATGTTTGG	Obtain RHA; *Asc*I site
c33910_SNP_F	GAAAGTATATTAGTAGTGATTTTTCTATTAAGAG	Introduces SNP in stop codon
c33910_SNP_R	CTCTTAATAGAAAAATCACTACTAATATACTTTC	Introduces SNP in stop codon
c33910_scr_F	GAGCCTGTAGAACTAGGTGG	Screening
c33910_scr_R	ATGTAAGCTATCTGAAGTGG	Screening
c33910_fl_F	TTGGGGGTAAAAGTAATAGG	Sequencing
spoIIE_scr_F	GGTAAGTGATGGGGTTCTAGATTATGAAAGTAG	Screening/sequencing (ClosTron)
spoIIE_scr_R	CTACTATTATTTTGTCATTGGCTTCAATCATATC	Screening/sequencing (ClosTron)
spoIIR_scr_F	ATGAAAAGAATATTAGTTTTTATGTTAAG	Screening/sequencing (ClosTron)
spoIIR_scr_R	TTATTTTTTTAATTTATTAATCTCTTCTACAAC	Screening/sequencing (ClosTron)
gerKA4_scr_F	ATGAAAGAAATAATTAAGGATGAAATATAT	Screening/sequencing (ClosTron)
gerKA4_scr_R	TTATCTTTGTCTTATTTTGTCTCTAG	Screening/sequencing (ClosTron)

Vectors used in the RFP reporter assay

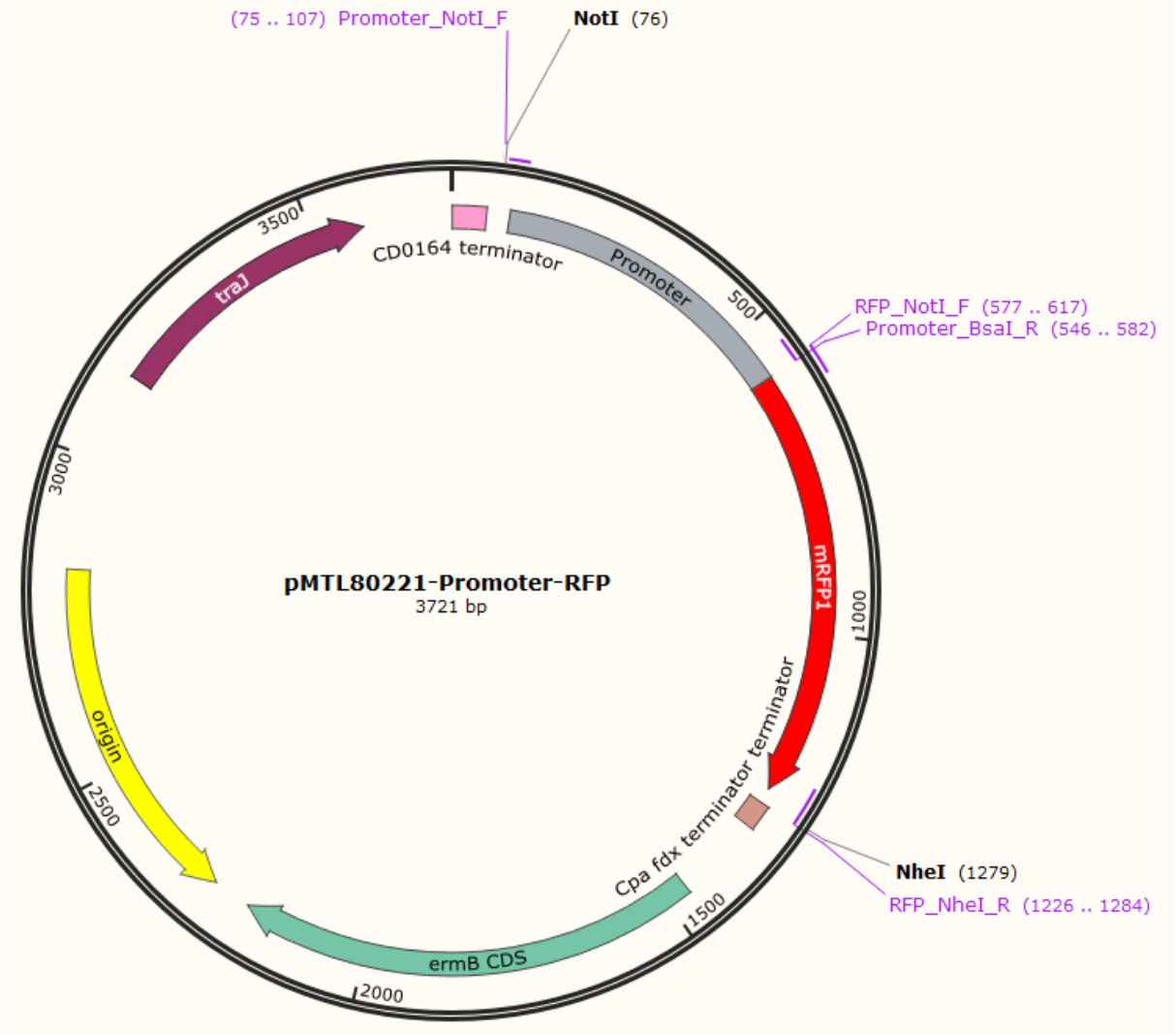


Figure S1: Generic representation of the pMTL80221-Promoter-RFP vectors used in the RFP reporter assay. The promoter drives the expression of the gene encoding the RFP. Promoter and gene are flanked by *Not*I and *Nhe*I restriction sites. The primers used to clone the construct are represented.

Vector used to express the *hsdS* and *hsdM* genes from the type I restriction system

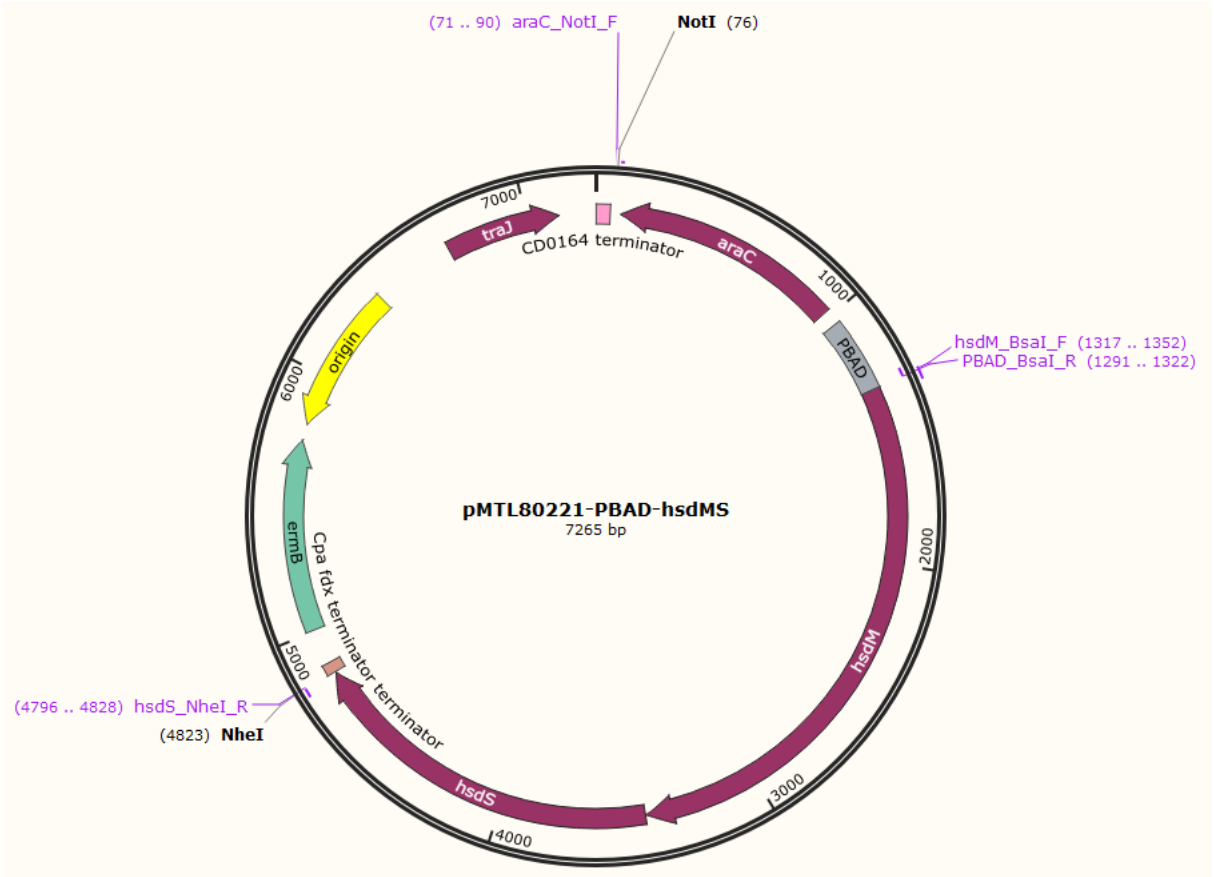


Figure S2: Representation of the pMTL80221-P_{BAD}-*hsdMS* vector used to express the *hsdS* and *hsdM* genes from the type I restriction system. The *araC* gene encodes the transcriptional regulator AraC. The P_{BAD} promoter drives the expression of the *hsdM-hsdS* operon. The promoter and the operon are flanked by the restriction sites *Not*I and *Nhe*I.

Vectors containing sequences potentially recognised by the type I restriction system

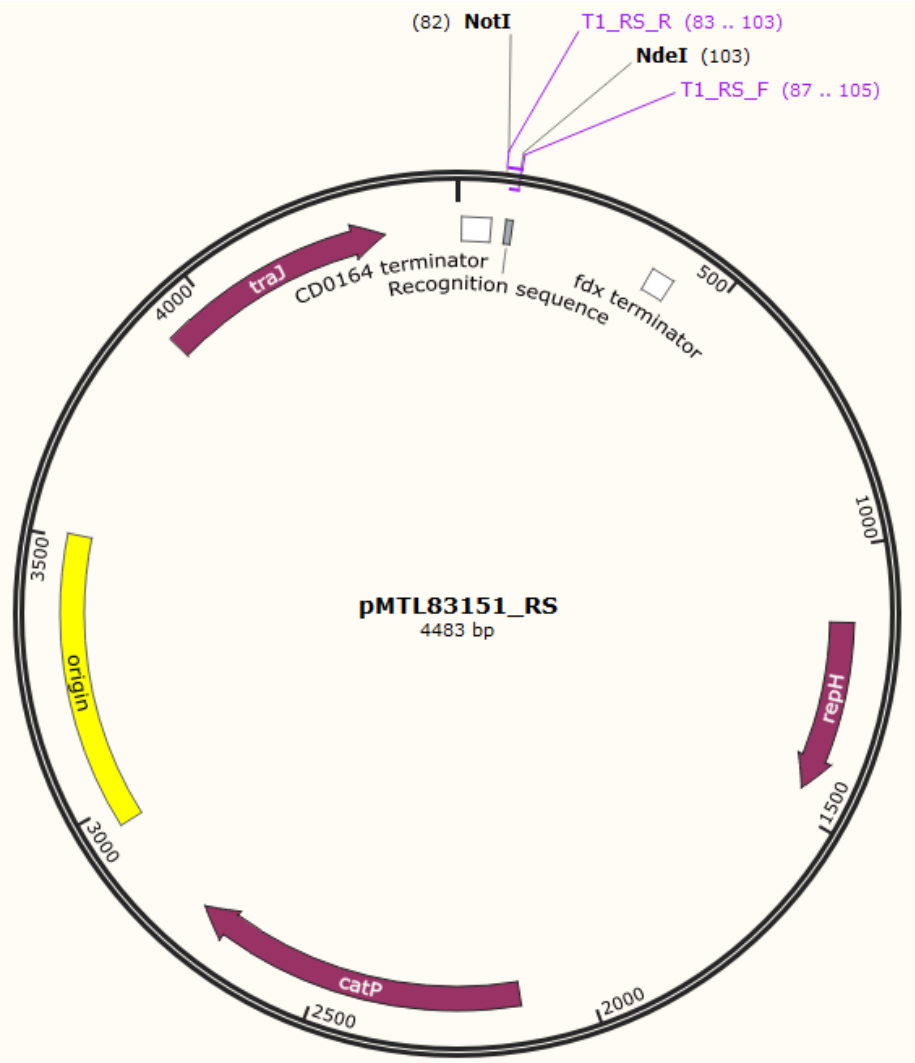


Figure S3: Generic representation of the pMTL83151_RS vectors. The sequence potentially recognised by the type I restriction system is flanked by the restriction enzymes *Not*I and *Nde*I.

CRISPR/Cas9 vectors

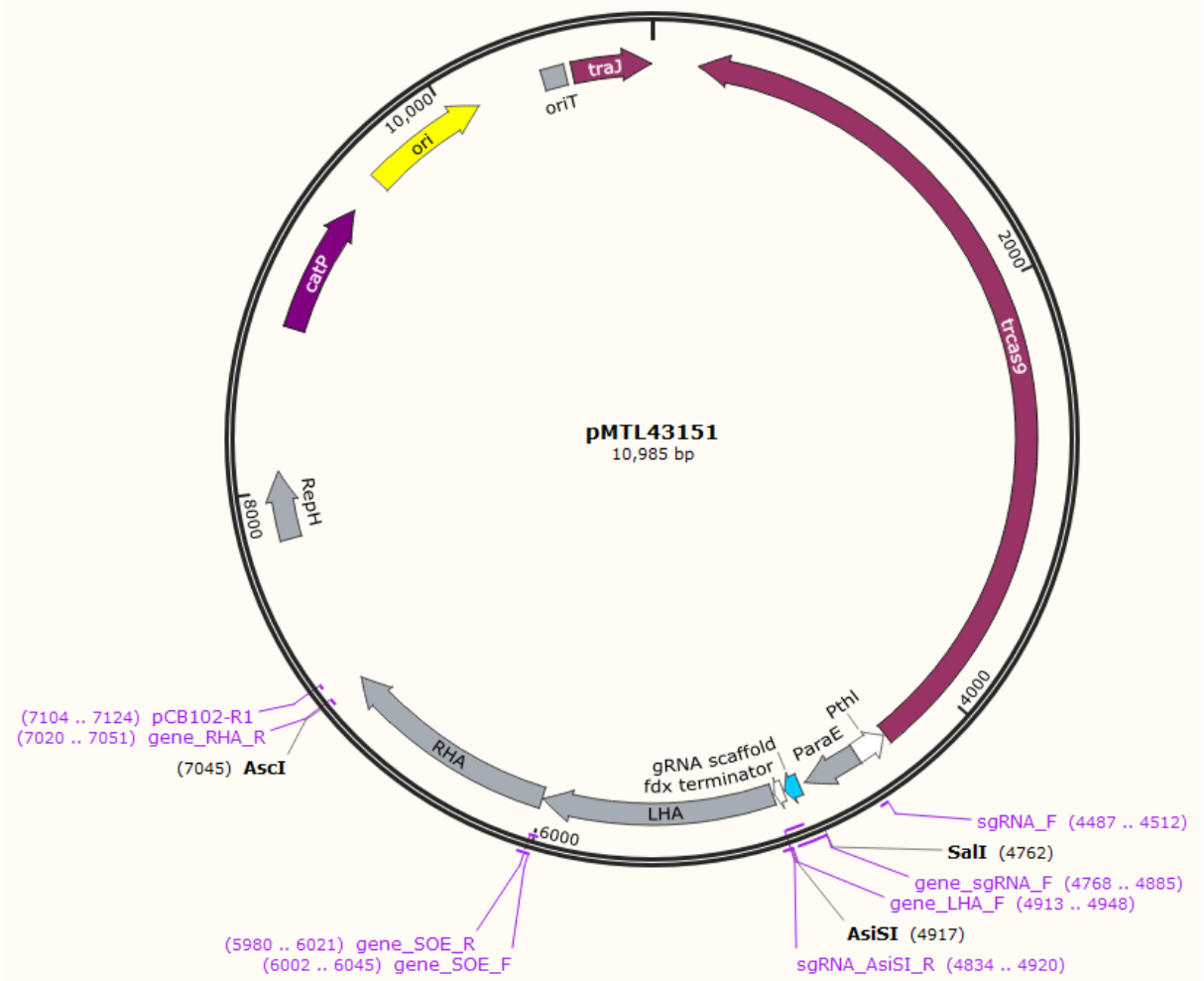


Figure S4: Generic representation of the pMTL43151 vectors with the CRISPR/Cas9 system. The left homology arm (LHA) and right homology arm (RHA) are flanked by *Asi*SI and *Asc*I restriction sites, while the specific sgRNA region is flanked by *Sal*I and *Asi*SI. The P_{thl} promoter drives the expression of the truncated Cas9 (*trcas9*), and the P_{araE} promoter drives the expression of the sgRNA region. The primers used to clone and screen the vector are represented.

RiboCas vectors

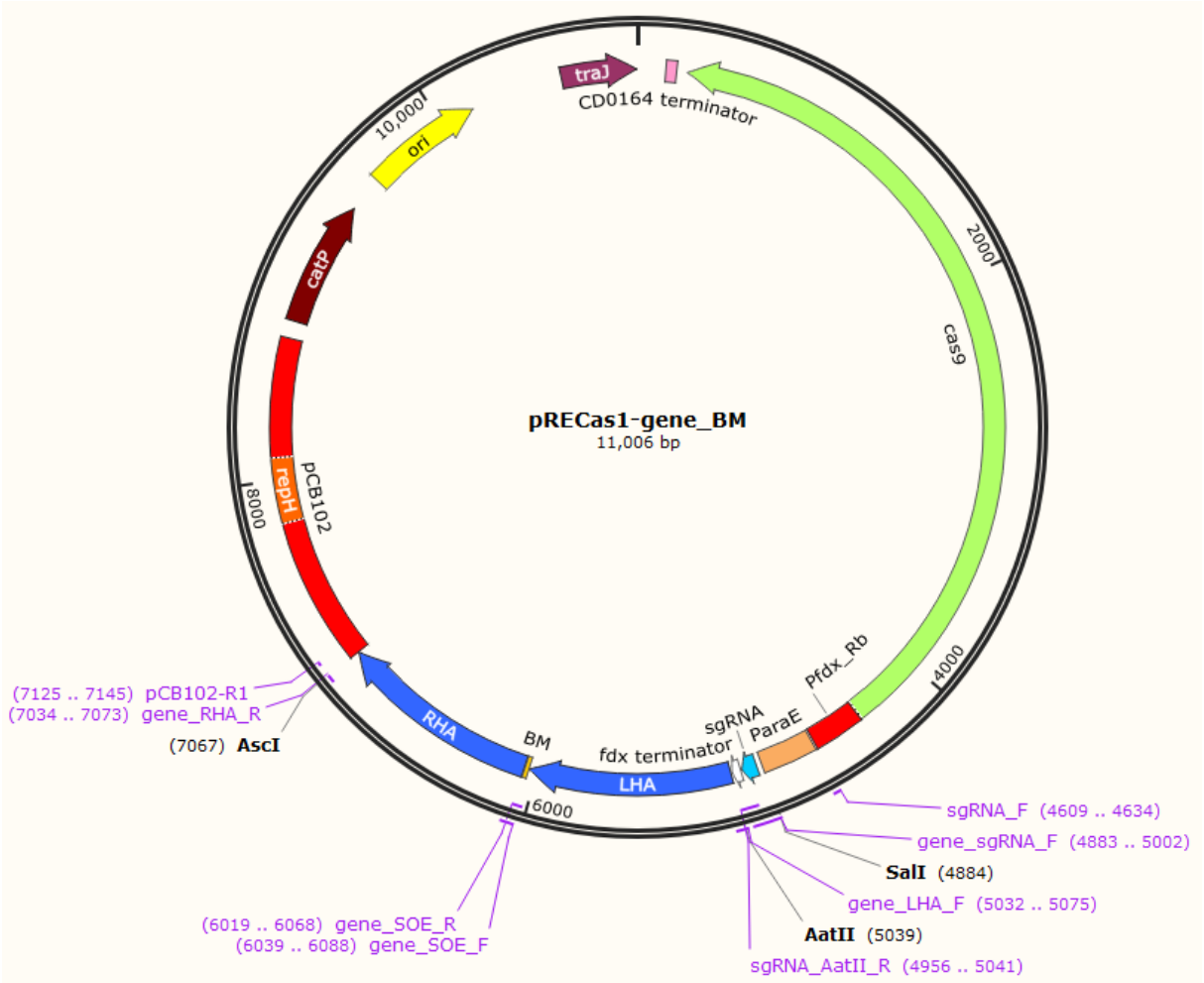


Figure S5: Generic representation of the pRECas1-gene_BM vectors with the RiboCas system. The left homology arm (LHA) and right homology arm (RHA) are flanked by *Aat*II and *Asc*I restriction sites, while the specific sgRNA region is flanked by *Sal*I and *Aat*II. A bookmark was included between the LHA and the RHA. The P_{fdx_Rb} promoter drives the expression of *cas9*, and the P_{araE} promoter drives the expression of the sgRNA region. The primers used to clone and screen the vector are represented.

RiboCas vectors used to revert genes to their original locus



Figure S6: Generic representation of the pRECas1-gene_rev vectors with the RiboCas system. The left homology arm (LHA) and right homology arm (RHA) are flanked by *Aat*II and *Asc*I restriction sites, while the specific sgRNA region is flanked by *Sal*I and *Aat*II. The gene was included between the LHA and the RHA, and the sgRNA targets the bookmark. The P_{fdx_Rb} promoter drives the expression of *cas9*, and the P_{araE} promoter drives the expression of the sgRNA region. The primers used to clone and screen the vector are represented.

Chapter 4: Supporting information

Insertion index distributions

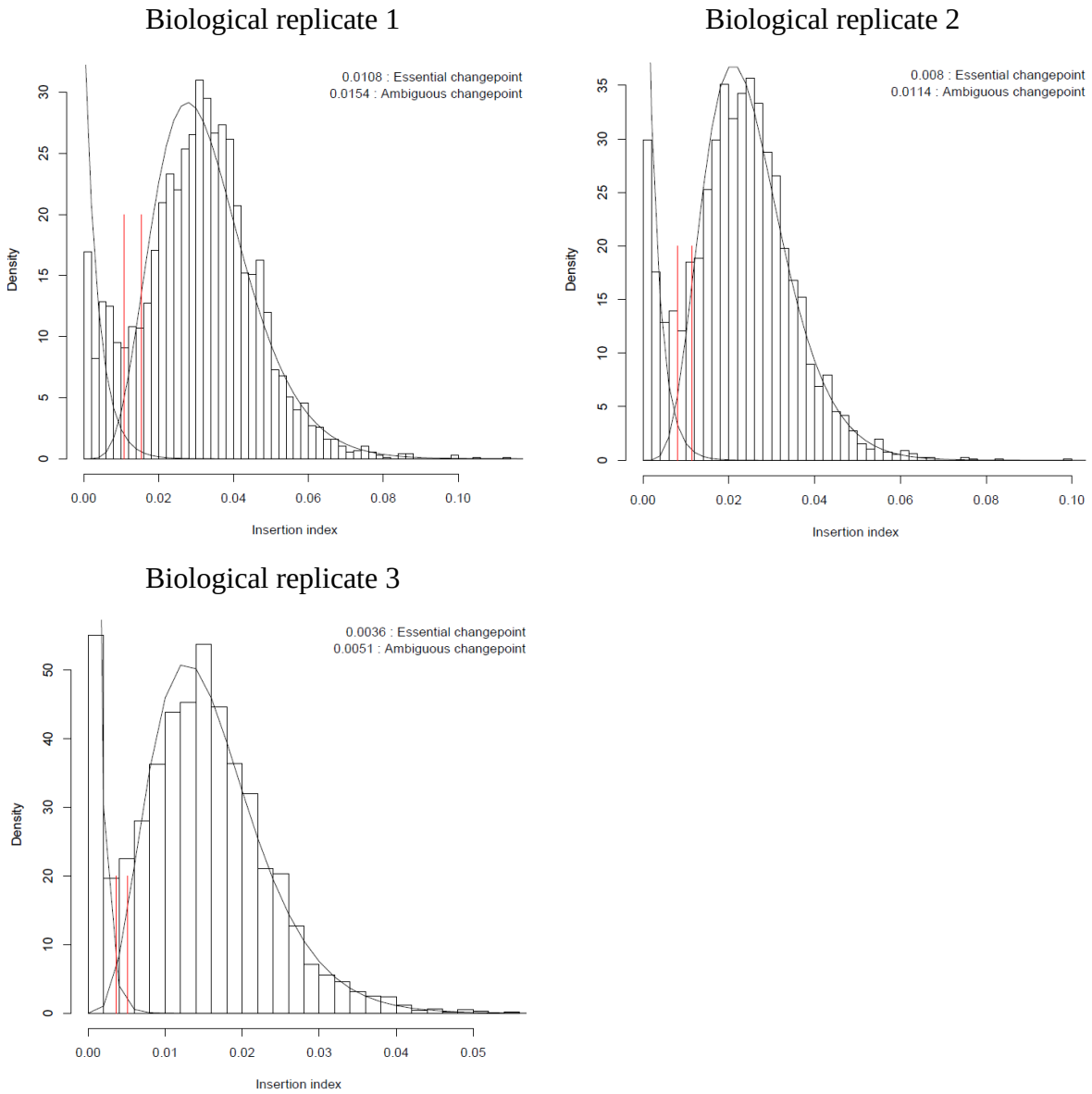


Figure S7: Plots demonstrating the identification of genes essential for vegetative growth of *C. sporogenes*. The plot shows the insertion index distribution from the vegetative growth library. The black lines represent the fitted gamma distributions. The peak on the left represents essential genes, and the first red line represents the insertion index cut-off for essential genes. The broad peak to the right of the second red line represents genes that contain several insertions and are not essential. The second red line represents the cut-off for ambiguous genes.

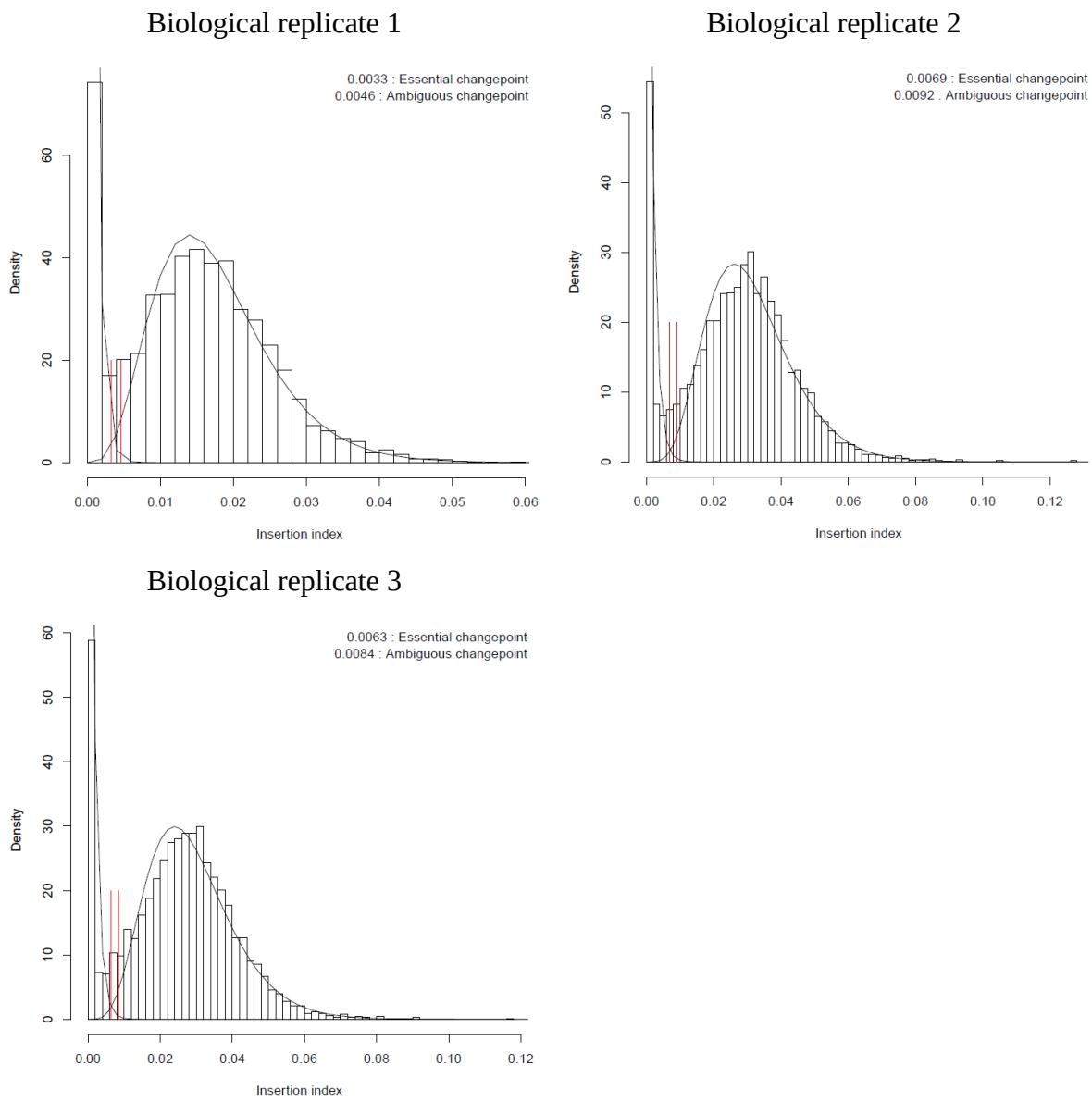


Figure S8: Plot demonstrating the identification of genes essential for sporulation of *C. sporogenes*. The plot shows the insertion index distribution from the sporulation library. The black lines represent the fitted gamma distributions. The peak on the left represents essential genes, and the first red line represents the insertion index cut-off for essential genes. The broad peak to the right of the second red line represents genes that contain several insertions and are not essential. The second red line represents the cut-off for ambiguous genes.

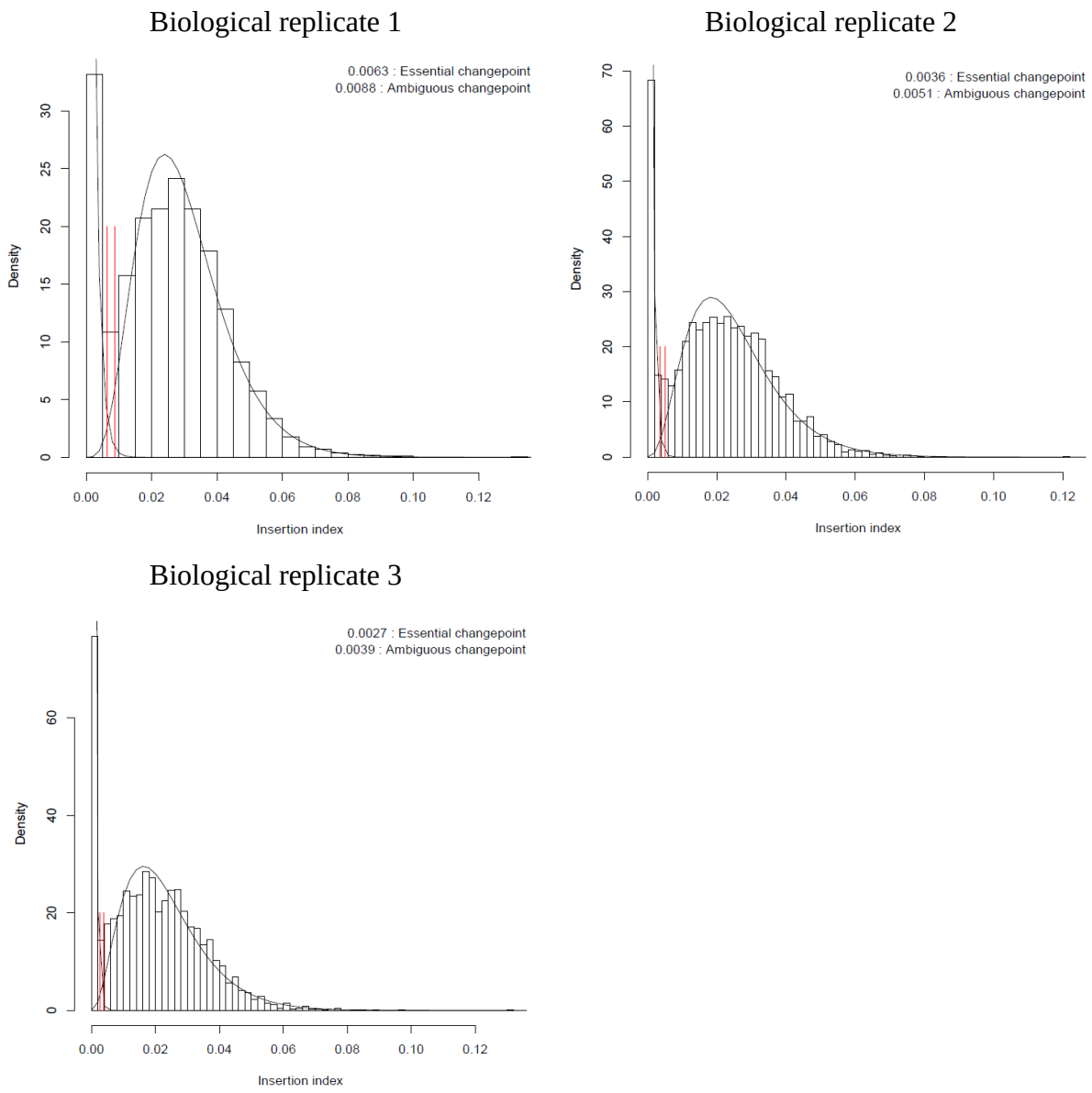


Figure S9: Plot demonstrating the identification of genes essential for germination of *C. sporogenes*. The plot shows the insertion index distribution from the germination library. The black lines represent the fitted gamma distributions. The peak on the left represents essential genes, and the first red line represents the insertion index cut-off for essential genes. The broad peak to the right of the second red line represents genes that contain several insertions and are not essential. The second red line represents the cut-off for ambiguous genes.

List of genes essential for survival

Table S4: List of candidate genes essential for the survival of *C. sporogenes*.

Gene ID	Gene name	Insertion index	Gene function
CLSPO_c00010	*dnaA*	0.0049	chromosomal replication initiator protein DnaA
CLSPO_c00020	*dnaN*	0.0030	DNA polymerase III subunit beta
CLSPO_c00040	*recF1*	0.0071	DNA replication and repair protein RecF
CLSPO_c00060	*gyrB1*	0.0035	DNA gyrase subunit B
CLSPO_c00070	*gyrA1*	0.0049	DNA gyrase subunit A
CLSPO_c00260	*psd*	0.0038	phosphatidylserine decarboxylase proenzyme
CLSPO_c00280	CLSPO_c00280	0.0104	hypothetical protein
CLSPO_c00340	*dnaX*	0.0075	DNA polymerase III subunit gamma/tau
CLSPO_c00350	CLSPO_c00350	0.0032	nucleoid-associated protein
CLSPO_c00360	*recR*	0.0093	recombination protein RecR
CLSPO_c00510	CLSPO_c00510	0.0074	uracil-DNA glycosylase family protein
CLSPO_c00560	CLSPO_c00560	0.0062	hypothetical protein
CLSPO_c00820	CLSPO_c00820	0.0016	thymidylate kinase
CLSPO_c00840	CLSPO_c00840	0.0071	DNA polymerase III subunit delta'
CLSPO_c00870	CLSPO_c00870	0.0000	ferredoxin
CLSPO_c00900	CLSPO_c00900	0.0041	NlpC/P60 family protein
CLSPO_c00910	*abrB*	0.0090	transition state regulatory protein AbrB
CLSPO_c00940	*ispF*	0.0047	2-C-methyl-D-erythritol 2
CLSPO_c01020	*metG*	0.0063	methionine--tRNA ligase MetG
CLSPO_c01200	*nrdD*	0.0093	anaerobic ribonucleoside-triphosphate reductase NrdD
CLSPO_c01210	*nrdG*	0.0086	anaerobic ribonucleoside-triphosphate reductase-activating protein NrdG
CLSPO_c01230	CLSPO_c01230	0.0098	PadR family transcriptional regulator
CLSPO_c01250	*ileS*	0.0025	isoleucine--tRNA ligase IleS
CLSPO_c01460	*ispE*	0.0040	4-diphosphocytidyl-2-C-methyl-D-erythritol kinase IspE
CLSPO_c01520	*ugtP*	0.0045	processive diacylglycerol glucosyltransferase UgtP
CLSPO_c01640	*prfA*	0.0052	peptide chain release factor 1
CLSPO_c01670	*sua*	0.0053	threonylcarbamoyl-AMP synthase Sua
CLSPO_c01690	*rpiB*	0.0076	ribose-5-phosphate isomerase B
CLSPO_c01710	CLSPO_c01710	0.0041	hypothetical protein
CLSPO_c01720	CLSPO_c01720	0.0046	hypothetical protein

Table S4: List of candidate genes essential for the survival of *C. sporogenes* (continuation).

Gene ID	Gene name	Insertion index	Gene function
CLSPO_c01770	*tagO*	0.0064	putative undecaprenyl-phosphate N-acetylglucosaminyl 1-phosphate transferase TagO
CLSPO_c01780	CLSPO_c01780	0.0038	putative UDP-N-acetylglucosamine 2-epimerase
CLSPO_c01860	*atpD1*	0.0088	F0F1-type ATP synthase subunit beta
CLSPO_c01870	*atpC1*	0.0083	F0F1-type ATP synthase epsilon chain
CLSPO_c02010	*metK*	0.0028	S-adenosylmethionine synthase MetK
CLSPO_c02060	*secA*	0.0044	protein translocase subunit SecA
CLSPO_c02130	CLSPO_c02130	0.0016	Ser/Thr protein phosphatase family protein
CLSPO_c02180	*leuS*	0.0041	leucine--tRNA ligase LeuS
CLSPO_c02260	*fba*	0.0036	fructose-bisphosphate aldolase Fba
CLSPO_c02390	*cysE*	0.0093	serine acetyltransferase CysE
CLSPO_c02580	*gap2*	0.0033	glyceraldehyde-3-phosphate dehydrogenase Gap
CLSPO_c02590	*pgk*	0.0000	phosphoglycerate kinase Pgk
CLSPO_c02600	*tpiA*	0.0059	triosephosphate isomerase TpiA
CLSPO_c02610	*gpmI*	0.0029	2,3-bisphosphoglycerate-independent phosphoglycerate mutase GpmI
CLSPO_c02620	*eno*	0.0034	enolase Eno
CLSPO_c02660	*smpB*	0.0047	SsrA-binding protein SmpB
CLSPO_c03250	CLSPO_c03250	0.0099	hypothetical protein
CLSPO_c03620	CLSPO_c03620	0.0082	hypothetical protein
CLSPO_c03690	CLSPO_c03690	0.0082	cyclic lactone autoinducer
CLSPO_c03970	*yycG1*	0.0088	sensor histidine kinase YycG
CLSPO_c04010	*rnfC*	0.0042	electron transport complex protein RnfC
CLSPO_c04020	*rnfD*	0.0083	electron transport complex protein RnfD
CLSPO_c04030	*rnfG*	0.0000	electron transport complex protein RnfG
CLSPO_c04040	*rnfE*	0.0037	electron transport complex protein RnfE
CLSPO_c04050	*rnfA*	0.0077	electron transport complex protein RnfA

Table S4: List of candidate genes essential for the survival of *C. sporogenes* (continuation).

Gene ID	Gene name	Insertion index	Gene function
CLSPO_c04060	*rnfB*	0.0040	electron transport complex protein RnfB
CLSPO_c04100	*apbE1*	0.0075	thiamine biosynthesis lipoprotein ApbE
CLSPO_c04380	CLSPO_c04380	0.0060	tetratricopeptide repeat/DNA binding domain-containing protein
CLSPO_c04420	CLSPO_c04420	0.0061	CobB/CobQ family glutamine amidotransferase
CLSPO_c04430	CLSPO_c04430	0.0025	putative UDP-N-acetylmuramyl tripeptide synthetase
CLSPO_c04540	CLSPO_c04540	0.0031	putative 1-acyl-sn-glycerol-3-phosphate acyltransferase
CLSPO_c04640	*addB*	0.0068	ATP-dependent helicase/deoxyribonuclease subunit B
CLSPO_c04700	*addA*	0.0067	ATP-dependent helicase/nuclease subunit A
CLSPO_c04770	CLSPO_c04770	0.0072	glyoxalase
CLSPO_c04780	CLSPO_c04780	0.0105	hypothetical protein
CLSPO_c04860	*ddl*	0.0037	D-alanine--D-alanine ligase Ddl
CLSPO_c04950	CLSPO_c04950	0.0054	GntR family transcriptional regulator
CLSPO_c05190	CLSPO_c05190	0.0064	hypothetical protein
CLSPO_c05250	CLSPO_c05250	0.0036	ABC-type multidrug transport system
CLSPO_c05260	CLSPO_c05260	0.0069	ABC-type Na+ efflux pump
CLSPO_c05270	CLSPO_c05270	0.0039	ABC-type multidrug transport system
CLSPO_c05510	CLSPO_c05510	0.0000	hypothetical protein
CLSPO_c05730	*nadE*	0.0064	glutamine-dependent NAD(+) synthetase NadE
CLSPO_c06120	CLSPO_c06120	0.0075	hypothetical protein
CLSPO_c06260	CLSPO_c06260	0.0077	DegV domain-containing protein
CLSPO_c06310	*asnS*	0.0016	asparagine--tRNA ligase AsnS
CLSPO_c06450	*vanS4*	0.0040	sensor protein VanS
CLSPO_c06530	*csaA*	0.0101	protein CsaA
CLSPO_c06570	CLSPO_c06570	0.0096	ferredoxin
CLSPO_c07130	*nifH*	0.0104	nitrogenase iron protein NifH
CLSPO_c07410	*macB2*	0.0065	macrolide export ATP-binding/permease protein MacB
CLSPO_c07520	CLSPO_c07520	0.0092	AraC family transcriptional regulator

Table S4: List of candidate genes essential for the survival of *C. sporogenes* (continuation).

Gene ID	Gene name	Insertion index	Gene function
CLSPO_c07600	*luxS*	0.0098	S-ribosylhomocysteine lyase LuxS
CLSPO_c07710	CLSPO_c07710	0.0000	putative DNA-binding protein
CLSPO_c07760	CLSPO_c07760	0.0100	hypothetical protein DUF1659
CLSPO_c07770	CLSPO_c07770	0.0097	hypothetical protein DUF2922
CLSPO_c08030	CLSPO_c08030	0.0093	hypothetical protein
CLSPO_c08260	CLSPO_c08260	0.0000	coat F domain-containing protein
CLSPO_c08320	CLSPO_c08320	0.0028	exopolysaccharide biosynthesis protein
CLSPO_c08330	*degT*	0.0050	pleiotropic regulatory protein DegT
CLSPO_c08350	CLSPO_c08350	0.0057	putative transcriptional regulator
CLSPO_c08370	CLSPO_c08370	0.0000	hypothetical protein
CLSPO_c08570	CLSPO_c08570	0.0070	transcriptional regulator
CLSPO_c08640	CLSPO_c08640	0.0069	pyridoxamine 5'-phosphate oxidase-like FMN-binding protein
CLSPO_c08650	CLSPO_c08650	0.0078	membrane protein
CLSPO_c08660	CLSPO_c08660	0.0047	putative nimA protein
CLSPO_c08800	CLSPO_c08800	0.0043	hypothetical protein
CLSPO_c08880	CLSPO_c08880	0.0040	DNA-binding protein
CLSPO_c09250	CLSPO_c09250	0.0095	putative acetyltransferase
CLSPO_c09490	CLSPO_c09490	0.0105	hypothetical protein
CLSPO_c10330	CLSPO_c10330	0.0106	Fe2+ transport system protein A
CLSPO_c10850	*argS*	0.0020	arginine--tRNA ligase ArgS
CLSPO_c10900	*feoA2*	0.0100	ferrous iron transport protein A
CLSPO_c11040	*gltX*	0.0030	glutamate--tRNA ligase GltX
CLSPO_c11270	CLSPO_c11270	0.0093	hypothetical protein
CLSPO_c11300	*mgtE2*	0.0098	Mg/Co/Ni transporter MgtE
CLSPO_c11540	CLSPO_c11540	0.0079	hypothetical protein
CLSPO_c11710	*fchA*	0.0070	methenyltetrahydrofolate cyclohydrolase FchA
CLSPO_c11720	*folD*	0.0105	bifunctional 5
CLSPO_c11900	CLSPO_c11900	0.0066	hypothetical protein DUF3787
CLSPO_c12140	CLSPO_c12140	0.0053	beta-amylase
CLSPO_c12340	CLSPO_c12340	0.0077	hypothetical protein
CLSPO_c12520	*sspH2*	0.0058	small, acid-soluble spore protein H 2
CLSPO_c12540	CLSPO_c12540	0.0000	hypothetical protein
CLSPO_c12690	*grdA*	0.0093	glycine/sarcosine/betaine reductase complex component A

Table S4: List of candidate genes essential for the survival of *C. sporogenes* (continuation).

Gene ID	Gene name	Insertion index	Gene function
CLSPO_c13170	CLSPO_c13170	0.0000	putative zinc finger domain-containing protein-containing protein
CLSPO_c13410	CLSPO_c13410	0.0000	hypothetical protein
CLSPO_c13750	CLSPO_c13750	0.0053	hypothetical protein
CLSPO_c14280	CLSPO_c14280	0.0000	cold-shock DNA-binding protein family
CLSPO_c14530	*pncB*	0.0052	nicotinate phosphoribosyltransferase PncB
CLSPO_c14600	*cpsA2*	0.0028	thermostable carboxypeptidase 1
CLSPO_c14630	*spoVB1*	0.0041	stage V sporulation protein B
CLSPO_c14660	CLSPO_c14660	0.0054	hypothetical protein
CLSPO_c14840	*mraZ*	0.0103	protein MraZ
CLSPO_c14860	*ftsL1*	0.0066	putative cell division protein FtsL
CLSPO_c14870	*spoVD1*	0.0066	stage V sporulation protein D
CLSPO_c14880	*murE*	0.0031	UDP-N-acetylmuramoyl-L-alanyl-D-glutamate--2
CLSPO_c14890	*murF*	0.0016	UDP-N-acetylmuramoyl-tripeptide--D-alanyl-D-alanine ligase MurF
CLSPO_c14900	*mraY2*	0.0000	phospho-N-acetylmuramoyl-pentapeptide-transferase MraY
CLSPO_c14910	*ftsW*	0.0030	lipid II flippase FtsW
CLSPO_c14920	*divIB*	0.0029	cell division protein DivIB
CLSPO_c14960	CLSPO_c14960	0.0066	hypothetical protein UPF0001
CLSPO_c14970	*sepF*	0.0074	cell division protein SepF
CLSPO_c15020	*mtnN*	0.0032	5'-methylthioadenosine/S-adenosylhomocysteine nucleosidase MtnN
CLSPO_c15180	CLSPO_c15180	0.0071	hypothetical protein
CLSPO_c16000	*queC*	0.0034	7-cyano-7-deazaguanine synthase QueC
CLSPO_c16790	CLSPO_c16790	0.0080	hypothetical protein
CLSPO_c17580	CLSPO_c17580	0.0000	hypothetical protein
CLSPO_c17830	*xerC1*	0.0068	tyrosine recombinase XerC
CLSPO_c17850	*lexA*	0.0018	LexA repressor
CLSPO_c18380	*drrA*	0.0105	daunorubicin/doxorubicin resistance ATP-binding protein DrrA
CLSPO_c18710	*argR*	0.0098	arginine repressor

Table S4: List of candidate genes essential for the survival of *C. sporogenes* (continuation).

Gene ID	Gene name	Insertion index	Gene function
CLSPO_c18720	*ppnK*	0.0040	putative inorganic polyphosphate/ATP-NAD kinase PpnK
CLSPO_c18740	*dxs*	0.0018	1-deoxy-D-xylulose-5-phosphate synthase Dxs
CLSPO_c18800	CLSPO_c18800	0.0028	alkaline shock protein
CLSPO_c18900	*efp*	0.0000	elongation factor P
CLSPO_c19040	*dapF*	0.0055	diaminopimelate epimerase DapF
CLSPO_c19550	*ppsA*	0.0104	putative phosphoenolpyruvate synthase PpsA
CLSPO_c19560	CLSPO_c19560	0.0073	putative phage DNA packaging protein
CLSPO_c19580	CLSPO_c19580	0.0067	transposase
CLSPO_c19610	*stc7*	0.0076	signal-transduction and transcriptional-control protein Stc
CLSPO_c19620	CLSPO_c19620	0.0087	glutaredoxin-like protein
CLSPO_c19650	*gltT2*	0.0087	proton/sodium-glutamate symport protein GltT
CLSPO_c19660	*patB2*	0.0066	cystathionine beta-lyase PatB
CLSPO_c19670	CLSPO_c19670	0.0034	hypothetical protein UPF0597
CLSPO_c19860	CLSPO_c19860	0.0103	putative redox protein C_GCAxxG_C_C family
CLSPO_c20360	CLSPO_c20360	0.0050	hypothetical protein
CLSPO_c20440	CLSPO_c20440	0.0092	proline dipeptidase
CLSPO_c20620	CLSPO_c20620	0.0084	hypothetical protein
CLSPO_c20660	*xylB*	0.0098	xylulose kinase XylB
CLSPO_c20870	CLSPO_c20870	0.0062	hypothetical protein DUF2620
CLSPO_c21050	CLSPO_c21050	0.0068	hypothetical protein DUF2442
CLSPO_c21080	CLSPO_c21080	0.0105	putative AAA-ATPase
CLSPO_c21190	CLSPO_c21190	0.0038	ethanolamine utilization protein PduA/PduJ
CLSPO_c21200	*pduL*	0.0102	phosphate propanoyltransferase PduL
CLSPO_c21210	*adhE2*	0.0097	aldehyde-alcohol dehydrogenase AdhE
CLSPO_c21280	CLSPO_c21280	0.0099	GTP-binding protein
CLSPO_c21290	*pduU*	0.0094	propanediol utilization protein PduU
CLSPO_c21340	CLSPO_c21340	0.0075	microcompartment family protein
CLSPO_c21350	CLSPO_c21350	0.0083	hypothetical protein

Table S4: List of candidate genes essential for the survival of *C. sporogenes* (continuation).

Gene ID	Gene name	Insertion index	Gene function
CLSPO_c21780	*thiS*	0.0056	thiamine biosynthesis protein ThiS
CLSPO_c21930	CLSPO_c21930	0.0101	PTS system IIA component
CLSPO_c22110	CLSPO_c22110	0.0060	phenazine biosynthesis protein
CLSPO_c22280	*bcd1*	0.0098	acyl-CoA dehydrogenase
CLSPO_c22420	CLSPO_c22420	0.0069	multimeric flavodoxin WrbA
CLSPO_c22630	*hypA1*	0.0032	hydrogenase nickel incorporation protein HypA
CLSPO_c22710	CLSPO_c22710	0.0038	hypothetical protein DUF4352
CLSPO_c22860	CLSPO_c22860	0.0082	hypothetical protein
CLSPO_c23120	CLSPO_c23120	0.0102	macro domain-containing protein
CLSPO_c23800	CLSPO_c23800	0.0103	AraC family transcriptional regulator
CLSPO_c23830	CLSPO_c23830	0.0067	alpha/beta hydrolase family protein
CLSPO_c23860	CLSPO_c23860	0.0047	transcription activator effector binding protein
CLSPO_c24030	CLSPO_c24030	0.0000	hypothetical protein
CLSPO_c24090	CLSPO_c24090	0.0000	hypothetical protein DUF2922
CLSPO_c24240	CLSPO_c24240	0.0077	hypothetical protein
CLSPO_c24310	CLSPO_c24310	0.0022	phage protein
CLSPO_c24840	CLSPO_c24840	0.0063	hypothetical protein
CLSPO_c24870	CLSPO_c24870	0.0091	putative phage protein
CLSPO_c24880	CLSPO_c24880	0.0036	hypothetical protein
CLSPO_c24890	CLSPO_c24890	0.0072	phage regulatory protein
CLSPO_c24900	CLSPO_c24900	0.0077	hypothetical protein
CLSPO_c24910	*znuC*	0.0058	zinc import ATP-binding protein ZnuC
CLSPO_c24920	CLSPO_c24920	0.0089	putative transcriptional regulator
CLSPO_c25020	*spoVS*	0.0043	stage V sporulation protein S
CLSPO_c25030	*rny*	0.0007	ribonuclease Y
CLSPO_c25040	CLSPO_c25040	0.0042	RecA/RadA recombinase
CLSPO_c25090	*dapG*	0.0046	aspartokinase 1
CLSPO_c25100	CLSPO_c25100	0.0081	PRC-barrel domain-containing protein
CLSPO_c25120	*pnp*	0.0037	polyribonucleotide nucleotidyltransferase Pnp
CLSPO_c25130	*rpsO*	0.0000	30S ribosomal protein S15
CLSPO_c25140	*ribC*	0.0060	riboflavin biosynthesis protein RibC
CLSPO_c25160	*nrnA*	0.0069	bifunctional oligoribonuclease and PAP phosphatase NrnA

Table S4: List of candidate genes essential for the survival of *C. sporogenes* (continuation).

Gene ID	Gene name	Insertion index	Gene function
CLSPO_c25180	*infB*	0.0043	translation initiation factor IF-2
CLSPO_c25200	CLSPO_c25200	0.0000	putative nucleic-acid-binding protein
CLSPO_c25210	*nusA*	0.0049	transcription termination/antitermination protein NusA
CLSPO_c25230	*polC*	0.0031	DNA polymerase III PolC-type
CLSPO_c25240	*ispG*	0.0021	4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase IspG
CLSPO_c25260	*dxr*	0.0038	1-deoxy-D-xylulose 5-phosphate reductoisomerase Dxr
CLSPO_c25280	*cdsA*	0.0083	phosphatidate cytidylyltransferase CdsA
CLSPO_c25290	*uppS2*	0.0000	isoprenyl transferase UppS
CLSPO_c25300	*frr*	0.0080	ribosome-recycling factor Frr
CLSPO_c25310	*pyrH*	0.0031	uridylate kinase PyrH
CLSPO_c25320	*tsf*	0.0024	elongation factor Ts
CLSPO_c25330	*rpsB*	0.0047	30S ribosomal protein S2
CLSPO_c25340	*codY*	0.0043	GTP-sensing transcriptional pleiotropic repressor CodY
CLSPO_c25350	*topA*	0.0048	DNA topoisomerase 1
CLSPO_c25400	*rbgA*	0.0025	ribosome biogenesis GTPase A
CLSPO_c25410	*lepB*	0.0106	putative signal peptidase I-1
CLSPO_c25420	*rplS*	0.0032	50S ribosomal protein L19
CLSPO_c25430	*trmD*	0.0031	tRNA (guanine-N(1)-)-methyltransferase TrmD
CLSPO_c25440	*rimM*	0.0000	ribosome maturation factor RimM
CLSPO_c25450	CLSPO_c25450	0.0049	hypothetical protein UPF0109
CLSPO_c25460	rpsP	0.0000	30S ribosomal protein S16
CLSPO_c25470	ffh	0.0008	signal recognition particle protein Ffh
CLSPO_c25480	CLSPO_c25480	0.0035	hypothetical protein UPF0122
CLSPO_c25490	*ftsY*	0.0012	signal recognition particle receptor FtsY
CLSPO_c25550	*plsX*	0.0033	phosphate acyltransferase PlsX
CLSPO_c25560	*rpmF*	0.0000	50S ribosomal protein L32
CLSPO_c25580	*ackA*	0.0093	acetate kinase AckA
CLSPO_c25590	*pta*	0.0044	phosphate acetyltransferase Pta
CLSPO_c25610	*prdB2*	0.0093	D-proline reductase subunit gamma

Table S4: List of candidate genes essential for the survival of *C. sporogenes* (continuation).

Gene ID	Gene name	Insertion index	Gene function
CLSPO_c25620	*prdH1*	0.0080	D-proline reducatse cluster associated protein
CLSPO_c25770	*prdA4*	0.0101	D-proline reductase proprotein PrdA
CLSPO_c25820	CLSPO_c25820	0.0078	proline reductase operon associated protein
CLSPO_c25880	*coaD*	0.0022	phosphopantetheine adenylyltransferase CoaD
CLSPO_c25910	CLSPO_c25910	0.0041	DAK2 domain-containing protein
CLSPO_c25920	CLSPO_c25920	0.0095	hypothetical protein
CLSPO_c25930	*rpmB*	0.0000	50S ribosomal protein L28
CLSPO_c25950	CLSPO_c25950	0.0069	thiamine pyrophosphokinase
CLSPO_c25960	*rpe*	0.0085	ribulose-phosphate 3-epimerase Rpe
CLSPO_c25980	CLSPO_c25980	0.0017	putative serine/threonine-protein kinase
CLSPO_c26030	*fmt*	0.0024	methionyl-tRNA formyltransferase Fmt
CLSPO_c26040	*def3*	0.0100	peptide deformylase Def
CLSPO_c26050	*priA*	0.0060	primosomal protein N'
CLSPO_c26060	*coaBC*	0.0028	putative coenzyme A biosynthesis bifunctional protein CoaBC
CLSPO_c26070	*rpoZ*	0.0051	DNA-directed RNA polymerase subunit omega
CLSPO_c26080	*gmk*	0.0018	guanylate kinase Gmk
CLSPO_c26090	CLSPO_c26090	0.0000	hypothetical protein UPF0296
CLSPO_c26150	*der*	0.0051	GTPase Der
CLSPO_c26300	*ftsZ*	0.0040	cell division protein FtsZ
CLSPO_c26440	*fur*	0.0098	ferric uptake regulation protein Fur
CLSPO_c26460	CLSPO_c26460	0.0000	putative holliday junction resolvase
CLSPO_c26470	CLSPO_c26470	0.0044	hypothetical protein UPF0297
CLSPO_c26480	*alaS*	0.0051	alanine--tRNA ligase AlaS
CLSPO_c26510	*nifU*	0.0078	NifU-like protein NifU
CLSPO_c26520	*iscS1*	0.0037	cysteine desulfurase IscS
CLSPO_c26530	*cymR2*	0.0050	HTH-type transcriptional regulator CymR
CLSPO_c26650	*arcB2*	0.0033	ornithine carbamoyltransferase ArcB
CLSPO_c26660	*arcC*	0.0106	carbamate kinase ArcC
CLSPO_c26990	*atpD2*	0.0034	V-type ATP synthase subunit D

Table S4: List of candidate genes essential for the survival of *C. sporogenes* (continuation).

Gene ID	Gene name	Insertion index	Gene function
CLSPO_c27000	*atpB2*	0.0080	V-type ATP synthase beta chain
CLSPO_c27010	*atpA2*	0.0081	V-type ATP synthase alpha chain
CLSPO_c27030	*atpC2*	0.0078	V-type sodium ATPase subunit C
CLSPO_c27040	*atpE*	0.0056	V-type proton ATPase subunit E
CLSPO_c27050	*atpK*	0.0000	V-type sodium ATPase subunit K
CLSPO_c27060	*atpI*	0.0074	V-type sodium ATPase subunit I
CLSPO_c27070	*atpG2*	0.0068	putative V-type ATPase
CLSPO_c27130	CLSPO_c27130	0.0027	cobalt-dependent inorganic pyrophosphatase
CLSPO_c27170	CLSPO_c27170	0.0094	hypothetical protein
CLSPO_c27450	CLSPO_c27450	0.0090	hypothetical protein
CLSPO_c27730	CLSPO_c27730	0.0063	putative motility protein
CLSPO_c28050	*fliT*	0.0065	flagellar protein FliT
CLSPO_c28320	CLSPO_c28320	0.0102	hypothetical protein
CLSPO_c28330	*nifJ2*	0.0031	pyruvate-flavodoxin oxidoreductase NifJ
CLSPO_c28340	*murG*	0.0042	UDP-N-acetylglucosamine--N-acetylmuramyl-(pentapeptide) pyrophosphoryl-undecaprenol N-acetylglucosamine transferase MurG
CLSPO_c28350	CLSPO_c28350	0.0000	small acid-soluble spore protein
CLSPO_c28450	*gyrA2*	0.0058	DNA gyrase subunit A
CLSPO_c28460	*gyrB2*	0.0028	DNA gyrase subunit B
CLSPO_c28800	CLSPO_c28800	0.0054	universal stress protein family protein
CLSPO_c28870	*nrdB*	0.0043	ribonucleoside-diphosphate reductase subunit beta
CLSPO_c28880	*nrdA*	0.0048	ribonucleoside-diphosphate reductase subunit alpha
CLSPO_c29030	CLSPO_c29030	0.0000	putative flagellin B
CLSPO_c29050	CLSPO_c29050	0.0076	MutT/nudix family protein
CLSPO_c29220	*sudA3*	0.0083	sulfide dehydrogenase subunit alpha
CLSPO_c29230	*hypB2*	0.0083	hydrogenase nickel incorporation protein HypB
CLSPO_c29260	*lolD2*	0.0097	lipoprotein-releasing system ATP-binding protein LolD
CLSPO_c29270	*resE9*	0.0034	sensor histidine kinase ResE
CLSPO_c29280	*regX3*	0.0065	sensory transduction protein RegX3

Table S4: List of candidate genes essential for the survival of *C. sporogenes* (continuation).

Gene ID	Gene name	Insertion index	Gene function
CLSPO_c29770	*mnmA2*	0.0042	tRNA-specific 2-thiouridylase MnmA 2
CLSPO_c30000	CLSPO_c30000	0.0030	hypothetical protein
CLSPO_c30060	CLSPO_c30060	0.0065	hypothetical protein
CLSPO_c30070	CLSPO_c30070	0.0038	hypothetical protein
CLSPO_c30130	CLSPO_c30130	0.0100	hypothetical protein DUF1659
CLSPO_c30210	CLSPO_c30210	0.0050	putative transcriptional regulator
CLSPO_c30280	*sigA*	0.0010	RNA polymerase sigma factor SigA
CLSPO_c30290	*dnaG*	0.0050	DNA primase DnaG
CLSPO_c30350	*recO*	0.0029	DNA repair protein RecO
CLSPO_c30360	*era*	0.0087	GTPase Era
CLSPO_c30390	*ybeY*	0.0044	putative rRNA maturation factor YbeY
CLSPO_c30440	CLSPO_c30440	0.0000	SSU ribosomal protein S21P
CLSPO_c30490	*dnaJ*	0.0029	chaperone protein DnaJ
CLSPO_c30500	*dnaK*	0.0047	chaperone protein DnaK
CLSPO_c30510	*grpE*	0.0017	protein GrpE
CLSPO_c30520	*hrcA*	0.0065	heat-inducible transcription repressor HrcA
CLSPO_c30580	*rpsT*	0.0083	30S ribosomal protein S20
CLSPO_c30740	CLSPO_c30740	0.0078	HDIG domain-containing protein
CLSPO_c30750	*nadD*	0.0037	putative nicotinate-nucleotide adenylyltransferase NadD
CLSPO_c30770	*obg*	0.0026	GTPase Obg
CLSPO_c30780	*rpmA*	0.0037	50S ribosomal protein L27
CLSPO_c30790	CLSPO_c30790	0.0034	putative ribosomal protein
CLSPO_c30800	*rplU*	0.0028	50S ribosomal protein L21
CLSPO_c30870	*spoVE*	0.0030	stage V sporulation protein E
CLSPO_c30880	*minE*	0.0042	cell division topological specificity factor MinE
CLSPO_c30910	CLSPO_c30910	0.0060	penicillin-binding protein
CLSPO_c30930	*mreC*	0.0039	cell shape-determining protein MreC
CLSPO_c30940	*mreB*	0.0011	Rod shape-determining protein MreB
CLSPO_c31030	*coaE*	0.0069	dephospho-CoA kinase CoaE
CLSPO_c31390	CLSPO_c31390	0.0027	DNA-binding helix-turn-helix protein

Table S4: List of candidate genes essential for the survival of *C. sporogenes* (continuation).

Gene ID	Gene name	Insertion index	Gene function
CLSPO_c31440	CLSPO_c31440	0.0093	hypothetical protein
CLSPO_c31460	CLSPO_c31460	0.0085	hypothetical protein
CLSPO_c31470	CLSPO_c31470	0.0026	sigma-E factor regulator
CLSPO_c31480	*aspS2*	0.0075	aspartate--tRNA ligase AspS
CLSPO_c31490	*hisS*	0.0080	histidine--tRNA ligase HisS
CLSPO_c31590	CLSPO_c31590	0.0000	hypothetical protein
CLSPO_c31640	*ruvB*	0.0022	holliday junction ATP-dependent DNA helicase RuvB
CLSPO_c31650	*ruvA*	0.0037	holliday junction ATP-dependent DNA helicase RuvA
CLSPO_c31820	*spoIIAB*	0.0105	anti-sigma F factor
CLSPO_c31880	*tagA*	0.0077	putative N-acetylmannosaminyltransferase TagA
CLSPO_c31890	*rfbP*	0.0051	undecaprenyl-phosphate galactose phosphotransferase RfbP
CLSPO_c31900	CLSPO_c31900	0.0086	putative polysaccharide transporter
CLSPO_c31920	CLSPO_c31920	0.0100	glycosyl transferase
CLSPO_c31930	CLSPO_c31930	0.0018	exopolysaccharide biosynthesis family protein
CLSPO_c31950	CLSPO_c31950	0.0028	group 1 glycosyl transferase
CLSPO_c31960	CLSPO_c31960	0.0021	succinoglycan biosynthesis protein
CLSPO_c31970	CLSPO_c31970	0.0054	lipid A core-O-antigen ligase
CLSPO_c31980	CLSPO_c31980	0.0050	sugar-binding protein
CLSPO_c32000	CLSPO_c32000	0.0040	group 1 glycosyl transferase
CLSPO_c32010	CLSPO_c32010	0.0045	putative acetyltransferase
CLSPO_c32020	CLSPO_c32020	0.0040	NAD-binding rossmann fold family oxidoreductase
CLSPO_c32030	*wbpA*	0.0034	UDP-N-acetyl-D-glucosamine 6-dehydrogenase WbpA
CLSPO_c32040	*epsC*	0.0024	putative polysaccharide biosynthesis protein EpsC
CLSPO_c32200	*mutS4*	0.0084	DNA mismatch repair protein MutS
CLSPO_c32290	*zapA*	0.0055	cell division protein ZapA
CLSPO_c32300	*pheT*	0.0033	phenylalanine--tRNA ligase beta subunit
CLSPO_c32310	*pheS*	0.0076	phenylalanine--tRNA ligase alpha subunit
CLSPO_c32350	*rplT*	0.0062	50S ribosomal protein L20
CLSPO_c32360	*rpmI*	0.0000	50S ribosomal protein L35

Table S4: List of candidate genes essential for the survival of *C. sporogenes* (continuation).

Gene ID	Gene name	Insertion index	Gene function
CLSPO_c32370	*infC*	0.0045	translation initiation factor IF-3
CLSPO_c32380	*thrS*	0.0017	threonine--tRNA ligase ThrS
CLSPO_c32560	*asd2*	0.0034	aspartate-semialdehyde dehydrogenase Asd
CLSPO_c32570	*dapA*	0.0063	4-hydroxy-tetrahydrodipicolinate synthase DapA
CLSPO_c32580	*dapB2*	0.0029	4-hydroxy-tetrahydrodipicolinate reductase DapB
CLSPO_c32590	*patA*	0.0104	putative N-acetyl-LL-diaminopimelate aminotransferase PatA
CLSPO_c32600	*dapH*	0.0063	2,3,4,5-tetrahydropyridine-2,6-dicarboxylate N-acetyltransferase DapH
CLSPO_c32680	*valS*	0.0029	valine--tRNA ligase ValS
CLSPO_c32690	CLSPO_c32690	0.0088	GNAT family acetyltransferase
CLSPO_c33060	CLSPO_c33060	0.0000	hypothetical protein
CLSPO_c33160	*rbr3A*	0.0061	reverse rubrerythrin-1
CLSPO_c33300	*engB*	0.0000	putative GTP-binding protein EngB
CLSPO_c33870	*gatB*	0.0031	aspartyl/glutamyl-tRNA(Asn/Gln) amidotransferase subunit B
CLSPO_c33880	*gatA*	0.0084	glutamyl-tRNA(Gln) amidotransferase subunit A
CLSPO_c33890	*gatC*	0.0077	aspartyl/glutamyl-tRNA(Asn/Gln) amidotransferase subunit C
CLSPO_c33920	*ligA*	0.0028	DNA ligase LigA
CLSPO_c34000	*pgi*	0.0033	glucose-6-phosphate isomerase Pgi
CLSPO_c34010	CLSPO_c34010	0.0000	putative zinc finger domain-containing protein-containing protein
CLSPO_c34040	*fldH*	0.0100	phenyllactate dehydrogenase FldH
CLSPO_c34180	*groL*	0.0048	60 kDa chaperonin
CLSPO_c34190	*groS*	0.0038	10 kDa chaperonin
CLSPO_c34330	*tsaD*	0.0011	tRNA N6-adenosine threonylcarbamoyltransferase TsaD
CLSPO_c34340	CLSPO_c34340	0.0036	endodeoxyribonuclease RusA family protein
CLSPO_c34450	*tsaB*	0.0062	tRNA threonylcarbamoyladenosine biosynthesis protein TsaB

Table S4: List of candidate genes essential for the survival of *C. sporogenes* (continuation).

Gene ID	Gene name	Insertion index	Gene function
CLSPO_c34670	*proS1*	0.0031	proline--tRNA ligase ProS
CLSPO_c34790	CLSPO_c34790	0.0085	insect kinin peptide
CLSPO_c34820	*celD*	0.0037	cellobiose-specific phosphotransferase enzyme IIA component CelD
CLSPO_c35040	CLSPO_c35040	0.0028	hypothetical protein DUF4284
CLSPO_c35150	CLSPO_c35150	0.0105	putative NADH-flavin reductase
CLSPO_c35280	*pyk*	0.0069	pyruvate kinase Pyk
CLSPO_c35290	*pfkA*	0.0046	6-phosphofructokinase PfkA
CLSPO_c35300	*dnaE*	0.0028	DNA polymerase III subunit alpha
CLSPO_c35310	*whiA*	0.0105	putative sporulation transcription regulator WhiA
CLSPO_c35340	*murB*	0.0012	UDP-N-acetylenolpyruvoylglucosamine reductase MurB
CLSPO_c35400	CLSPO_c35400	0.0101	FtsW-like protein
CLSPO_c35460	*ftsX*	0.0062	cell division protein FtsX
CLSPO_c35470	*ftsE*	0.0000	cell division ATP-binding protein FtsE
CLSPO_c35500	CLSPO_c35500	0.0059	transketolase
CLSPO_c35510	CLSPO_c35510	0.0027	transketolase
CLSPO_c35540	*alr3*	0.0067	alanine racemase 1
CLSPO_c35570	*acpS*	0.0087	holo-[acyl-carrier-protein] synthase AcpS
CLSPO_c35720	*glmS*	0.0030	glutamine--fructose-6-phosphate aminotransferase
CLSPO_c35780	*srlA*	0.0061	glucitol/sorbitol permease IIC component SrlA
CLSPO_c35820	*glmM*	0.0000	phosphoglucosamine mutase GlmM
CLSPO_c35850	*vorB*	0.0083	ketoisovalerate oxidoreductase subunit VorB
CLSPO_c35960	*trxA2*	0.0105	thioredoxin TrxA
CLSPO_c36070	*rpsI*	0.0000	30S ribosomal protein S9
CLSPO_c36080	*rplM*	0.0026	50S ribosomal protein L13
CLSPO_c36120	*ecfA1*	0.0092	energy-coupling factor transporter ATP-binding protein EcfA
CLSPO_c36130	*rplQ*	0.0000	50S ribosomal protein L17
CLSPO_c36140	*rpoA*	0.0047	DNA-directed RNA polymerase subunit alpha

Table S4: List of candidate genes essential for the survival of *C. sporogenes* (continuation).

Gene ID	Gene name	Insertion index	Gene function
CLSPO_c36150	*rpsD1*	0.0018	30S ribosomal protein S4
CLSPO_c36160	*rpsK*	0.0028	30S ribosomal protein S11
CLSPO_c36170	*rpsM*	0.0030	30S ribosomal protein S13
CLSPO_c36180	CLSPO_c36180	0.0000	LSU ribosomal protein L36P
CLSPO_c36190	*infA*	0.0051	translation initiation factor IF-1
CLSPO_c36200	CLSPO_c36200	0.0039	LSU ribosomal protein L14E
CLSPO_c36210	*map1*	0.0030	methionine aminopeptidase Map
CLSPO_c36220	*adk*	0.0051	adenylate kinase Adk
CLSPO_c36230	*secY*	0.0035	protein translocase subunit SecY
CLSPO_c36240	*rplO*	0.0000	50S ribosomal protein L15
CLSPO_c36260	*rpsE*	0.0022	30S ribosomal protein S5
CLSPO_c36270	*rplR*	0.0000	50S ribosomal protein L18
CLSPO_c36280	*rplF*	0.0041	50S ribosomal protein L6
CLSPO_c36290	*rpsH*	0.0083	30S ribosomal protein S8
CLSPO_c36300	*rpsZ*	0.0060	30S ribosomal protein S14 type Z
CLSPO_c36310	*rplE*	0.0020	50S ribosomal protein L5
CLSPO_c36320	*rplX*	0.0070	50S ribosomal protein L24
CLSPO_c36330	*rplN*	0.0060	50S ribosomal protein L14
CLSPO_c36340	*rpsQ*	0.0000	30S ribosomal protein S17
CLSPO_c36350	*rpmC*	0.0000	50S ribosomal protein L29
CLSPO_c36360	*rplP*	0.0000	50S ribosomal protein L16
CLSPO_c36370	*rpsC*	0.0017	30S ribosomal protein S3
CLSPO_c36380	*rplV*	0.0066	50S ribosomal protein L22
CLSPO_c36390	*rpsS*	0.0000	30S ribosomal protein S19
CLSPO_c36400	*rplB*	0.0053	50S ribosomal protein L2
CLSPO_c36410	*rplW*	0.0075	50S ribosomal protein L23
CLSPO_c36420	*rplD*	0.0018	50S ribosomal protein L4
CLSPO_c36430	*rplC*	0.0035	50S ribosomal protein L3
CLSPO_c36440	*rpsJ*	0.0036	30S ribosomal protein S10
CLSPO_c36450	*tuf2*	0.0047	elongation factor Tu
CLSPO_c36460	*fusA1*	0.0021	elongation factor G
CLSPO_c36470	*rpsG*	0.0047	30S ribosomal protein S7
CLSPO_c36480	*rpsL*	0.0000	30S ribosomal protein S12
CLSPO_c36490	CLSPO_c36490	0.0000	ribosomal protein L7Ae family protein
CLSPO_c36500	*rpoC*	0.0025	DNA-directed RNA polymerase subunit beta'
CLSPO_c36510	*rpoB*	0.0045	DNA-directed RNA polymerase subunit beta

Table S4: List of candidate genes essential for the survival of *C. sporogenes* (continuation).

Gene ID	Gene name	Insertion index	Gene function
CLSPO_c36520	*rplL*	0.0000	50S ribosomal protein L7/L12
CLSPO_c36530	*rplJ*	0.0022	50S ribosomal protein L10
CLSPO_c36540	*rplA*	0.0064	50S ribosomal protein L1
CLSPO_c36550	*rplK*	0.0000	50S ribosomal protein L11
CLSPO_c36560	*nusG*	0.0000	transcription termination/antitermination protein NusG
CLSPO_c36570	*secE*	0.0000	protein translocase subunit SecE
CLSPO_c36580	*rpmG*	0.0074	50S ribosomal protein L33
CLSPO_c36620	*sigH*	0.0017	RNA polymerase sigma-H factor SigH
CLSPO_c36650	*thyX*	0.0029	thymidylate synthase ThyX
CLSPO_c36670	*cysS*	0.0095	cysteine--tRNA ligase CysS
CLSPO_c36700	*ispD*	0.0048	2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase IspD
CLSPO_c36920	*murD*	0.0073	UDP-N-acetylmuramoylalanine--D-glutamate ligase MurD
CLSPO_c36940	*glyQS*	0.0016	glycine--tRNA ligase
CLSPO_c36950	*lysS*	0.0073	lysine--tRNA ligase LysS
CLSPO_c36960	*greA2*	0.0046	transcription elongation factor GreA
CLSPO_c36990	*coaX*	0.0057	type III pantothenate kinase CoaX
CLSPO_c37030	*tilS*	0.0064	tRNA(Ile)-lysidine synthase TilS
CLSPO_c37130	*ftsL2*	0.0000	cell division protein FtsL
CLSPO_c37170	*hup*	0.0000	DNA-binding protein HU
CLSPO_c37230	*pth*	0.0097	peptidyl-tRNA hydrolase Pth
CLSPO_c37240	*htrB*	0.0076	serine protease Do-like HtrB
CLSPO_c37270	*prs*	0.0023	ribose-phosphate pyrophosphokinase Prs
CLSPO_c37280	*glmU*	0.0073	bifunctional protein GlmU
CLSPO_c37310	*murC*	0.0105	UDP-N-acetylmuramate--L-alanine ligase MurC
CLSPO_c37460	*murI*	0.0029	glutamate racemase MurI
CLSPO_c37490	*murA1*	0.0035	UDP-N-acetylglucosamine 1-carboxyvinyltransferase MurA
CLSPO_c37810	*fabZ*	0.0026	3-hydroxyacyl-[acyl-carrier-protein] dehydratase FabZ
CLSPO_c37830	*fabF*	0.0054	3-oxoacyl-[acyl-carrier-protein] synthase 2
CLSPO_c37850	*fabD*	0.0106	malonyl CoA-acyl carrier protein transacylase FabD

Table S4: List of candidate genes essential for the survival of *C. sporogenes* (continuation).

Gene ID	Gene name	Insertion index	Gene function
CLSPO_c37870	*acpP2*	0.0049	acyl carrier protein AcpP
CLSPO_c37880	*fabH2*	0.0102	3-oxoacyl-[acyl-carrier-protein] synthase 3
CLSPO_c37920	*dnaD*	0.0101	putative DNA replication protein DnaD
CLSPO_c37930	*dnaC1*	0.0045	DNA replication protein DnaC
CLSPO_c38120	CLSPO_c38120	0.0051	hypothetical protein DUF3006
CLSPO_c38140	*dnaC2*	0.0033	DNA replication protein DnaC
CLSPO_c38200	*rpsR*	0.0046	30S ribosomal protein S18
CLSPO_c38210	*ssb3*	0.0074	single-stranded DNA-binding protein 3
CLSPO_c38220	*rpsF*	0.0000	30S ribosomal protein S6
CLSPO_c38400	CLSPO_c38400	0.0051	membrane protein insertase
CLSPO_c38430	CLSPO_c38430	0.0000	hypothetical protein

List of genes essential for vegetative growth

Table S5: List of candidate genes essential for the vegetative growth of *C. sporogenes*. Insertion indexes below the essentiality cut-off are represented in bold, and values in the ambiguous zone are represented with *.

Gene ID	Gene name	Insertion index (replicate 1)	Insertion index (replicate 2)	Insertion index (replicate 3)	Gene function
CLSPO_c01760	CLSPO_c01760	**0.0068**	0.0091*	0.0045*	cytidine/deoxycytidylate deaminase family protein
CLSPO_c10400	*tlp*	**0.0097**	0.0097*	0.0049*	small acid-soluble spore protein Tlp
CLSPO_c18550	CLSPO_c18550	**0.0089**	0.0089*	0.0089	hypothetical protein
CLSPO_c18770	*xseB*	**0.0051**	0.0103*	0.0154	exodeoxyribonuclease 7 small subunit
CLSPO_c21770	*nodI2*	**0.0076**	0.0101*	0.0038*	Nod factor export ATP-binding protein I
CLSPO_c28100	*csrA*	**0.0101**	0.0101*	0.0101	carbon storage regulator
CLSPO_c29210	*hypA2*	**0.0097**	0.0097*	0.0065	hydrogenase nickel incorporation protein HypA
CLSPO_c30340	CLSPO_c30340	**0.0107**	0.0089*	0.0072	UBA/TS-N domain-containing protein
CLSPO_c30720	CLSPO_c30720	**0.0058**	0.0087*	0.0058	RutC family protein
CLSPO_c31420	CLSPO_c31420	**0.0054**	0.0109*	0.0054	small, acid-soluble spore protein, alpha/beta type
CLSPO_c06110	CLSPO_c06110	0.0632	**0.0000**	0.0105	hypothetical protein
CLSPO_c08630	CLSPO_c08630	0.0134*	**0.0079**	0.0087	GntR family transcriptional regulator
CLSPO_c11230	*birA*	0.0169	**0.0056**	0.0045*	bition-(acetyl-CoA_carboxylase) ligase BirA
CLSPO_c12010	CLSPO_c12010	0.0113*	**0.0075**	0.0094	putative hydrolase
CLSPO_c12070	CLSPO_c12070	0.0139*	**0.0069**	0.0069	hypothetical protein
CLSPO_c12080	CLSPO_c12080	0.0309	**0.0062**	0.0062	hypothetical protein
CLSPO_c12450	*cotJB*	0.0163	**0.0041**	0.0041*	spore coat peptide assembly protein CotJB
CLSPO_c12820	CLSPO_c12820	0.0121*	**0.0061**	0.0061	hypothetical protein DUF4363
CLSPO_c12890	*ytfJ*	0.0150*	**0.0075**	0.0075	sporulation protein YtfJ

Table S5: List of candidate genes essential for the vegetative growth of *C. sporogenes* (continuation).

Gene ID	Gene name	Insertion index (replicate 1)	Insertion index (replicate 2)	Insertion index (replicate 3)	Gene function
CLSPO_c13180	CLSPO_c13180	0.0109*	**0.0000**	0.0109	hypothetical protein
CLSPO_c13510	CLSPO_c13510	0.0156	**0.0078**	0.0078	putative transcriptional regulator
CLSPO_c14610	CLSPO_c14610	0.0191	**0.0064**	0.0064	N-acetylmuramoyl-L-alanine amidase
CLSPO_c16720	CLSPO_c16720	0.0284	**0.0000**	0.0071	hypothetical protein
CLSPO_c17180	CLSPO_c17180	0.0215	**0.0072**	0.0108	transcriptional regulator PadR family protein
CLSPO_c17600	*safA*	0.0209	**0.0070**	0.0070	spore coat assembly protein SafA
CLSPO_c17880	*miaA*	0.0166	**0.0059**	0.0047*	tRNA dimethylallyltransferase MiaA
CLSPO_c18200	CLSPO_c18200	0.0133*	**0.0044**	0.0044*	ubiquinone oxidoreductase 24 kD subunit
CLSPO_c18340	CLSPO_c18340	0.0137*	**0.0068**	0.0068	hypothetical protein
CLSPO_c19900	*nhaP*	0.0179	**0.0078**	0.0062	K(+)/H(+) antiporter NhaP2
CLSPO_c19950	*treA*	0.0172	**0.0073**	0.0046*	trehalose-6-phosphate hydrolase TreA
CLSPO_c20480	CLSPO_c20480	0.0114*	**0.0071**	0.0071	PhzF family phenazine biosynthesis protein
CLSPO_c23290	CLSPO_c23290	0.0133*	**0.0044**	0.0089	hypothetical protein
CLSPO_c23730	CLSPO_c23730	0.0190	**0.0071**	0.0095	xanthine dehydrogenase
CLSPO_c23900	CLSPO_c23900	0.0110*	**0.0055**	0.0110	hypothetical protein
CLSPO_c24190	CLSPO_c24190	0.0236	**0.0079**	0.0118	DNA-binding protein
CLSPO_c24360	CLSPO_c24360	0.0150*	**0.0075**	0.0150	hypothetical protein
CLSPO_c25500	*smc*	0.0124*	**0.0074**	0.0047*	chromosome partition protein Smc
CLSPO_c25890	CLSPO_c25890	0.0139*	**0.0060**	0.0080	putative methyltransferase
CLSPO_c26100	CLSPO_c26100	0.0189	**0.0076**	0.0051*	TIGR00255 family protein
CLSPO_c26310	*ftsA2*	0.0186	**0.0080**	0.0080	cell division protein FtsA
CLSPO_c26430	*rnjA2*	0.0130*	**0.0045**	0.0065	ribonuclease J 1
CLSPO_c26820	CLSPO_c26820	0.0175	**0.0058**	0.0117	hypothetical protein
CLSPO_c26900	*glpX*	0.0171	**0.0080**	0.0068	fructose-1,6-bisphosphatase class 2
CLSPO_c26910	CLSPO_c26910	0.0127*	**0.0064**	0.0064	Spo0E like sporulation regulatory protein

Table S5: List of candidate genes essential for the vegetative growth of *C. sporogenes* (continuation).

Gene ID	Gene name	Insertion index (replicate 1)	Insertion index (replicate 2)	Insertion index (replicate 3)	Gene function
CLSPO_c30670	*selA*	0.0192	**0.0080**	0.0080	L-seryl-tRNA(Sec) selenium transferase SelA
CLSPO_c31580	CLSPO_c31580	0.0138*	**0.0065**	0.0041*	radical SAM domain-containing protein
CLSPO_c33490	CLSPO_c33490	0.0233	**0.0067**	0.0067	DsrE/DsrF-like family
CLSPO_c33850	CLSPO_c33850	0.0142*	**0.0071**	0.0071	hypothetical protein
CLSPO_c35410	CLSPO_c35410	0.0203	**0.0076**	0.0051*	FHA domain-containing protein
CLSPO_c37820	CLSPO_c37820	0.0116*	**0.0000**	0.0070	acetyl-CoA carboxylase
CLSPO_c38250	CLSPO_c38250	0.0112*	**0.0056**	0.0112	hypothetical protein
CLSPO_c38420	*rnpA*	0.0132*	**0.0066**	0.0066	ribonuclease P protein component RnpA
CLSPO_c03090	CLSPO_c03090	0.0279	0.0279	**0.0000**	cold shock protein
CLSPO_c04120	CLSPO_c04120	0.0115*	0.0115	**0.0000**	hypothetical protein
CLSPO_c05400	*fabG1*	0.0269	0.0149	**0.0030**	3-oxoacyl-[acyl-carrier-protein] reductase FabG
CLSPO_c05910	CLSPO_c05910	0.0262	0.0117	**0.0029**	GntR family transcriptional regulator
CLSPO_c07790	*arsR*	0.0240	0.0103*	**0.0034**	arsenical resistance operon repressor
CLSPO_c07920	CLSPO_c07920	0.0286	0.0143	**0.0024**	putative transporter
CLSPO_c09430	CLSPO_c09430	0.0289	0.0129	**0.0032**	ribosomal protein L7/L12
CLSPO_c09530	*btuR*	0.0234	0.0085*	**0.0021**	Cob(I)yrinic acid a
CLSPO_c09880	CLSPO_c09880	0.0174	0.0209	**0.0000**	PadR family transcriptional regulator
CLSPO_c10430	CLSPO_c10430	0.0164	0.0246	**0.0000**	hypothetical protein
CLSPO_c12230	CLSPO_c12230	0.0333	0.0182	**0.0030**	protease family protein
CLSPO_c12830	*ytrH*	0.0196	0.0163	**0.0033**	sporulation protein YtrH
CLSPO_c13210	CLSPO_c13210	0.0290	0.0290	**0.0000**	hypothetical protein
CLSPO_c14440	*patM2*	0.0283	0.0133	**0.0017**	putative amino-acid ABC transporter permease protein PatM

Table S5: List of candidate genes essential for the vegetative growth of *C. sporogenes* (continuation).

Gene ID	Gene name	Insertion index (replicate 1)	Insertion index (replicate 2)	Insertion index (replicate 3)	Gene function
CLSPO_c14850	*rsmH*	0.0227	0.0131	**0.0036**	ribosomal RNA small subunit methyltransferase H
CLSPO_c14940	CLSPO_c14940	0.0158	0.0127	**0.0032**	hypothetical protein
CLSPO_c14950	CLSPO_c14950	0.0228	0.0152	**0.0030**	hypothetical protein
CLSPO_c15370	*proX*	0.0271	0.0226	**0.0023**	prolyl-tRNA editing protein ProX
CLSPO_c15680	CLSPO_c15680	0.0317	0.0159	**0.0000**	hypothetical protein DUF1657
CLSPO_c16870	CLSPO_c16870	0.0229	0.0115	**0.0029**	putative aspartyl protease
CLSPO_c17220	*adh1*	0.0133*	0.0114	**0.0029**	NADPH-dependent butanol dehydrogenase Adh
CLSPO_c17800	CLSPO_c17800	0.0109*	0.0109*	**0.0000**	small acid-soluble spore protein alpha
CLSPO_c17860	CLSPO_c17860	0.0190	0.0087*	**0.0026**	aluminum resistance protein
CLSPO_c18020	*gluD*	0.0140*	0.0105*	**0.0035**	NAD-specific glutamate dehydrogenase GluD
CLSPO_c18270	CLSPO_c18270	0.0298	0.0128	**0.0000**	hydrogenase expression/synthesis hypA family
CLSPO_c18370	CLSPO_c18370	0.0121*	0.0145	**0.0000**	transcription activator effector binding protein
CLSPO_c19300	CLSPO_c19300	0.0200	0.0125	**0.0025**	MarR family transcriptional regulator
CLSPO_c19590	CLSPO_c19590	0.0300	0.0178	**0.0033**	leucine rich repeat protein
CLSPO_c19850	CLSPO_c19850	0.0157	0.0118	**0.0000**	hypothetical protein
CLSPO_c19970	*fruA2*	0.0252	0.0202	**0.0025**	PTS system fructose-specific EIIABC component FruA
CLSPO_c20190	CLSPO_c20190	0.0225	0.0090*	**0.0015**	SCP-like extracellular protein
CLSPO_c20430	CLSPO_c20430	0.0165	0.0107*	**0.0033**	amino acid/peptide transporter
CLSPO_c20680	CLSPO_c20680	0.0137*	0.0126	**0.0032**	sorbitol dehydrogenase
CLSPO_c21240	*eutN*	0.0168	0.0210	**0.0000**	ethanolamine utilization protein EutN
CLSPO_c21260	*eutJ*	0.0157	0.0131	**0.0026**	ethanolamine utilization protein EutJ
CLSPO_c21270	CLSPO_c21270	0.0137*	0.0107*	**0.0030**	hypothetical protein

Table S5: List of candidate genes essential for the vegetative growth of *C. sporogenes* (continuation).

Gene ID	Gene name	Insertion index (replicate 1)	Insertion index (replicate 2)	Insertion index (replicate 3)	Gene function
CLSPO_c21650	CLSPO_c21650	0.0210	0.0105*	**0.0028**	putative AAA-ATPase
CLSPO_c21660	CLSPO_c21660	0.0144*	0.0103*	**0.0031**	dehydrogenase
CLSPO_c22180	CLSPO_c22180	0.0258	0.0155	**0.0026**	hypothetical protein DUF2935
CLSPO_c22640	*hypB1*	0.0295	0.0243	**0.0017**	hydrogenase nickel incorporation protein HypB
CLSPO_c23140	CLSPO_c23140	0.0174	0.0130	**0.0000**	TIGR03987 family protein
CLSPO_c23500	CLSPO_c23500	0.0200	0.0100*	**0.0033**	stress responsive A/B barrel domain-containing protein
CLSPO_c24050	CLSPO_c24050	0.0377	0.0094*	**0.0000**	hypothetical protein
CLSPO_c24060	CLSPO_c24060	0.0298	0.0119	**0.0000**	hypothetical protein
CLSPO_c24070	CLSPO_c24070	0.0198	0.0123	**0.0025**	hypothetical protein
CLSPO_c24120	CLSPO_c24120	0.0230	0.0115	**0.0000**	hypothetical protein
CLSPO_c24140	CLSPO_c24140	0.0180	0.0126	**0.0036**	methyltransferase family protein
CLSPO_c24290	CLSPO_c24290	0.0121*	0.0121	**0.0000**	hypothetical protein
CLSPO_c24540	CLSPO_c24540	0.0251	0.0219	**0.0031**	hypothetical protein
CLSPO_c25170	*rbfA*	0.0214	0.0092*	**0.0031**	ribosome-binding factor A
CLSPO_c26420	*typA*	0.0195	0.0116	**0.0012**	GTP-binding protein TypA/BipA
CLSPO_c26700	CLSPO_c26700	0.0123*	0.0165	**0.0000**	hypothetical protein
CLSPO_c27020	*atpF2*	0.0139*	0.0105*	**0.0035**	V-type ATP synthase subunit F
CLSPO_c28790	CLSPO_c28790	0.0212	0.0189	**0.0024**	hypothetical protein
CLSPO_c30680	*selD*	0.0165	0.0144	**0.0021**	selenide, water dikinase SelD
CLSPO_c31320	CLSPO_c31320	0.0156	0.0156	**0.0000**	hypothetical protein
CLSPO_c31550	*recJ2*	0.0152*	0.0139	**0.0025**	single-stranded-DNA-specific exonuclease RecJ
CLSPO_c35860	CLSPO_c35860	0.0159	0.0159	**0.0000**	2-oxoacid:acceptor oxidoreductase subunit delta

Table S5: List of candidate genes essential for the vegetative growth of *C. sporogenes* (continuation).

Gene ID	Gene name	Insertion index (replicate 1)	Insertion index (replicate 2)	Insertion index (replicate 3)	Gene function
CLSPO_c37210	*prsA*	0.0187	0.0110*	**0.0011**	foldase protein PrsA
CLSPO_c01100	CLSPO_c01100	0.0160	**0.0053**	**0.0027**	putative lipoprotein
CLSPO_c01910	CLSPO_c01910	**0.0104**	0.0089*	**0.0030**	putative peptidase
CLSPO_c07780	CLSPO_c07780	**0.0106**	**0.0080**	0.0053	arsenate reductase family protein
CLSPO_c08710	CLSPO_c08710	**0.0079**	**0.0063**	0.0063	L-aminopeptidase/D-esterase
CLSPO_c10240	CLSPO_c10240	0.0144*	**0.0018**	**0.0018**	TetR family transcriptional regulator
CLSPO_c14140	CLSPO_c14140	0.0213	**0.0071**	**0.0000**	hypothetical protein
CLSPO_c14800	*grdI*	0.0109*	**0.0075**	**0.0033**	betaine reductase complex component B subunit alpha
CLSPO_c14930	CLSPO_c14930	0.0171	**0.0078**	**0.0016**	hypothetical protein
CLSPO_c18110	CLSPO_c18110	0.0180	**0.0077**	**0.0026**	thioredoxin family protein
CLSPO_c20490	CLSPO_c20490	0.0213	**0.0047**	**0.0024**	endoribonuclease family protein
CLSPO_c23490	CLSPO_c23490	**0.0077**	**0.0077**	0.0077	putative transcriptional regulator
CLSPO_c24460	CLSPO_c24460	**0.0087**	0.0087*	**0.0000**	phage protein
CLSPO_c30840	CLSPO_c30840	**0.0000**	0.0119	**0.0000**	hypothetical protein
CLSPO_c21230	*eutQ*	0.0134*	**0.0067**	**0.0017**	ethanolamine utilization protein EutQ
CLSPO_c21420	CLSPO_c21420	0.0183	**0.0073**	**0.0000**	hypothetical protein DUF1963
CLSPO_c22690	CLSPO_c22690	0.0116*	**0.0058**	**0.0000**	hypothetical protein
CLSPO_c24700	CLSPO_c24700	0.0132*	**0.0000**	**0.0000**	hypothetical protein
CLSPO_c24760	CLSPO_c24760	0.0131*	**0.0052**	**0.0000**	hypothetical protein DUF1064
CLSPO_c25220	*rimP*	0.0192	**0.0072**	**0.0000**	ribosome maturation factor RimP
CLSPO_c25540	*acpP1*	0.0142*	**0.0000**	**0.0000**	acyl carrier protein AcpP
CLSPO_c25810	CLSPO_c25810	0.0208	**0.0069**	**0.0000**	transposase
CLSPO_c25970	*rsgA*	0.0125*	**0.0075**	**0.0012**	putative ribosome biogenesis GTPase RsgA

Table S5: List of candidate genes essential for the vegetative growth of *C. sporogenes* (continuation).

Gene ID	Gene name	Insertion index (replicate 1)	Insertion index (replicate 2)	Insertion index (replicate 3)	Gene function
CLSPO_c26330	CLSPO_c26330	0.0246	**0.0000**	**0.0000**	hypothetical protein
CLSPO_c26950	CLSPO_c26950	0.0238	**0.0000**	**0.0000**	hypothetical protein
CLSPO_c31990	CLSPO_c31990	0.0199	**0.0021**	**0.0031**	UDP-N-acetylglucosamine 2-epimerase
CLSPO_c34150	*guaA*	0.0167	**0.0051**	**0.0014**	GMP synthase
CLSPO_c34260	*rex*	0.0246	**0.0070**	**0.0018**	redox-sensing transcriptional repressor Rex
CLSPO_c35390	*pbpA2*	0.0124*	**0.0062**	**0.0016**	penicillin-binding protein A
CLSPO_c35680	CLSPO_c35680	0.0138*	**0.0069**	**0.0023**	coat F domain-containing protein
CLSPO_c35890	*ptb2*	0.0134*	**0.0049**	**0.0024**	phosphate butyryltransferase Ptb
CLSPO_c36100	*ecfT*	0.0318	**0.0069**	**0.0014**	energy-coupling factor transporter transmembrane protein EcfT
CLSPO_c36590	*tuf1*	0.0130*	**0.0037**	**0.0019**	elongation factor Tu
CLSPO_c37860	CLSPO_c37860	0.0120*	**0.0072**	**0.0036**	nitronate monooxygenase
CLSPO_c00050	CLSPO_c00050	**0.0084**	**0.0000**	**0.0000**	hypothetical protein
CLSPO_c01600	*rpmE*	**0.0051**	**0.0000**	**0.0000**	50S ribosomal protein L31
CLSPO_c02070	*prfB*	**0.0090**	**0.0045**	**0.0000**	peptide chain release factor 2
CLSPO_c19870	CLSPO_c19870	**0.0100**	**0.0067**	**0.0033**	4-carboxymuconolactone decarboxylase
CLSPO_c21130	CLSPO_c21130	**0.0083**	**0.0033**	**0.0000**	hypothetical protein
CLSPO_c21490	CLSPO_c21490	**0.0080**	**0.0000**	**0.0000**	acetyltransferase
CLSPO_c24970	*hpr*	**0.0086**	**0.0043**	**0.0000**	phosphocarrier protein HPr
CLSPO_c25630	*prdA2*	**0.0084**	**0.0037**	**0.0016**	D-proline reductase proprotein PrdA
CLSPO_c30590	*holA*	**0.0108**	**0.0022**	**0.0022**	DNA polymerase III
CLSPO_c30920	*mreD*	**0.0023**	**0.0023**	**0.0023**	Rod shape-determining protein MreD

Table S5: List of candidate genes essential for the vegetative growth of *C. sporogenes* (continuation).

Gene ID	Gene name	Insertion index (replicate 1)	Insertion index (replicate 2)	Insertion index (replicate 3)	Gene function
CLSPO_c33930	*pcrA2*	**0.0075**	**0.0050**	**0.0035**	ATP-dependent DNA helicase PcrA
CLSPO_c36250	*rpmD*	**0.0000**	**0.0000**	**0.0000**	50S ribosomal protein L30
CLSPO_c37780	*accA*	**0.0026**	**0.0026**	**0.0000**	acetyl-coenzyme A carboxylase carboxyl transferase subunit alpha
CLSPO_c37790	*accD*	**0.0089**	**0.0077**	**0.0026**	acetyl-coenzyme A carboxylase carboxyl transferase subunit beta
CLSPO_c37800	*accC*	**0.0074**	**0.0008**	**0.0000**	biotin carboxylase AccC
CLSPO_c37840	*fabG2*	**0.0074**	**0.0074**	**0.0000**	3-oxoacyl-[acyl-carrier-protein] reductase FabG
CLSPO_c37900	CLSPO_c37900	**0.0084**	**0.0063**	**0.0031**	2-nitropropane dioxygenase family oxidoreductase

List of genes essential for sporulation

Table S6: List of candidate genes essential for sporulation in *C. sporogenes*. Insertion indexes below the essentiality cut-off are represented in bold, and values in the ambiguous zone are represented with *. The log_2FC of mutant frequencies in the sporulation library relative to the vegetative cells library is also presented, as well as the respective q-value (bold numbers represent significant differences: log_2FC < -2 and q-value < 0.05).

		Insertion index					
Locus tag	**Gene name**	**Replicate 1**	**Replicate 2**	**Replicate 3**	**Log_2FC**	**q-value**	**Function**
CLSPO_c00200	*serS1*	**0.0000**	**0.0017**	**0.0009**	**-5.7620**	**1.15E-28**	serine--tRNA ligase SerS
CLSPO_c00800	CLSPO_c00800	0.0052	0.0104	0.0104	**-2.5426**	**2.08E-06**	inhibitor of sigma-G Gin
CLSPO_c00810	CLSPO_c00810	**0.0008**	0.0130	0.0122	-1.1773	**7.01E-03**	Orn/Lys/Arg decarboxylase
CLSPO_c00850	CLSPO_c00850	**0.0012**	0.0073*	**0.0049**	**-2.2512**	**5.36E-09**	PSP1 domain-containing protein
CLSPO_c01080	CLSPO_c01080	0.0046*	0.0165	0.0119	**-3.3712**	**2.00E-49**	hypothetical protein
CLSPO_c01090	CLSPO_c01090	0.0038*	0.0189	0.0170	**-3.6345**	**2.78E-50**	hypothetical protein
CLSPO_c01280	CLSPO_c01280	0.0125	0.0229	0.0249	**-2.4840**	**7.12E-37**	mannose-1-phosphate guanylyltransferase ManC
CLSPO_c01350	CLSPO_c01350	**0.0022**	0.0165	0.0088	**-4.5481**	**3.90E-72**	CotS family spore coat protein
CLSPO_c01360	CLSPO_c01360	0.0057	0.0114	**0.0000**	**-3.5099**	**2.75E-06**	hypothetical protein
CLSPO_c01400	CLSPO_c01400	**0.0011**	**0.0044**	0.0066*	**-6.6086**	**1.42E-127**	putative spore coat protein
CLSPO_c01430	CLSPO_c01430	**0.0028**	0.0071*	0.0064*	-1.1076	**1.79E-07**	LysM domain-containing protein
CLSPO_c01510	*spoIIR*	0.0140	0.0281	0.0316	**-2.8498**	**6.64E-41**	stage II sporulation protein R
CLSPO_c01540	CLSPO_c01540	**0.0000**	**0.0023**	**0.0000**	**-5.7428**	**2.60E-21**	putative tRNA(5-methylaminomethyl-2-thiouridylate) methyltransferase
CLSPO_c01650	CLSPO_c01650	0.0182	0.0364	0.0364	**-2.3538**	**1.01E-31**	hypothetical protein
CLSPO_c01880	CLSPO_c01880	**0.0031**	0.0219	0.0172	**-2.9983**	**6.56E-20**	hypothetical protein DUF1779
CLSPO_c01890	*murA2*	0.0035*	0.0115	0.0106	**-3.9942**	**2.69E-79**	UDP-N-acetylglucosamine 1-carboxyvinyltransferase MurA
CLSPO_c01900	*spoIID*	**0.0021**	**0.0062**	**0.0052**	**-3.7163**	**1.42E-61**	stage II sporulation protein D

Table S6: List of candidate genes essential for sporulation in *C. sporogenes* (continuation).

Locus tag	Gene name	Insertion index			Log_2FC	q-value	Function
		Replicate 1	Replicate 2	Replicate 3			
CLSPO_c01920	*spoIIID*	0.0130	0.0174	0.0217	**-2.7790**	**2.23E-23**	stage III sporulation protein D
CLSPO_c02310	*pssA*	**0.0000**	**0.0064**	**0.0043**	**-3.0257**	**2.68E-07**	CDP-diacylglycerol--serine O-phosphatidyltransferase PssA
CLSPO_c03660	*resE7*	0.0036*	0.0276	0.0222	**-2.2745**	**1.05E-11**	sensor histidine kinase ResE
CLSPO_c04390	CLSPO_c04390	**0.0000**	**0.0000**	**0.0000**	**-10.0248**	**2.93E-33**	tetratricopeptide repeat protein
CLSPO_c04460	*uppP1*	0.0091	0.0169	0.0104	**-3.9957**	**1.32E-37**	undecaprenyl-diphosphatase UppP
CLSPO_c05110	*ziaR*	**0.0000**	0.0091*	0.0121	**-3.8865**	**1.36E-05**	transcriptional repressor SmtB
CLSPO_c05120	*cadA*	**0.0000**	**0.0010**	**0.0010**	**-7.1997**	**5.42E-49**	cadmium, zinc and cobalt-transporting ATPase CadA
CLSPO_c05340	CLSPO_c05340	0.0091	0.0205	0.0160	**-2.3151**	**1.10E-07**	hypothetical protein
CLSPO_c05400	*fabG1*	**0.0015**	0.0090*	**0.0045**	**-2.7699**	**3.23E-06**	3-oxoacyl-[acyl-carrier-protein] reductase FabG
CLSPO_c05620	*asnC*	0.0072	0.0119	0.0143	**-2.0101**	**2.25E-08**	regulatory protein AsnC
CLSPO_c05690	*chrA1*	0.0083	0.0393	0.0351	**-2.1448**	**9.87E-08**	chromate transport protein ChrA
CLSPO_c05700	*chrA2*	**0.0021**	0.0248	0.0165	-1.7585	**1.16E-05**	chromate transport protein ChrA
CLSPO_c08250	CLSPO_c08250	0.0061	**0.0061**	0.0121	0.3142	7.65E-01	hypothetical protein
CLSPO_c09460	*cbiG*	**0.0031**	0.0093	0.0108	-0.4445	3.48E-01	cobalt-precorrin 5A hydrolase CbiG
CLSPO_c09480	*cbiJ*	**0.0014**	**0.0043**	0.0072*	**-3.4152**	**3.95E-22**	putative cobalt-precorrin-6A reductase CbiJ
CLSPO_c09530	*btuR*	**0.0000**	**0.0000**	**0.0000**	**-6.2064**	**1.32E-04**	Cob(I)yrinic acid a,c-diamide adenosyltransferase BtuR
CLSPO_c09720	CLSPO_c09720	**0.0000**	**0.0008**	**0.0008**	**-5.5375**	**9.44E-19**	sensor histidine kinase
CLSPO_c12640	*grdX*	**0.0000**	**0.0029**	**0.0000**	**-5.9349**	**6.53E-09**	glycine reducatse operon associated protein GrdX

Table S6: List of candidate genes essential for sporulation in *C. sporogenes* (continuation).

Locus tag	Gene name	Insertion index			Log_2FC	q-value	Function
		Replicate 1	Replicate 2	Replicate 3			
CLSPO_c12650	*grdE*	0.0035*	0.0129	0.0086	**-3.4319**	**1.08E-44**	glycine reductase complex component B subunits alpha and beta
CLSPO_c12660	*grdB*	**0.0017**	0.0085*	**0.0034**	**-2.9398**	**5.53E-14**	glycine reductase complex component B subunit gamma
CLSPO_c12670	*trxB3*	**0.0000**	**0.0012**	**0.0000**	**-7.0769**	**1.70E-31**	thioredoxin reductase TrxB
CLSPO_c12680	*trxA1*	**0.0000**	0.0069*	**0.0000**	**-4.8988**	**1.23E-13**	thioredoxin TrxA
CLSPO_c12690	*grdA*	**0.0000**	**0.0023**	**0.0000**	**-5.9402**	**4.06E-12**	glycine/sarcosine/betaine reductase complex component A
CLSPO_c12700	*grdC*	**0.0000**	**0.0014**	**0.0000**	**-7.5550**	**4.34E-46**	glycine/sarcosine/betaine reductase complex component C subunit beta
CLSPO_c12710	*grdD*	**0.0000**	**0.0010**	**0.0010**	**-7.2303**	**2.51E-34**	glycine/sarcosine/betaine reductase complex component C subunit alpha
CLSPO_c13700	CLSPO_c13700	0.0157	0.0262	0.0157	**-3.0892**	**8.68E-17**	RpiR family transcriptional regulator
CLSPO_c14830	*ychF*	**0.0010**	0.0111	**0.0061**	-1.5002	**2.64E-04**	GTP-binding protein YchF
CLSPO_c14850	*rsmH*	**0.0012**	**0.0036**	**0.0060**	**-2.7327**	**1.53E-09**	ribosomal RNA small subunit methyltransferase H
CLSPO_c15060	*lspA*	**0.0025**	0.0125	0.0125	**-3.3386**	**6.64E-15**	lipoprotein signal peptidase LspA
CLSPO_c15070	CLSPO_c15070	**0.0012**	**0.0024**	**0.0012**	-1.1865	**4.91E-02**	pseudouridine synthase
CLSPO_c16610	*ytaF*	**0.0018**	0.0071*	0.0142	-1.7632	**2.76E-08**	putative sporulation protein YtaF
CLSPO_c16620	CLSPO_c16620	0.0132	0.0263	0.0197	**-2.2550**	**1.69E-08**	hypothetical protein
CLSPO_c17010	CLSPO_c17010	**0.0032**	0.0095	0.0074*	-0.2845	5.15E-01	putative type I restriction enzyme MjaXP M protein

Table S6: List of candidate genes essential for sporulation in *C. sporogenes* (continuation).

Locus tag	Gene name	Insertion index			Log_2FC	q-value	Function
		Replicate 1	Replicate 2	Replicate 3			
CLSPO_c17020	*prrB*	0.0040*	**0.0064**	0.0064*	0.2479	8.38E-01	type-1 restriction enzyme ecoprrI specificity protein PrrB
CLSPO_c17180	CLSPO_c17180	**0.0000**	0.0179	0.0072*	-1.6389	**2.12E-03**	transcriptional regulator PadR family protein
CLSPO_c17560	CLSPO_c17560	**0.0023**	0.0091*	0.0080*	**-5.1096**	**2.93E-40**	hypothetical protein
CLSPO_c17860	CLSPO_c17860	**0.0009**	0.0069*	**0.0035**	**-2.2996**	**1.39E-06**	aluminum resistance protein
CLSPO_c17880	*miaA*	**0.0000**	**0.0047**	**0.0024**	**-3.0192**	**6.45E-07**	tRNA dimethylallyltransferase MiaA
CLSPO_c17920	*lysN*	**0.0000**	**0.0048**	**0.0008**	**-5.3777**	**2.84E-29**	2-aminoadipate transaminase LysN
CLSPO_c17940	*aspB1*	**0.0028**	0.0103	0.0065*	-1.3197	**1.40E-04**	aspartate aminotransferase AspB
CLSPO_c17990	*cmk*	**0.0000**	**0.0068**	**0.0034**	**-2.4952**	**3.70E-05**	cytidylate kinase Cmk
CLSPO_c18220	CLSPO_c18220	**0.0009**	**0.0019**	**0.0019**	**-6.6114**	**1.71E-56**	sporulation domain-containing protein
CLSPO_c18510	*scpA*	**0.0030**	0.0342	0.0253	-0.3463	4.71E-01	segregation and condensation protein A
CLSPO_c18600	*spoIIM*	0.0051	**0.0051**	0.0068*	**-2.3250**	**2.53E-11**	stage II sporulation protein M
CLSPO_c18650	*spo0A*	**0.0000**	**0.0000**	**0.0000**	**-6.1293**	**1.14E-04**	stage 0 sporulation protein A
CLSPO_c18660	*spoIVB*	0.0049	0.0113	0.0065*	**-2.0244**	**1.39E-11**	SpoIVB peptidase
CLSPO_c18710	*argR*	**0.0000**	0.0074*	**0.0049**	**-4.2459**	**2.86E-09**	arginine repressor
CLSPO_c18730	CLSPO_c18730	**0.0014**	0.0095	**0.0041**	**-2.2928**	**6.14E-06**	putative rRNA methylase
CLSPO_c18810	*spoIIIAH*	**0.0021**	**0.0042**	**0.0021**	**-5.4140**	**5.34E-28**	stage III sporulation protein AH
CLSPO_c18820	*spoIIIAG*	**0.0019**	**0.0056**	**0.0019**	0.4758	2.54E-01	stage III sporulation protein AG
CLSPO_c18830	*spoIIIAF*	**0.0019**	**0.0019**	**0.0038**	**-5.0500**	**1.43E-22**	stage III sporulation protein AF
CLSPO_c18840	*spoIIIAE*	**0.0019**	**0.0067**	0.0086	-1.8891	**2.25E-13**	stage III sporulation protein AE
CLSPO_c18850	*spoIIIAD*	**0.0029**	**0.0058**	**0.0029**	0.1477	8.73E-01	stage III sporulation protein AD

Table S6: List of candidate genes essential for sporulation in *C. sporogenes* (continuation).

Locus tag	Gene name	Insertion index			Log_2FC	q-value	Function
		Replicate 1	Replicate 2	Replicate 3			
CLSPO_c18860	*spoIIIAC*	**0.0000**	0.0168	0.0112	**-3.4440**	**1.41E-11**	stage III sporulation protein AC
CLSPO_c18880	*spoIIIAA*	0.0036*	**0.0048**	0.0072*	-1.8858	**1.09E-08**	stage III sporulation protein AA
CLSPO_c18910	*aroQ*	0.0050	0.0250	0.0125	**-2.1023**	**2.11E-07**	3-dehydroquinate dehydratase AroQ
CLSPO_c19050	CLSPO_c19050	**0.0000**	**0.0032**	**0.0045**	**-4.3812**	**1.84E-25**	fibronectin-binding protein
CLSPO_c19300	CLSPO_c19300	**0.0025**	0.0100	**0.0025**	**-3.1966**	**9.88E-22**	MarR family transcriptional regulator
CLSPO_c19960	*treR*	**0.0032**	0.0239	0.0159	-1.5674	**1.80E-04**	trehalose operon transcriptional repressor
CLSPO_c20420	CLSPO_c20420	**0.0000**	0.0147	0.0082*	**-2.2059**	**1.68E-08**	proton-dependent oligopeptide transporter
CLSPO_c21250	CLSPO_c21250	**0.0000**	0.0147	0.0073*	**-2.2677**	**1.04E-04**	BMC domain-containing protein
CLSPO_c22240	*lcdC*	**0.0029**	**0.0029**	**0.0043**	**-6.0467**	**4.14E-41**	activator of lactoyl-CoA dehydratase LcdC
CLSPO_c22260	*etfA1*	**0.0032**	0.0085*	0.0063*	**-3.2766**	**2.27E-13**	electron transfer flavoprotein subunit alpha
CLSPO_c22270	*etfB1*	**0.0028**	0.0113	0.0071*	**-3.3013**	**5.88E-18**	electron transfer flavoprotein subunit beta
CLSPO_c22290	CLSPO_c22290	**0.0020**	**0.0020**	**0.0039**	**-6.0458**	**2.00E-05**	benzoyl-CoA reductase/2-hydroxyglutaryl-CoA dehydratase subunit
CLSPO_c22300	CLSPO_c22300	**0.0000**	**0.0009**	**0.0018**	**-7.5193**	**4.00E-38**	benzoyl-CoA reductase/2-hydroxyglutaryl-CoA dehydratase subunit
CLSPO_c22310	CLSPO_c22310	**0.0027**	**0.0064**	**0.0055**	-1.8620	**7.45E-08**	putative acyl-CoA transferases/carnitine dehydratase

Table S6: List of candidate genes essential for sporulation in *C. sporogenes* (continuation).

Locus tag	Gene name	Insertion index			Log_2FC	q-value	Function
		Replicate 1	Replicate 2	Replicate 3			
CLSPO_c22510	CLSPO_c22510	**0.0032**	0.0120	0.0080*	-1.1460	**5.62E-04**	ErfK/YbiS/YcfS/YnhG family protein
CLSPO_c23270	CLSPO_c23270	0.0037*	0.0118	0.0087	**-2.9551**	**2.69E-21**	M24 family metallopeptidase
CLSPO_c24290	CLSPO_c24290	**0.0000**	**0.0061**	**0.0000**	**-3.4553**	**3.49E-02**	hypothetical protein
CLSPO_c24710	CLSPO_c24710	**0.0000**	0.0088*	**0.0000**	**-2.9165**	1.18E-01	hypothetical protein
CLSPO_c24960	CLSPO_c24960	**0.0000**	0.0150	**0.0050**	**-3.6133**	**7.75E-08**	hypothetical protein
CLSPO_c25000	CLSPO_c25000	0.0035*	0.0234	0.0121	**-3.8771**	**2.98E-39**	4-amino-4-deoxy-L-arabinose transferase
CLSPO_c25010	CLSPO_c25010	0.0051	0.0205	0.0154	**-2.8656**	**5.89E-13**	decaprenyl-phosphate phosphoribosyltransferase
CLSPO_c25050	*pgsA*	**0.0000**	**0.0019**	**0.0000**	**-5.3831**	**2.49E-14**	CDP-diacylglycerol--glycerol-3-phosphate 3-phosphatidyltransferase PgsA
CLSPO_c25070	*ftsK*	**0.0010**	**0.0029**	**0.0034**	**-3.1555**	**2.25E-15**	DNA translocase FtsK
CLSPO_c25100	CLSPO_c25100	**0.0000**	**0.0000**	**0.0000**	**-6.5939**	**3.46E-06**	PRC-barrel domain-containing protein
CLSPO_c25150	*truB*	0.0064	0.0140	0.0128	**-2.2999**	**2.95E-10**	tRNA pseudouridine synthase B
CLSPO_c25250	CLSPO_c25250	**0.0022**	**0.0066**	0.0066*	**-3.3975**	**1.73E-17**	putative zinc metalloprotease
CLSPO_c25700	*prdF*	0.0033*	0.0088*	0.0099	**-2.6647**	**1.40E-13**	D-proline racemase PrdF
CLSPO_c25710	*prdE1*	**0.0024**	0.0142	0.0071*	-1.2403	**4.76E-04**	D-proline reductase operon protein PrdE
CLSPO_c25740	*prdH2*	**0.0000**	0.0161	**0.0000**	-1.9995	**4.91E-02**	D-proline reductase cluster associated protein
CLSPO_c25750	*prdA3*	0.0053	0.0138	0.0101	**-2.1264**	**1.12E-13**	D-proline reductase proprotein PrdA
CLSPO_c25840	*prdR1*	**0.0000**	**0.0006**	**0.0019**	**-6.3492**	**8.67E-34**	D-proline reductase associated sigma 54 transcriptional activator
CLSPO_c25990	*prpC*	**0.0000**	**0.0015**	**0.0015**	**-7.5459**	**7.46E-38**	protein phosphatase PrpC

Table S6: List of candidate genes essential for sporulation in *C. sporogenes* (continuation).

Locus tag	Gene name	Insertion index			Log_2FC	q-value	Function
		Replicate 1	Replicate 2	Replicate 3			
CLSPO_c26120	*spoIVA*	**0.0015**	**0.0022**	**0.0030**	**-4.1485**	**1.86E-33**	stage IV sporulation protein A
CLSPO_c26260	CLSPO_c26260	0.0042*	0.0168	0.0084	**-2.2250**	**2.73E-10**	PRC-barrel domain-containing protein
CLSPO_c26270	*sigG*	**0.0029**	0.0086*	0.0129	**-2.8378**	**4.20E-16**	RNA polymerase sigma-G factor SigG
CLSPO_c26280	*sigE*	**0.0031**	0.0125	0.0110	**-2.8254**	**3.74E-16**	RNA polymerase sigma-E factor SigE
CLSPO_c26290	*spoIIGA*	0.0070	0.0209	0.0251	**-3.7598**	**2.40E-39**	putative sporulation sigma-E factor-processing peptidase
CLSPO_c26310	*ftsA2*	**0.0009**	**0.0035**	**0.0027**	**-4.6075**	**1.35E-28**	cell division protein FtsA
CLSPO_c26340	*pilT*	**0.0011**	**0.0063**	**0.0032**	**-2.2447**	**1.20E-04**	twitching mobility protein PilT
CLSPO_c26360	*sigK*	**0.0031**	**0.0047**	**0.0047**	**-3.4751**	**7.35E-27**	RNA polymerase sigma-28 factor SigK
CLSPO_c26580	*yedF*	**0.0000**	**0.0057**	**0.0019**	**-4.0281**	**3.46E-12**	selenium metabolism protein YedF
CLSPO_c26780	*apeA*	0.0051	0.0108	0.0087	**-3.0294**	**8.99E-28**	putative M18 family aminopeptidase 1
CLSPO_c30560	*spoIIP*	0.0041*	0.0081*	**0.0051**	**-2.2823**	**2.24E-22**	stage II sporulation protein
CLSPO_c30660	*selB*	**0.0000**	**0.0012**	**0.0012**	**-6.1052**	**2.94E-21**	selenocysteine-specific elongation factor SelB
CLSPO_c30670	*selA*	**0.0008**	**0.0000**	**0.0016**	**-5.2976**	**8.62E-13**	L-seryl-tRNA(Sec) selenium transferase SelA
CLSPO_c30680	*selD*	**0.0000**	**0.0021**	**0.0010**	**-4.2036**	**2.98E-08**	selenide, water dikinase SelD
CLSPO_c30950	CLSPO_c30950	0.0048	0.0145	0.0129	**-2.4527**	**9.69E-14**	hypothetical protein UPF0758
CLSPO_c30980	CLSPO_c30980	**0.0032**	0.0127	**0.0032**	**-3.7268**	**6.11E-29**	hypothetical protein
CLSPO_c31020	CLSPO_c31020	0.0041*	0.0163	0.0122	**-3.8964**	**3.98E-26**	transglycosylase SLT domain protein
CLSPO_c31620	*tgt*	0.0039*	0.0177	0.0118	**-3.1257**	**3.60E-15**	queuine tRNA-ribosyltransferase Tgt

Table S6: List of candidate genes essential for sporulation in *C. sporogenes* (continuation).

Locus tag	Gene name	Insertion index			Log_2FC	q-value	Function
		Replicate 1	Replicate 2	Replicate 3			
CLSPO_c31630	*queA*	0.0043*	0.0184	0.0108	**-2.3326**	**8.07E-16**	S-adenosylmethionine:tRNA ribosyltransferase-isomerase QueA
CLSPO_c31660	CLSPO_c31660	**0.0018**	0.0073*	0.0073*	**-5.0708**	**9.64E-38**	hypothetical protein
CLSPO_c31830	*spoIIAA*	0.0066	0.0297	0.0231	**-2.6851**	**2.23E-13**	anti-sigma F factor antagonist
CLSPO_c32220	CLSPO_c32220	0.0074	0.0074*	0.0111	**-3.3688**	**8.99E-29**	hypothetical protein
CLSPO_c32230	CLSPO_c32230	0.0101	0.0277	0.0283	**-3.1096**	**5.28E-44**	hypothetical protein
CLSPO_c32250	CLSPO_c32250	0.0041*	0.0155	0.0113	**-3.3718**	**5.69E-26**	beta 1,4 glucosyltransferase
CLSPO_c32610	CLSPO_c32610	0.0141	0.0125	0.0109	**-3.6985**	**6.11E-16**	single-stranded DNA-binding protein
CLSPO_c32640	CLSPO_c32640	**0.0000**	0.0126	0.0084	**-2.1213**	**2.49E-05**	hypothetical protein DUF4364
CLSPO_c33110	*lgt*	0.0058	0.0190	0.0190	**-4.3797**	**7.17E-70**	prolipoprotein diacylglyceryl transferase Lgt
CLSPO_c33220	*perR*	**0.0026**	0.0207	0.0078*	**-4.0931**	**4.35E-21**	transcriptional regulator PerR
CLSPO_c33260	CLSPO_c33260	**0.0028**	0.0249	0.0331	**-2.2023**	**9.71E-08**	hypothetical protein
CLSPO_c33270	CLSPO_c33270	0.0144	0.0144	0.0337	**-3.4033**	**4.57E-16**	hypothetical protein
CLSPO_c33280	CLSPO_c33280	**0.0000**	**0.0054**	**0.0054**	**-4.1979**	**5.16E-21**	hypothetical protein
CLSPO_c33330	*clpX*	**0.0000**	**0.0000**	**0.0000**	**-10.0088**	**2.53E-41**	ATP-dependent Clp protease ATP-binding subunit ClpX
CLSPO_c33340	*clpP1*	**0.0000**	**0.0038**	**0.0019**	**-6.6955**	**1.36E-39**	ATP-dependent Clp protease proteolytic subunit
CLSPO_c33350	*tig*	**0.0017**	0.0112	0.0069*	**-3.3769**	**2.98E-24**	trigger factor Tig
CLSPO_c33450	CLSPO_c33450	**0.0015**	0.0175	0.0146	**-3.3042**	**2.96E-20**	(R)-hydroxyglutaryl-CoA dehydratase activator
CLSPO_c33470	CLSPO_c33470	0.0097	0.0185	0.0141	**-2.3952**	**1.33E-36**	putative R-2-hydroxyglutaryl-CoA dehydratase subunit
CLSPO_c33840	CLSPO_c33840	**0.0024**	0.0073*	0.0113	**-4.1635**	**2.13E-51**	hypothetical protein

Table S6: List of candidate genes essential for sporulation in *C. sporogenes* (continuation).

Locus tag	Gene name	Insertion index			Log_2FC	q-value	Function
		Replicate 1	Replicate 2	Replicate 3			
CLSPO_c33910	CLSPO_c33910	**0.0000**	**0.0000**	**0.0000**	**-9.7120**	**1.29E-33**	hypothetical protein
CLSPO_c34080	*mmgC*	0.0108	0.0137	0.0127	**-3.4902**	**1.11E-66**	acyl-CoA dehydrogenase MmgC
CLSPO_c34120	*fldA*	0.0036*	0.0125	**0.0063**	-0.9515	**3.04E-03**	E-cinnamoyl-CoA:R-phenyllactate CoA transferase FldA
CLSPO_c34420	CLSPO_c34420	**0.0024**	0.0147	0.0122	-1.6595	**2.00E-09**	manganese/zinc/iron chelate ABC transporter substrate-binding protein
CLSPO_c34540	CLSPO_c34540	**0.0014**	0.0220	0.0275	**-2.8367**	**1.17E-15**	HAD family hydrolase
CLSPO_c35320	CLSPO_c35320	**0.0000**	**0.0008**	**0.0017**	**-9.0224**	**1.45E-109**	putative gluconeogenesis factor
CLSPO_c35830	*vorA1*	**0.0000**	**0.0000**	**0.0000**	**-8.1486**	**3.92E-18**	ketoisovalerate oxidoreductase subunit VorA
CLSPO_c35840	*vorA2*	**0.0000**	**0.0000**	**0.0000**	**-9.5378**	**6.76E-30**	ketoisovalerate oxidoreductase subunit VorA
CLSPO_c35850	*vorB*	**0.0000**	**0.0010**	**0.0010**	**-6.3939**	**6.97E-29**	ketoisovalerate oxidoreductase subunit VorB
CLSPO_c35860	CLSPO_c35860	**0.0000**	**0.0053**	**0.0000**	**-6.0385**	**2.28E-07**	2-oxoacid:acceptor oxidoreductase subunit delta
CLSPO_c35870	CLSPO_c35870	**0.0017**	**0.0017**	**0.0034**	**-5.9463**	**6.50E-33**	CobQ/CobB/MinD/ParA nucleotide binding domain
CLSPO_c35880	*buk*	**0.0021**	**0.0021**	**0.0041**	**-3.2438**	**3.99E-11**	butyrate kinase Buk
CLSPO_c36730	*disA*	0.0105	0.0241	0.0178	**-2.2813**	**4.09E-15**	DNA integrity scanning protein DisA
CLSPO_c37010	*ftsH3*	**0.0025**	**0.0068**	**0.0049**	**-4.7751**	**6.77E-87**	ATP-dependent zinc metalloprotease FtsH
CLSPO_c37140	*yabQ*	0.0111	0.0222	0.0167	**-3.4011**	**2.37E-55**	spore cortex biosynthesis protein YabQ
CLSPO_c37160	CLSPO_c37160	0.0093	0.0185	0.0139	**-2.0955**	**1.13E-06**	S4 domain-containing protein

Table S6: List of candidate genes essential for sporulation in *C. sporogenes* (continuation).

Locus tag	Gene name	Insertion index			Log_2FC	q-value	Function
		Replicate 1	Replicate 2	Replicate 3			
CLSPO_c37390	CLSPO_c37390	0.0099	0.0208	0.0159	**-2.3156**	**3.27E-16**	putative phage head-tail adaptor
CLSPO_c37720	CLSPO_c37720	**0.0000**	0.0183	0.0160	-0.4944	4.18E-01	hypothetical protein
CLSPO_c37740	*rbcR*	**0.0000**	**0.0000**	**0.0012**	**-6.7873**	**4.81E-19**	putative RuBisCO transcriptional regulator
CLSPO_c38080	CLSPO_c38080	**0.0030**	0.0267	0.0193	**-2.1611**	**1.76E-10**	acyl-ACP thioesterase family protein
CLSPO_c38190	CLSPO_c38190	0.0314	0.0348	0.0453	**-2.3474**	**3.09E-18**	MazG-like family
CLSPO_c38270	CLSPO_c38270	**0.0000**	0.0180	0.0135	**-4.0921**	**2.50E-18**	hypothetical protein DUF3343
CLSPO_c38280	*csd3*	**0.0010**	**0.0038**	**0.0048**	**-5.7905**	**3.90E-72**	putative cysteine desulfurase Csd
CLSPO_c38370	*mnmG*	**0.0006**	**0.0012**	**0.0024**	**-5.7531**	**4.33E-42**	tRNA uridine 5-carboxymethylaminomethyl modification enzyme MnmG
CLSPO_c38380	*mnmE*	**0.0008**	**0.0048**	**0.0040**	**-4.4294**	**9.27E-31**	tRNA modification GTPase MnmE

List of genes essential for germination

Table S7: List of candidate genes essential for germination in *C. sporogenes*. Insertion indexes below the essentiality cut-off are represented in bold, and values in the ambiguous zone are represented with *. The log_2FC of mutant frequencies in the germination library relative to the sporulation library is also presented, as well as the respective q-value (bold numbers represent significant differences: $log_2FC < -2$ and q-value < 0.05).

		Insertion index					
Locus tag	**Gene name**	**Replicate 1**	**Replicate 2**	**Replicate 3**	**Log_2FC**	**q-value**	**Function**
CLSPO_c01340	CLSPO_c01340	**0.0030**	0.0040*	0.0030*	**-4.9120**	**3.71E-32**	putative mannosyltransferase
CLSPO_c01370	CLSPO_c01370	**0.0014**	**0.0029**	0.0029*	**-8.3949**	**1.05E-66**	hypothetical protein
CLSPO_c01390	CLSPO_c01390	**0.0010**	**0.0000**	**0.0000**	**-7.7605**	**2.65E-53**	glycosyl transferase, group 1 family protein
CLSPO_c01480	*gerKA4*	0.0190	0.0152	0.0106	**-4.3084**	**1.57E-59**	spore germination protein KA
CLSPO_c01490	CLSPO_c01490	0.0272	0.0151	0.0211	**-4.7167**	**5.21E-74**	spore germination protein
CLSPO_c01500	CLSPO_c01500	0.0141	0.0113	0.0094	**-3.8094**	**1.81E-32**	germination protein, GerC family
CLSPO_c01630	*prmC*	**0.0052**	**0.0013**	**0.0026**	**-2.7041**	**1.70E-03**	Peptide chain release factor N(5)-glutamine methyltransferase PrmC
CLSPO_c01840	*atpA1*	0.0081*	0.0117	**0.0022**	**-3.9637**	**1.67E-16**	F0F1-type ATP synthase subunit alpha
CLSPO_c01850	*atpG1*	0.0078*	**0.0026**	**0.0013**	**-3.7755**	**1.69E-19**	F0F1-type ATP synthase gamma chain
CLSPO_c02270	CLSPO_c02270	**0.0000**	0.0039*	0.0039*	**-6.5687**	**6.48E-28**	hypothetical protein DUF4363
CLSPO_c02280	CLSPO_c02280	**0.0000**	**0.0000**	0.0050	**-6.3842**	**4.37E-37**	hypothetical protein
CLSPO_c02320	CLSPO_c02320	0.0306	0.0281	0.0357	**-2.2737**	**2.93E-23**	CBS domain-containing protein
CLSPO_c06980	CLSPO_c06980	**0.0058**	0.0117	0.0029*	-1.2449	**1.47E-03**	penicillinase repressor
CLSPO_c09910	*add*	**0.0055**	0.0044*	0.0033*	**-4.7660**	**8.87E-45**	adenosine deaminase Add
CLSPO_c10820	CLSPO_c10820	0.0115	0.0097	0.0080	**-2.3921**	**7.22E-12**	pyridine nucleotide-disulfide oxidoreductase
CLSPO_c11050	*gap1*	0.0262	0.0175	0.0164	**-2.6628**	**6.66E-20**	glyceraldehyde-3-phosphate dehydrogenase Gap

Table S7: List of candidate genes essential for germination in *C. sporogenes* (continuation).

Locus tag	Gene name	Insertion index			Log_2FC	q-value	Function
		Replicate 1	Replicate 2	Replicate 3			
CLSPO_c13320	CLSPO_c13320	**0.0059**	0.0148	0.0118	-0.9960	**2.49E-04**	hypothetical protein
CLSPO_c14460	CLSPO_c14460	0.0101	0.0152	0.0101	**-2.1875**	**2.13E-09**	putative DNA-binding protein
CLSPO_c14470	*dacF1*	0.0096	0.0052	0.0070	**-2.4026**	**7.38E-16**	D-alanyl-D-alanine carboxypeptidase DacF
CLSPO_c14510	CLSPO_c14510	0.0246	0.0000	0.0049	-1.3610	5.43E-02	hypothetical protein
CLSPO_c14620	*rluC*	**0.0058**	0.0047*	0.0058	**-2.0835**	**2.66E-04**	ribosomal large subunit pseudouridine synthase C
CLSPO_c15000	CLSPO_c15000	**0.0051**	**0.0034**	0.0034*	-1.6279	**6.09E-03**	cell-division initiation protein DivIVA
CLSPO_c15010	*aroB*	0.0112	0.0102	0.0061	**-2.5559**	**3.80E-11**	3-dehydroquinate synthase AroB
CLSPO_c15410	CLSPO_c15410	0.0342	0.0302	0.0302	**-2.7075**	**3.82E-18**	hypothetical protein
CLSPO_c16090	*hisG*	0.0291	0.0240	0.0171	**-2.1381**	**2.57E-12**	ATP phosphoribosyltransferase HisG
CLSPO_c16760	*metF*	**0.0063**	0.0050*	0.0088	-1.9270	**3.78E-06**	5,10-methylenetetrahydrofolate reductase MetF
CLSPO_c17030	CLSPO_c17030	**0.0026**	0.0045*	**0.0008**	**-2.7709**	**3.85E-08**	type I site-specific restriction-modification system R subunit
CLSPO_c17430	*dppC1*	0.0257	0.0243	0.0243	**-2.7671**	**3.51E-27**	dipeptide transport system permease protein DppC
CLSPO_c17490	*hsp18*	0.0076*	0.0050*	0.0101	**-3.0123**	**2.83E-09**	18 kDa heat shock protein
CLSPO_c17590	CLSPO_c17590	0.0175	0.0066	0.0109	**-2.7869**	**8.36E-08**	PfkB family kinase
CLSPO_c17600	*safA*	**0.0052**	0.0087	0.0052	**-2.1936**	**2.71E-04**	spore coat assembly protein SafA
CLSPO_c17690	*cheV*	0.0236	0.0161	0.0137	**-2.1292**	**8.72E-11**	chemotaxis protein CheV
CLSPO_c17710	CLSPO_c17710	0.0371	0.0305	0.0225	**-2.2981**	**8.34E-13**	RNA-binding protein
CLSPO_c17810	CLSPO_c17810	0.0110	**0.0000**	**0.0000**	**-2.1916**	**2.51E-03**	small, acid-soluble spore protein beta

Table S7: List of candidate genes essential for germination in *C. sporogenes* (continuation).

Locus tag	Gene name	Insertion index			Log_2FC	q-value	Function
		Replicate 1	Replicate 2	Replicate 3			
CLSPO_c17840	CLSPO_c17840	0.0187	0.0094	0.0164	**-2.3306**	**4.74E-14**	hypothetical protein
CLSPO_c17910	*miaB*	0.0230	0.0156	0.0107	**-2.3228**	**5.07E-10**	(dimethylallyl)adenosine tRNA methylthiotransferase MiaB
CLSPO_c18030	CLSPO_c18030	**0.0063**	**0.0025**	0.0038*	-1.7552	**8.91E-05**	SIS domain-containing protein
CLSPO_c18360	CLSPO_c18360	0.0129	0.0090	0.0090	**-2.1275**	**5.85E-09**	AraC family transcriptional regulator
CLSPO_c18370	CLSPO_c18370	0.0072*	0.0072	**0.0024**	**-2.1581**	**5.38E-04**	transcription activator effector binding protein
CLSPO_c18470	*dacB1*	**0.0029**	**0.0020**	**0.0020**	**-5.0683**	**1.73E-28**	D-alanyl-D-alanine carboxypeptidase DacB
CLSPO_c18500	*scpB*	0.0134	0.0076	0.0057	**-2.4188**	**1.41E-09**	segregation and condensation protein B
CLSPO_c18570	*punA*	0.0082*	**0.0014**	0.0041	-1.5130	**9.21E-04**	purine nucleoside phosphorylase 1
CLSPO_c18750	CLSPO_c18750	0.0097	0.0039*	0.0058	**-2.4873**	**2.34E-11**	hypothetical protein
CLSPO_c18760	CLSPO_c18760	0.0102	0.0064	0.0051	**-2.6699**	**1.11E-07**	farnesyl diphosphate synthase
CLSPO_c18790	*nusB*	0.0103	0.0051	0.0103	**-2.9617**	**1.26E-10**	N utilization substance protein B
CLSPO_c18920	*aroK*	0.0108	0.0152	0.0108	**-2.1339**	**3.69E-07**	shikimate kinase AroK
CLSPO_c19080	CLSPO_c19080	0.0205	0.0176	0.0147	**-3.1213**	**3.68E-53**	HIT family protein
CLSPO_c19120	CLSPO_c19120	0.0194	0.0112	0.0102	**-2.3771**	**3.60E-09**	hypothetical protein
CLSPO_c19140	CLSPO_c19140	0.0236	0.0146	0.0079	**-2.3570**	**3.78E-09**	anti-sigma factor
CLSPO_c19200	*msrA2*	0.0211	0.0117	0.0141	**-2.3148**	**8.24E-12**	peptide methionine sulfoxide reductase MsrA
CLSPO_c19250	CLSPO_c19250	0.0196	0.0130	0.0065	**-2.0787**	**2.07E-06**	ABC transporter, permease protein
CLSPO_c19290	CLSPO_c19290	0.0162	0.0124	0.0110	**-2.6540**	**4.45E-25**	putative ABC transporter ATP-binding/permease protein

Table S7: List of candidate genes essential for germination in *C. sporogenes* (continuation).

Locus tag	Gene name	Insertion index			Log_2FC	q-value	Function
		Replicate 1	Replicate 2	Replicate 3			
CLSPO_c19310	*bltD*	0.0126	0.0101	0.0151	**-2.1680**	**1.07E-07**	spermine/spermidine acetyltransferase BltD
CLSPO_c19420	CLSPO_c19420	0.0192	0.0112	0.0064	**-2.5211**	**7.49E-14**	ABC-type sugar transport system, ATPase component
CLSPO_c19440	CLSPO_c19440	0.0104	0.0104	0.0052	**-2.7780**	**4.75E-13**	ABC-type tungstate transport system, permease component
CLSPO_c19540	CLSPO_c19540	**0.0000**	**0.0035**	0.0035*	**-2.8463**	**1.73E-03**	iron-sulfur flavoprotein
CLSPO_c19600	CLSPO_c19600	0.0091	0.0075	0.0037*	**-2.3442**	**3.42E-13**	carbohydrate binding protein
CLSPO_c19630	*aldR*	**0.0060**	0.0090	0.0060	-0.6565	4.12E-01	putative regulator AldR
CLSPO_c19640	CLSPO_c19640	0.0130	0.0043*	0.0087	**-3.1267**	**4.97E-06**	glutaredoxin-like protein
CLSPO_c19680	*inlA2*	0.0167	0.0145	0.0123	**-2.1205**	**3.86E-11**	internalin-A
CLSPO_c19770	*gerKA2*	0.0230	0.0207	0.0170	**-2.0635**	**8.13E-10**	spore germination protein KA
CLSPO_c19800	CLSPO_c19800	0.0244	0.0112	0.0092	**-2.3551**	**7.38E-15**	spore germination protein
CLSPO_c19850	CLSPO_c19850	**0.0039**	0.0039*	0.0079	**-2.3017**	**4.92E-02**	hypothetical protein
CLSPO_c19950	*treA*	**0.0026**	**0.0026**	0.0040	**-3.1305**	**8.66E-13**	trehalose-6-phosphate hydrolase TreA
CLSPO_c20130	CLSPO_c20130	0.0203	0.0142	0.0122	**-2.2404**	**3.81E-11**	adenine phosphoribosyltransferase
CLSPO_c20250	*artP*	**0.0054**	0.0108	0.0054	-1.0895	**6.05E-03**	arginine-binding extracellular protein ArtP
CLSPO_c20340	CLSPO_c20340	0.0156	0.0104	0.0052	**-2.2035**	**4.06E-04**	putative nucleic-acid-binding protein
CLSPO_c20390	CLSPO_c20390	**0.0051**	0.0051	0.0051	-1.4254	**2.54E-02**	DNA-binding helix-turn-helix protein
CLSPO_c20600	*codA*	0.0184	0.0158	0.0123	**-2.1100**	**2.35E-26**	cytosine deaminase CodA
CLSPO_c20630	CLSPO_c20630	0.0118	0.0089	0.0079	**-2.0070**	**3.27E-08**	radical SAM domain-containing protein

Table S7: List of candidate genes essential for germination in *C. sporogenes* (continuation).

Locus tag	Gene name	Insertion index			Log_2FC	q-value	Function
		Replicate 1	Replicate 2	Replicate 3			
CLSPO_c20650	CLSPO_c20650	0.0077*	0.0093	**0.0015**	-0.1246	1.00E+00	ABC-type transport system involved in Fe-S cluster assembly, ATPase component
CLSPO_c20690	*xylR*	0.0089	0.0059	0.0039	**-2.0534**	**6.06E-09**	xylose repressor
CLSPO_c20710	CLSPO_c20710	**0.0054**	0.0036*	0.0036*	**-2.3163**	**5.47E-04**	TetR family transcriptional regulator
CLSPO_c20740	CLSPO_c20740	0.0129	0.0076	0.0083	**-2.0132**	**2.78E-07**	putative ABC transporter ATP-binding protein
CLSPO_c20760	CLSPO_c20760	0.0127	0.0091	**0.0018**	**-2.5389**	**2.35E-09**	conserved hypothetical integral membrane protein TIGR02185
CLSPO_c20820	CLSPO_c20820	0.0129	0.0119	0.0046	**-2.5778**	**1.35E-11**	putative mutase
CLSPO_c20910	CLSPO_c20910	0.0175	0.0175	0.0088	**-3.0746**	**5.73E-08**	hypothetical protein
CLSPO_c21110	CLSPO_c21110	0.0176	0.0088	0.0088	**-2.0772**	**1.14E-04**	hypothetical protein
CLSPO_c21240	*eutN*	**0.0042**	0.0042*	**0.0000**	**-2.5219**	**8.17E-04**	ethanolamine utilization protein EutN
CLSPO_c21320	*adhE3*	0.0089	0.0089	0.0067	**-2.5168**	**3.55E-16**	aldehyde-alcohol dehydrogenase AdhE
CLSPO_c21470	CLSPO_c21470	0.0152	0.0121	0.0121	**-2.1340**	**1.99E-11**	hypothetical protein
CLSPO_c21720	*kanC*	0.0205	0.0127	0.0127	**-2.1111**	**1.55E-15**	2-deoxy-scyllo-inososе synthase KanC
CLSPO_c21940	*tal2*	0.0124	0.0108	0.0139	**-2.2420**	**4.38E-11**	putative transaldolase Tal
CLSPO_c22000	CLSPO_c22000	0.0125	**0.0031**	0.0031*	**-2.9799**	**1.85E-06**	M24 family metallopeptidase
CLSPO_c22010	CLSPO_c22010	**0.0051**	**0.0000**	0.0051	**-3.5780**	**6.60E-04**	M24 family metallopeptidase
CLSPO_c22050	CLSPO_c22050	0.0255	0.0170	0.0170	**-2.0014**	**6.81E-08**	hypothetical protein
CLSPO_c22150	*phoR2*	0.0116	0.0058	0.0058	**-2.7345**	**6.50E-10**	alkaline phosphatase synthesis sensor protein PhoR
CLSPO_c22750	CLSPO_c22750	0.0164	0.0082	**0.0000**	-0.9037	3.72E-01	hypothetical protein

Table S7: List of candidate genes essential for germination in *C. sporogenes* (continuation).

Locus tag	Gene name	Insertion index			Log_2FC	q-value	Function
		Replicate 1	Replicate 2	Replicate 3			
CLSPO_c22790	CLSPO_c22790	0.0217	0.0160	0.0132	**-2.2705**	**1.11E-13**	ABC-type Na+ efflux pump, permease component
CLSPO_c22810	CLSPO_c22810	0.0252	0.0168	0.0252	**-2.5571**	**2.67E-05**	hypothetical protein
CLSPO_c22870	CLSPO_c22870	0.0349	0.0295	0.0241	**-2.1671**	**1.37E-13**	hypothetical protein DUF2809
CLSPO_c22950	CLSPO_c22950	0.0150	0.0111	0.0103	**-2.2791**	**8.80E-15**	amino acid permease family protein
CLSPO_c23050	CLSPO_c23050	0.0172	0.0172	0.0129	**-2.0412**	**4.67E-12**	saccharopine dehydrogenase
CLSPO_c23640	CLSPO_c23640	0.0345	**0.0000**	0.0230	-1.6334	**3.53E-02**	hypothetical protein
CLSPO_c23850	CLSPO_c23850	0.0151	0.0151	0.0101	**-2.0045**	**7.58E-10**	hypothetical protein
CLSPO_c24040	CLSPO_c24040	0.0244	0.0198	0.0158	**-2.2054**	**8.12E-16**	ribonuclease
CLSPO_c24950	*purB*	0.0179	0.0140	0.0140	**-3.1193**	**6.63E-26**	adenylosuccinate lyase PurB
CLSPO_c25850	CLSPO_c25850	0.0235	0.0172	0.0072	**-2.3798**	**3.84E-14**	hypothetical protein UPF0348
CLSPO_c25860	CLSPO_c25860	**0.0057**	**0.0010**	**0.0010**	**-2.4292**	**1.13E-17**	nucleoside recognition domain-containing protein
CLSPO_c26420	*typA*	**0.0018**	**0.0018**	**0.0012**	**-2.5486**	**1.46E-05**	GTP-binding protein TypA/BipA
CLSPO_c26450	CLSPO_c26450	0.0174	0.0174	0.0174	**-2.2770**	**3.64E-09**	hypothetical protein DUF1292
CLSPO_c26720	CLSPO_c26720	0.0116	0.0047*	**0.0023**	**-3.0377**	**2.03E-12**	putative lipoprotein
CLSPO_c26900	*glpX*	**0.0023**	**0.0034**	**0.0011**	**-3.5676**	**1.02E-09**	fructose-1,6-bisphosphatase class 2
CLSPO_c27880	*rfbH*	**0.0026**	0.0077	0.0034*	**-4.8783**	**2.71E-24**	lipopolysaccharide biosynthesis protein RfbH
CLSPO_c27890	*rfbG*	**0.0061**	0.0122	0.0051	**-4.7032**	**1.23E-25**	CDP-glucose 4,6-dehydratase RfbG
CLSPO_c27900	*rfbF*	0.0071*	0.0114	0.0028*	**-3.4167**	**6.29E-11**	glucose-1-phosphate cytidylyltransferase RfbF
CLSPO_c27910	CLSPO_c27910	0.0390	0.0260	0.0292	**-2.0454**	**2.11E-02**	ISCb1g3, transposase
CLSPO_c27930	CLSPO_c27930	0.0073*	0.0117	0.0037*	**-4.3008**	**4.84E-22**	hypothetical protein
CLSPO_c27940	CLSPO_c27940	0.0139	0.0160	0.0077	**-3.5334**	**8.07E-32**	hypothetical protein

Table S7: List of candidate genes essential for germination in *C. sporogenes* (continuation).

Locus tag	Gene name	Insertion index			Log_2FC	q-value	Function
		Replicate 1	Replicate 2	Replicate 3			
CLSPO_c27950	CLSPO_c27950	0.0126	0.0139	0.0063	**-3.6163**	**4.31E-23**	putative sugar epimerase
CLSPO_c27960	CLSPO_c27960	0.0161	0.0101	0.0080	**-3.7232**	**2.82E-32**	putative methyltransferase
CLSPO_c27980	*ilvB*	**0.0043**	0.0086	**0.0025**	**-3.9605**	**4.25E-13**	acetolactate synthase large subunit
CLSPO_c27990	*gmhA*	0.0104	0.0083	0.0062	**-4.5690**	**7.68E-34**	phosphoheptose isomerase GmhA
CLSPO_c28000	CLSPO_c28000	0.0067*	0.0108	**0.0027**	**-4.2535**	**2.01E-27**	N-acetylneuraminate synthase
CLSPO_c28010	CLSPO_c28010	0.0100	0.0132	0.0061	**-4.3618**	**2.47E-42**	glycosyl transferase family 2
CLSPO_c28030	CLSPO_c28030	0.0079*	0.0084	0.0053	**-4.3041**	**1.99E-32**	glycosyltransferase
CLSPO_c28700	CLSPO_c28700	0.0158	0.0065	0.0065	**-4.1124**	**2.80E-29**	spore germination protein, GerAC
CLSPO_c28720	*gerKA3*	0.0157	0.0112	0.0082	**-3.7982**	**5.26E-31**	spore germination protein KA
CLSPO_c30260	*trmK*	0.0193	0.0113	0.0064	**-3.0872**	**2.71E-19**	tRNA (adenine(22)-N(1))-methyltransferase TrmK
CLSPO_c30410	CLSPO_c30410	**0.0000**	**0.0019**	**0.0000**	**-7.4529**	**1.29E-47**	putative stage IV sporulation protein
CLSPO_c30540	*lepA*	**0.0037**	0.0043*	**0.0018**	**-3.0388**	**3.70E-10**	elongation factor 4
CLSPO_c30570	*gpr*	0.0068*	0.0091	0.0057	**-6.2323**	**5.75E-67**	germination protease Gpr
CLSPO_c30850	CLSPO_c30850	0.0300	0.0274	0.0248	**-3.0928**	**9.83E-40**	putative metalloprotease
CLSPO_c31780	CLSPO_c31780	**0.0031**	0.0062	**0.0000**	**-7.6113**	**3.71E-45**	SpoVA family protein
CLSPO_c31790	*spoVAD*	**0.0000**	**0.0000**	**0.0000**	**-7.6144**	**7.83E-60**	stage V sporulation protein AD
CLSPO_c31800	*spoVAC*	**0.0000**	**0.0000**	**0.0000**	**-5.9631**	**6.70E-20**	stage V sporulation protein AC
CLSPO_c31810	*sigF*	**0.0000**	**0.0029**	**0.0000**	**-7.7055**	**2.38E-53**	RNA polymerase sigma-F factor SigF
CLSPO_c31860	CLSPO_c31860	**0.0022**	0.0045*	**0.0022**	**-5.8832**	**1.02E-52**	polysaccharide deacetylase family protein
CLSPO_c32100	*pgcA1*	**0.0013**	**0.0000**	**0.0006**	**-6.9153**	**7.66E-51**	phosphoglucomutase PgcA
CLSPO_c32330	*ktrC*	0.0101	0.0101	0.0084	**-2.4043**	**8.94E-10**	Ktr system potassium uptake protein C

Table S7: List of candidate genes essential for germination in *C. sporogenes* (continuation).

Locus tag	Gene name	Insertion index			Log_2FC	q-value	Function
		Replicate 1	Replicate 2	Replicate 3			
CLSPO_c32340	*ktrB*	0.0107	0.0083	0.0083	**-2.9187**	**1.20E-17**	Ktr system potassium uptake protein B
CLSPO_c32730	*fldI3*	0.0078*	0.0047*	0.0052	**-2.9691**	**1.60E-19**	R-phenyllactate dehydratase activator
CLSPO_c33970	CLSPO_c33970	**0.0055**	0.0050*	0.0039*	**-5.4178**	**6.70E-41**	triple tyrosine motif-containing protein
CLSPO_c34630	CLSPO_c34630	0.0427	0.0305	0.0326	**-2.3681**	**1.31E-31**	hypothetical protein
CLSPO_c36050	CLSPO_c36050	**0.0063**	0.0126	0.0047	**-5.1676**	**1.50E-37**	N-acetylmuramoyl-L-alanine amidase
CLSPO_c36110	*ecfA2*	**0.0051**	**0.0013**	**0.0026**	**-2.6781**	**2.20E-07**	energy-coupling factor transporter ATP-binding protein EcfA
CLSPO_c37040	*spoIIE*	**0.0019**	**0.0019**	**0.0009**	**-7.8132**	**2.86E-98**	stage II sporulation protein E
CLSPO_c37150	*yabP*	**0.0000**	**0.0000**	**0.0000**	**-8.4141**	**3.26E-20**	sporulation protein YabP
CLSPO_c37190	*spoVB2*	**0.0029**	**0.0022**	**0.0007**	**-8.1847**	**1.60E-84**	stage V sporulation protein B
CLSPO_c37200	*spoVT*	0.0080*	**0.0020**	0.0040	**-4.5177**	**3.65E-17**	stage V sporulation protein T
CLSPO_c38130	*pgcA2*	**0.0052**	**0.0026**	0.0032*	**-3.7080**	**1.53E-17**	phosphoglucomutase PgcA
CLSPO_c38150	*lon3*	0.0179	0.0104	0.0104	**-2.2153**	**3.25E-14**	Lon protease

Chapter 5: Supporting information

RNA quality control

Table S8: Concentrations and quality of the total RNA used for RT-PCR. The RNA was extracted from *C. sporogenes* WT. Concentration and $A_{260/280}$ were measured using the Nanodrop. The ratio of absorbance ($A_{260/280}$) indicates the purity of the RNA, and optimal values should be between 1.8 and 2.2. RIN and 23S/16S were measured using the 2100 Bioanalyzer (Agilent). The RNA integrity number (RIN) indicates the quality of the RNA, and optimal values should be between 8 and 10. The 23S/16S ratio should be around 2.

Time after inoculation	Biological replicate	RNA concentration (ng/µL)	$A_{260/280}$	RIN	23S/16S
5 hours	1	1347.9	2.21	10	1.9
	2	1465.0	2.20	10	1.8
8 hours	1	1366.8	2.20	10	1.8
	2	1614.7	2.15	9.7	1.8
10 hours	1	1053.1	2.21	10	1.7
	2	1448.1	2.19	9.9	1.8
14 hours	1	1323.1	2.20	9.6	1.8
	2	1170.7	2.21	9.5	1.8
17 hours	1	952.1	2.21	9.7	1.6
	2	841.9	2.22	9.3	1.7
20 hours	1	698.0	2.21	9.9	1.7
	2	810.5	2.22	9.8	1.5
24 hours	1	816.8	2.22	8.9	1.9
	2	584.4	2.21	7.6	1.8

Table S9: Concentrations and quality of the total RNA used for RT-qPCR and RNA-Seq. The RNA was extracted from *C. sporogenes* WT and mutants. Concentration and $A_{260/280}$ were measured using the Nanodrop. The ratio of absorbance ($A_{260/280}$) indicates the purity of the RNA, and optimal values should be between 1.8 and 2.2. RIN and 23S/16S were measured using the 2100 Bioanalyzer (Agilent). The RNA integrity number (RIN) indicates the quality of the RNA, and optimal values should be between 8 and 10. The 23S/16S ratio should be around 2.

Strain	Time after inoculation	Biological replicate	RNA concentration (ng/µL)	$A_{260/280}$	RIN	23S/16S
***C. sporogenes* WT**	17 hours	1	1837.5	1.98	9.1	1.8
		2	1492.8	2.16	9.2	1.7
		3	1738.8	2.07	9.3	1.9
	20 hours	1	1158.8	2.19	9.5	1.9
		2	1449.9	2.19	9.1	1.8
		3	1652.3	2.12	8.6	1.8
	24 hours	1	859.7	2.19	9	1.8
		2	1279.1	2.20	10	1.6
		3	1138.5	2.21	8.5	1.8
C. sporogenes* Δ*sigF	17 hours	1	1868.6	1.95	10	1.8
		2	1616.9	2.14	9	2.2
		3	1472.2	2.18	8.4	1.9
C. sporogenes* Δ*sigE	17 hours	1	1838.8	1.97	10	1.8
		2	1798.0	2.01	9.3	2.1
		3	1413.2	2.19	9.3	1.4
C. sporogenes* Δ*sigG	20 hours	1	1679.4	2.11	8.5	2.6
		2	1189.6	2.21	8.5	1.5
		3	1804.9	1.96	9.9	1.6
C. sporogenes* Δ*sigK	24 hours	1	1391.6	2.19	8.6	1.7
		2	1268.3	2.20	9.7	1.8
		3	1508.8	2.16	8.5	1.7

qPCR efficiencies

Table S10: Efficiencies obtained during RT-qPCR experiments. Ideally, efficiency should be around 2 – amount of DNA should double at each cycle.

Target gene	Efficiency
rrfA	2.011
gyrA	2.049
spoIID	1.957
spoIIM	1.984
gpr	1.979
spoIVB	1.899
sspH1	1.971
cotA	1.728
cotI	1.928

RNA-Seq: Sequencing data

Table S11: Summary of the results obtained using the NextSeq Illumina platform. Three biological replicates per strain were analysed. For each sample, the strain name is followed by the number of the replicate and the time point of RNA extraction.

Samples	Trimmed Reads (pairs)	Mapped Reads		Uniquely Mapped Reads	
		Counts (pairs)	Percentage	Counts (pairs)	Percentage
sigE1-17h	21,518,063	21,059,728	97.87%	20,456,880	95.07%
sigE2-17h	20,213,733	19,609,342	97.01%	18,544,864	91.74%
sigE3-17h	22,413,688	21,884,725	97.64%	20,420,637	91.11%
sigF1-17h	22,220,802	21,791,941	98.07%	21,136,016	95.12%
sigF2-17h	23,404,827	22,815,025	97.48%	21,744,720	92.91%
sigF3-17h	23,488,456	21,581,193	91.88%	19,461,818	82.86%
sigG1-20h	22,219,097	21,452,538	96.55%	19,022,307	85.61%
sigG2-20h	22,252,952	21,718,881	97.60%	20,338,066	91.39%
sigG3-20h	23,180,112	22,607,563	97.53%	21,736,606	93.77%
sigK1-24h	23,356,155	22,435,922	96.06%	20,402,926	87.36%
sigK2-24h	22,007,482	21,417,681	97.32%	20,257,983	92.05%
sigK3-24h	22,232,216	21,725,321	97.72%	20,203,790	90.88%
WT1-17h	21,959,696	21,382,156	97.37%	20,482,194	93.27%
WT1-20h	24,523,680	23,577,066	96.14%	21,920,669	89.39%
WT1-24h	23,588,018	22,991,241	97.47%	21,012,434	89.08%
WT2-17h	22,992,398	20,079,261	87.33%	17,796,325	77.40%
WT2-20h	23,810,161	23,150,620	97.23%	22,093,054	92.79%
WT2-24h	23,592,493	23,002,681	97.50%	22,162,861	93.94%
WT3-17h	21,090,278	20,592,547	97.64%	19,851,068	94.12%
WT3-20h	23,184,496	22,628,068	97.60%	21,127,425	91.13%
WT3-24h	23,009,858	22,310,358	96.96%	18,694,046	81.24%

RNA-Seq: PCA analysis

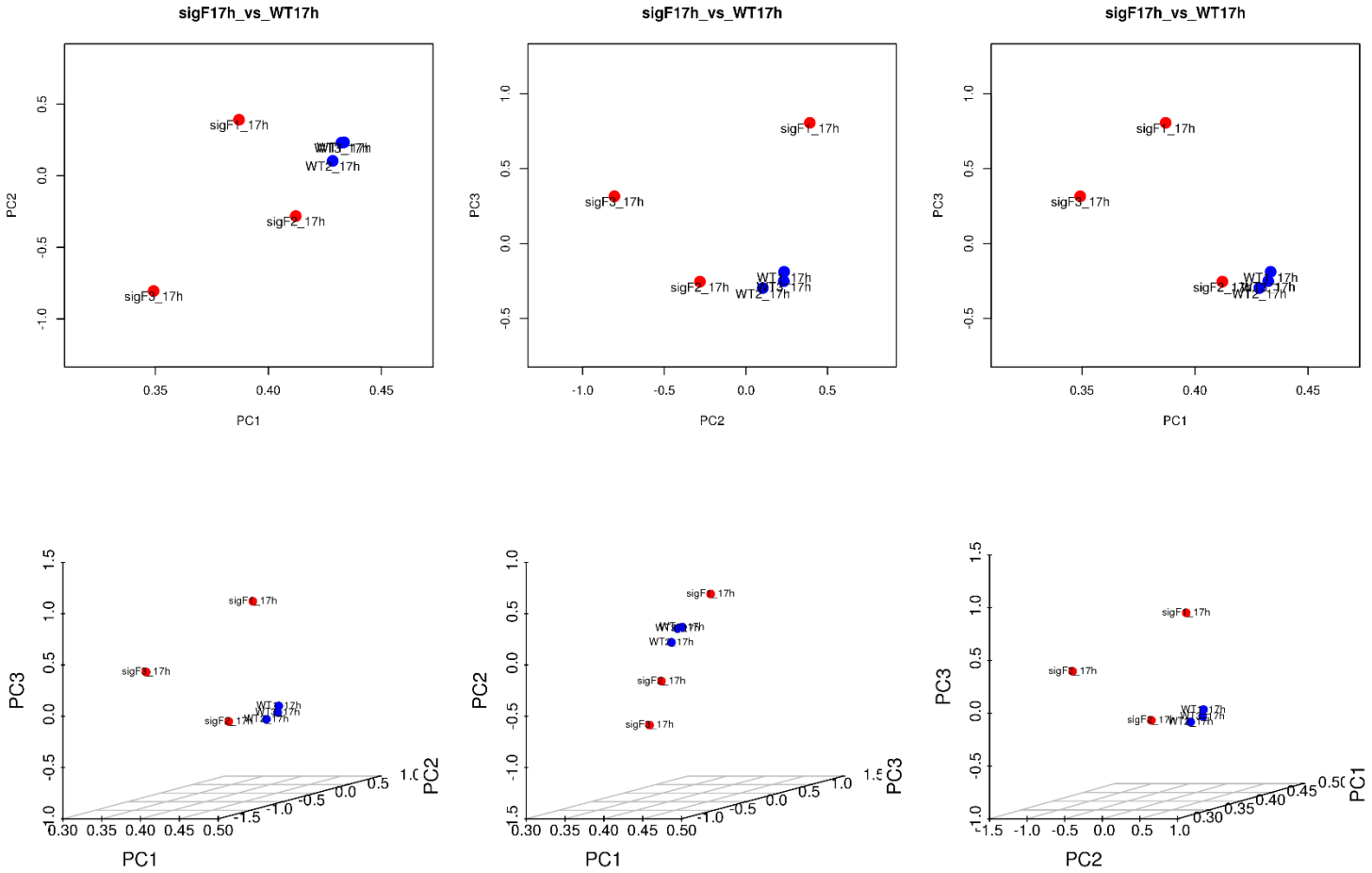


Figure S10: PCA analysis (*sigF* mutant). The three biological replicates from the *sigF* mutant (17 hours) are compared to the three biological replicates from wild type (17 hours).

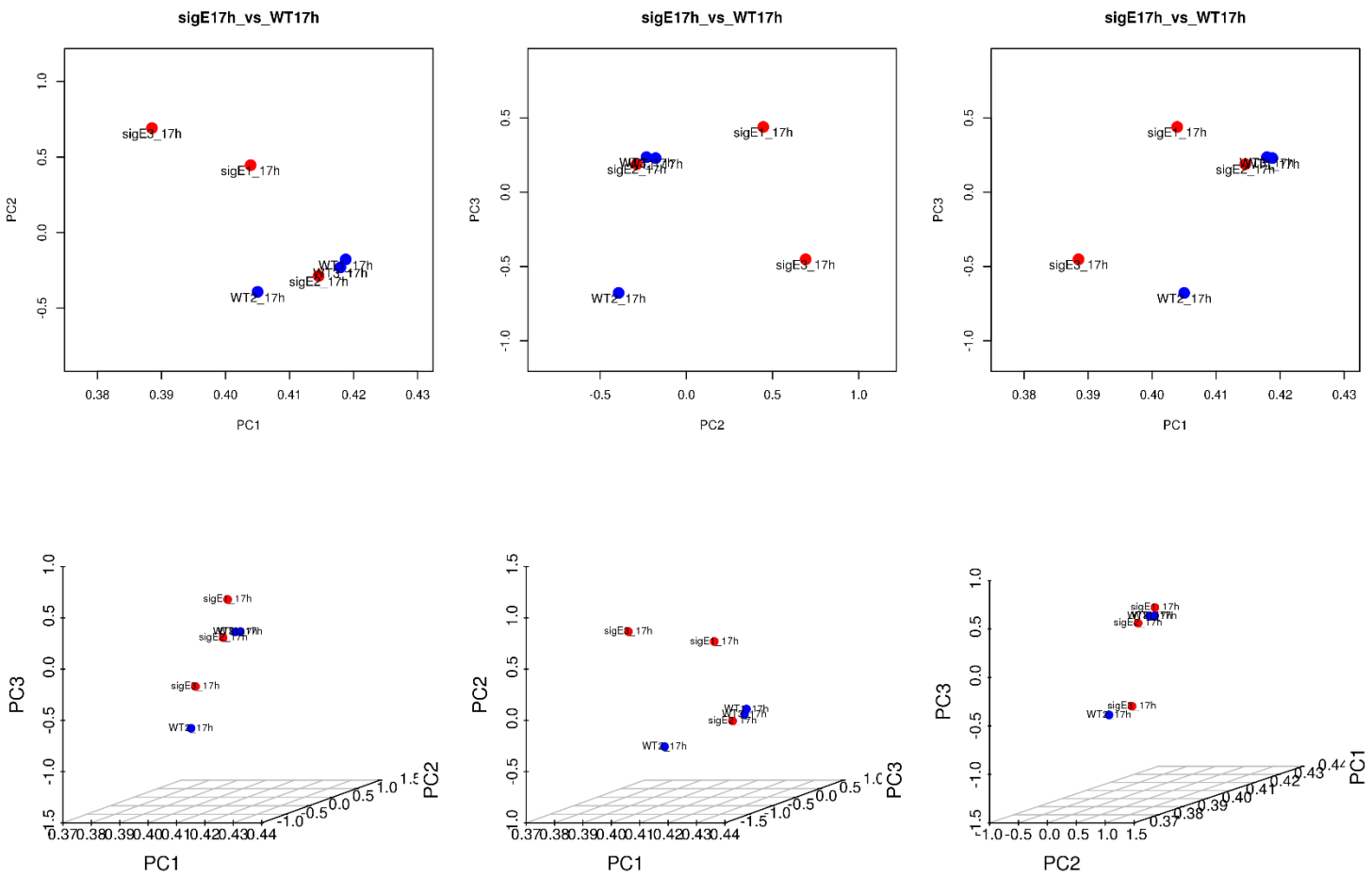


Figure S11: PCA analysis (*sigE* mutant). The three biological replicates from the *sigE* mutant (17 hours) are compared to the three biological replicates from wild type (17 hours).

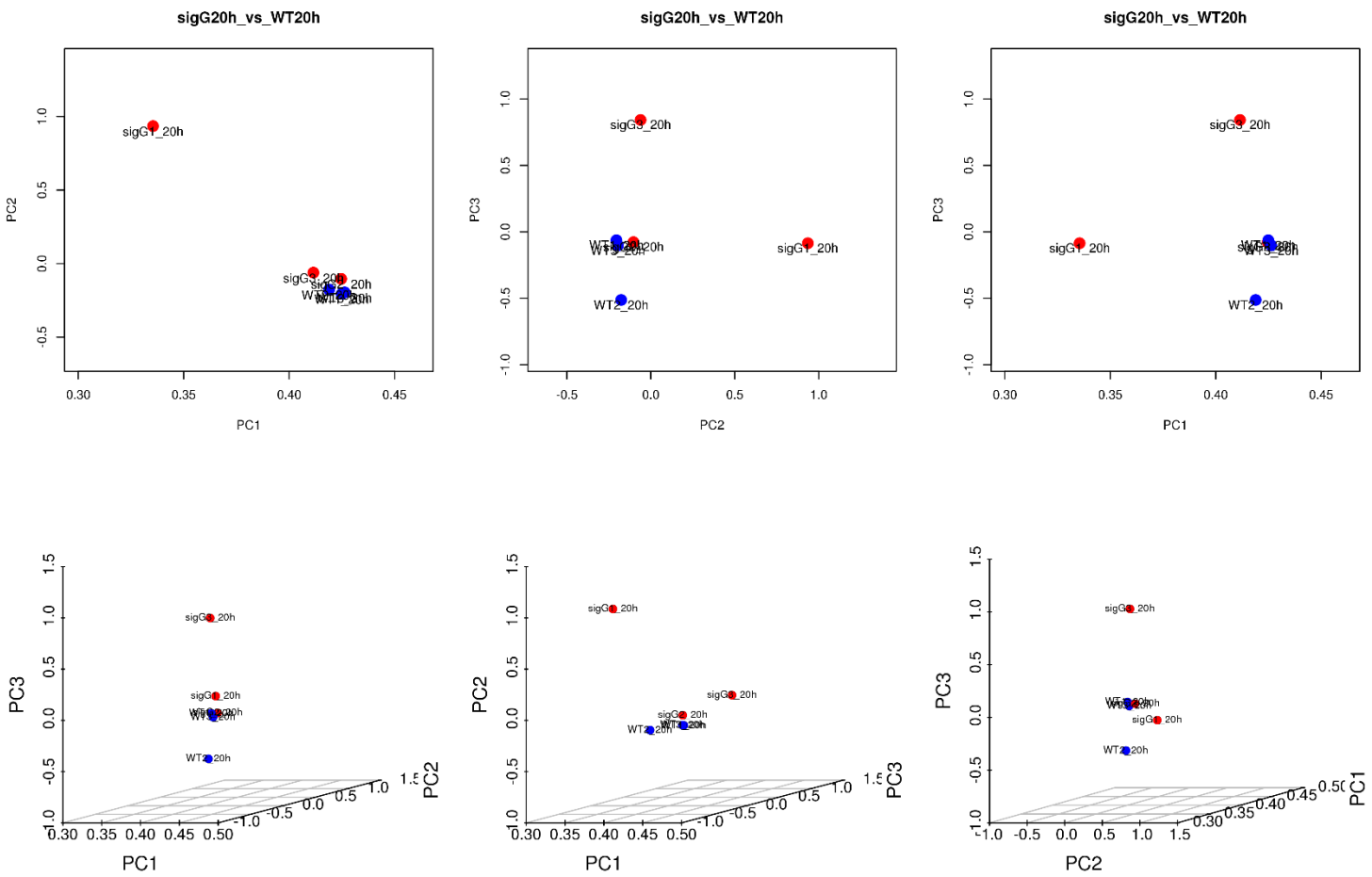


Figure S12: PCA analysis (*sigG* mutant). The three biological replicates from the *sigG* mutant (20 hours) are compared to the three biological replicates from wild type (20 hours).

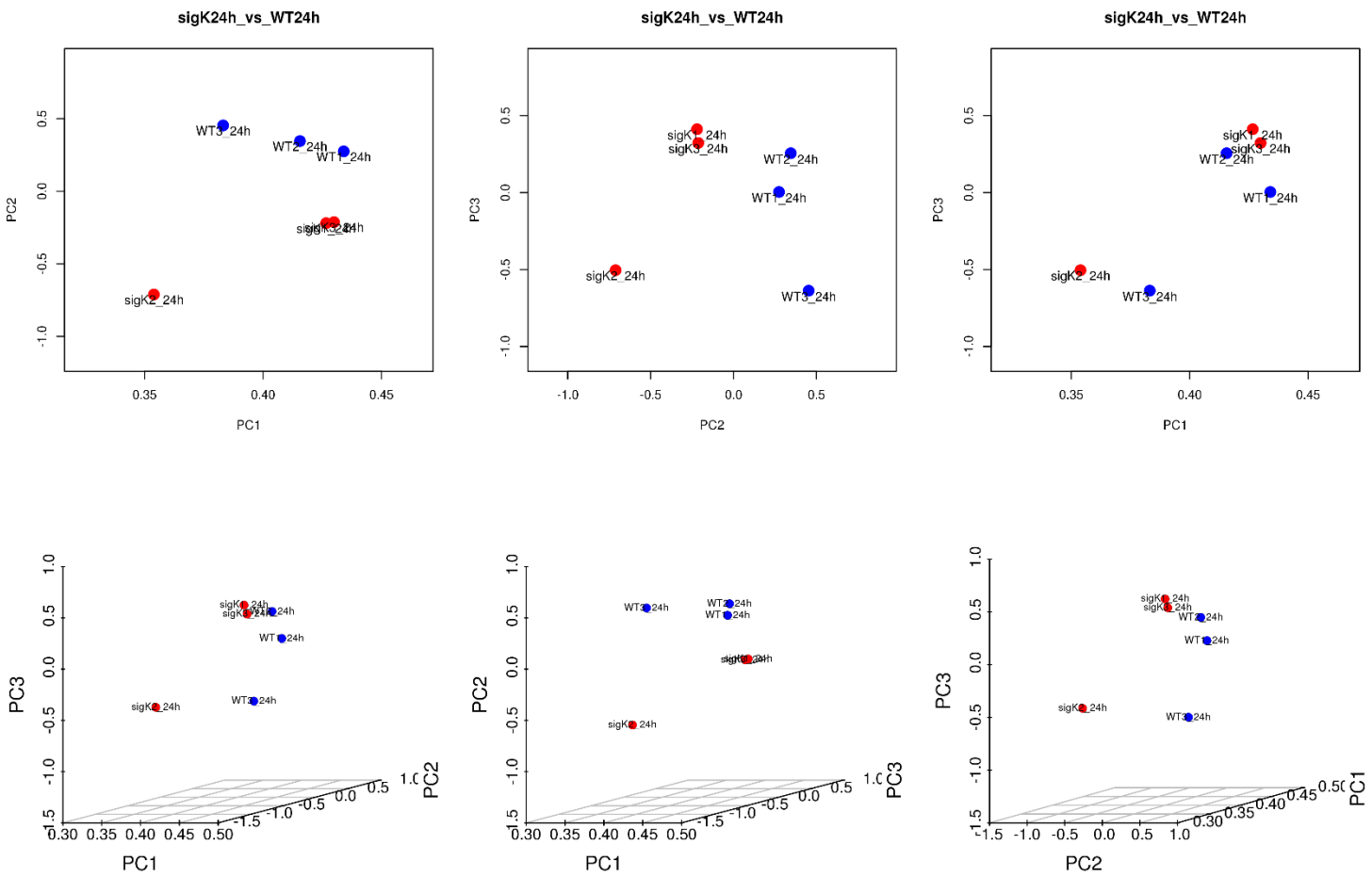


Figure S13: PCA analysis (*sigK* mutant). The three biological replicates from the *sigK* mutant (24 hours) are compared to the three biological replicates from wild type (24 hours).

RNA-Seq: Heatmaps

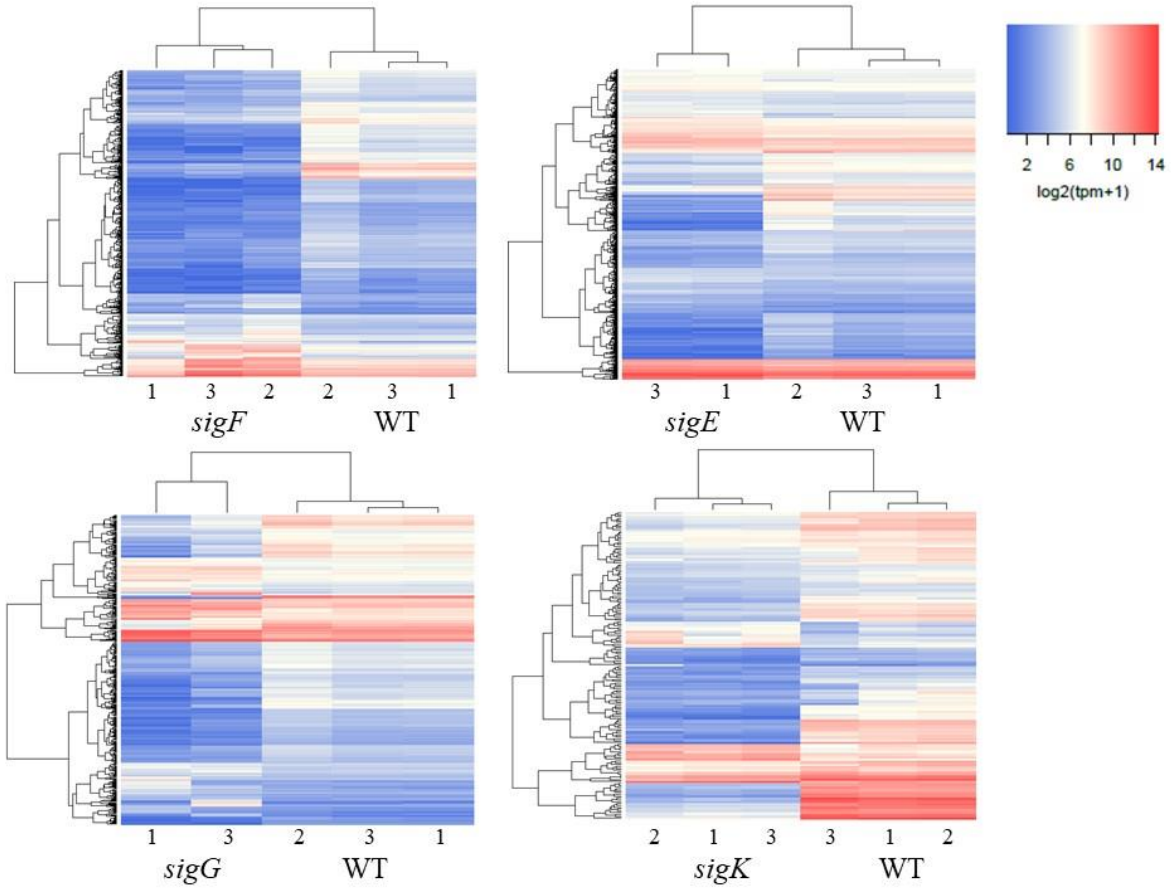


Figure S14: Whole transcriptome heatmaps. The heatmaps represent gene expression levels as Log_{10} (FPKM+1), and genes are grouped into dendograms using hierarquical clustering. The numbers under each heatmap correspond to the biological replicate, and the strain is also indicated.

RNA-Seq: List of genes from the σ^F regulon

Table S12: List of genes from the σ^F regulon. All genes have a $log_2FC < -1$ and adjusted p-value ≤ 0.05.

Gene ID	Gene name	Gene function	Log_2FC	P-value
CLSPO_c04270	*ssb*	single-stranded DNA-binding protein Ssb	-2.17	2.10E-04
CLSPO_c09570	*asrB*	anaerobic sulfite reductase subunit B	-2.85	1.45E-02
CLSPO_c13480	*uppS1*	isoprenyl transferase UppS	-3.01	9.20E-05
CLSPO_c16540	CLSPO_c16540	ABC transporter ATP-binding/permease protein	-1.77	1.19E-02
CLSPO_c19180	CLSPO_c19180	Two-component sensor kinase	-1.02	1.86E-04
CLSPO_c19450	CLSPO_c19450	Na+-driven multidrug efflux pump	-1.09	1.56E-05
CLSPO_c19590	CLSPO_c19590	leucine rich repeat protein	-1.09	8.49E-07
CLSPO_c20330	*catB2*	chloramphenicol acetyltransferase CatB	-1.03	5.08E-04
CLSPO_c20600	*codA*	cytosine deaminase CodA	-1.57	3.19E-05
CLSPO_c20610	*codB*	cytosine permease CodB	-1.26	3.14E-02
CLSPO_c21190	CLSPO_c21190	ethanolamine utilization protein PduA/PduJ	-3.06	3.42E-08
CLSPO_c21200	*pduL*	phosphate propanoyltransferase PduL	-3.37	3.71E-09
CLSPO_c21210	*adhE2*	aldehyde-alcohol dehydrogenase AdhE	-3.31	9.94E-10
CLSPO_c21220	CLSPO_c21220	microcompartment family protein	-3.67	6.50E-10
CLSPO_c21230	*eutQ*	ethanolamine utilization protein EutQ	-3.32	2.50E-08
CLSPO_c21240	*eutN*	ethanolamine utilization protein EutN	-3.28	8.89E-07
CLSPO_c21250	CLSPO_c21250	BMC domain-containing protein	-3.55	2.15E-08
CLSPO_c21260	*eutJ*	ethanolamine utilization protein EutJ	-3.42	9.57E-10
CLSPO_c21270	CLSPO_c21270	hypothetical protein	-3.94	7.16E-11
CLSPO_c21280	CLSPO_c21280	GTP-binding protein, EutP/PduV family	-3.88	1.05E-10
CLSPO_c21290	*pduU*	propanediol utilization protein PduU	-3.07	4.85E-06
CLSPO_c21300	CLSPO_c21300	glycyl-radical enzyme activating protein family protein	-3.62	8.89E-09
CLSPO_c21310	*pflD1*	formate acetyltransferase 2	-4.33	2.67E-14
CLSPO_c21320	*adhE3*	aldehyde-alcohol dehydrogenase AdhE	-4.78	1.11E-15
CLSPO_c21330	CLSPO_c21330	BMC domain-containing protein	-4.26	5.76E-10

Table S12: List of genes from the σ^F regulon (continuation).

Gene ID	Gene name	Gene function	Log_2FC	P-value
CLSPO_c21340	CLSPO_c21340	microcompartment family protein	-5.13	2.30E-15
CLSPO_c21350	CLSPO_c21350	hypothetical protein	-4.95	6.46E-09
CLSPO_c22850	CLSPO_c22850	hypothetical protein	-1.03	7.15E-03
CLSPO_c22950	CLSPO_c22950	amino acid permease family protein	-1.79	2.62E-02
CLSPO_c23700	CLSPO_c23700	putative peptidoglycan binding protein	-4.69	2.12E-04
CLSPO_c25620	*prdH1*	D-proline reducatse cluster associated protein	-1.63	4.55E-03
CLSPO_c25660	CLSPO_c25660	hypothetical protein	-1.29	3.65E-02
CLSPO_c26370	*spoVD2*	stage V sporulation protein D	-2.31	1.90E-09
CLSPO_c30010	CLSPO_c30010	hypothetical protein	-1.42	1.59E-02
CLSPO_c30120	CLSPO_c30120	hypothetical protein DUF2922	-1.56	1.26E-03
CLSPO_c30690	CLSPO_c30690	putative comE operon protein 1	-2.56	7.12E-07
CLSPO_c32520	*oppD*	oligopeptide transport ATP-binding protein OppD	-1.25	2.47E-04
CLSPO_c32960	CLSPO_c32960	hypothetical protein	-3.69	4.01E-07
CLSPO_c33860	*oppA2*	oligopeptide-binding protein OppA	-2.59	8.20E-07

RNA-Seq: List of genes from the σ^E regulon

Table S13: List of genes from the σ^E regulon. All genes have a $log_2FC < -1$ and adjusted p-value ≤ 0.05.

Gene ID	Gene name	Gene function	Log_2FC	P-value
CLSPO_c00390	*sfsA*	sugar fermentation stimulation protein SfsA	-1.17	2.80E-04
CLSPO_c00400	CLSPO_c00400	NUDIX family hydrolase	-1.69	1.75E-04
CLSPO_c00970	*spmA*	spore maturation protein A	-4.02	3.32E-08
CLSPO_c00980	*spmB*	spore maturation protein B	-4.04	9.09E-13
CLSPO_c01050	CLSPO_c01050	3D/G5 domain-containing protein	-2.09	4.36E-12
CLSPO_c01080	CLSPO_c01080	hypothetical protein	-3.90	3.00E-21
CLSPO_c01090	CLSPO_c01090	hypothetical protein	-6.35	3.07E-25
CLSPO_c01270	CLSPO_c01270	AAA domain-containing protein	-1.59	2.70E-08
CLSPO_c01280	*manC*	mannose-1-phosphate guanylyltransferase ManC	-1.39	5.96E-07
CLSPO_c01340	CLSPO_c01340	putative mannosyltransferase	-5.19	2.08E-29
CLSPO_c01360	CLSPO_c01360	hypothetical protein	-5.29	1.13E-02
CLSPO_c01370	CLSPO_c01370	hypothetical protein	-2.35	1.80E-05
CLSPO_c01400	CLSPO_c01400	putative spore coat protein	-8.13	1.13E-76
CLSPO_c01430	CLSPO_c01430	LysM domain-containing protein	-6.46	6.90E-58
CLSPO_c01660	CLSPO_c01660	ZIP transporter family protein	-1.03	1.22E-03
CLSPO_c01900	*spoIID*	stage II sporulation protein D	-5.82	3.18E-26
CLSPO_c01910	CLSPO_c01910	putative peptidase	-1.62	9.37E-03
CLSPO_c01920	*spoIIID*	stage III sporulation protein D	-1.66	3.71E-02
CLSPO_c01940	*yyaC*	putative sporulation protein YyaC	-1.27	1.20E-03
CLSPO_c02080	CLSPO_c02080	peptidase family protein	-4.05	1.50E-21
CLSPO_c02100	*fmnP*	riboflavin transporter FmnP	-3.71	3.59E-94
CLSPO_c02340	CLSPO_c02340	hypothetical protein	-1.87	4.57E-02
CLSPO_c02910	*resE1*	sensor histidine kinase ResE	-1.02	8.94E-03
CLSPO_c03390	CLSPO_c03390	Na+-driven multidrug efflux pump	-1.06	4.46E-02

Table S13: List of genes from the σ^E regulon (continuation).

Gene ID	Gene name	Gene function	Log_2FC	P-value
CLSPO_c03530	CLSPO_c03530	peptide ABC transporter permease	-1.19	4.49E-04
CLSPO_c03640	CLSPO_c03640	GNAT family acetyltransferase	-1.34	5.13E-03
CLSPO_c03900	CLSPO_c03900	MATE efflux family protein	-3.31	7.83E-17
CLSPO_c03930	CLSPO_c03930	hypothetical protein UPF0324	-1.89	1.96E-03
CLSPO_c03990	*yeiL*	DNA-binding transcriptional activator YeiL	-1.27	3.59E-02
CLSPO_c04010	*rnfC*	electron transport complex protein RnfC	-1.11	9.89E-06
CLSPO_c04280	*mntP2*	putative manganese efflux pump MntP 1	-1.63	1.90E-03
CLSPO_c04440	*manA*	mannose-6-phosphate isomerase ManA	-6.48	2.63E-27
CLSPO_c04730	CLSPO_c04730	phospholipase family protein	-1.24	2.17E-10
CLSPO_c04990	CLSPO_c04990	hypothetical protein	-1.54	6.36E-06
CLSPO_c05000	CLSPO_c05000	cell autolysis regulator	-1.19	1.63E-06
CLSPO_c05160	*phoH*	putative phosphate starvation-inducible protein PhoH	-3.74	1.98E-20
CLSPO_c05680	CLSPO_c05680	calcium-transporting ATPase 1	-1.02	2.28E-03
CLSPO_c05900	CLSPO_c05900	SecC motif-containing protein	-1.24	3.31E-02
CLSPO_c06010	CLSPO_c06010	putative heme/steroid binding protein	-5.59	1.32E-26
CLSPO_c06250	CLSPO_c06250	hypothetical protein	-4.11	3.34E-21
CLSPO_c07730	CLSPO_c07730	phage protein	-2.26	1.72E-03
CLSPO_c07740	CLSPO_c07740	sigma-70 family RNA polymerase sigma factor	-3.00	1.62E-11
CLSPO_c07750	CLSPO_c07750	hypothetical protein	-1.42	3.06E-05
CLSPO_c08150	*brnQ*	branched-chain amino acid transport system 2 carrier protein BrnQ	-3.23	3.37E-24
CLSPO_c08930	*ilvE*	branched-chain-amino-acid aminotransferase IlvE	-4.44	3.30E-44
CLSPO_c08950	CLSPO_c08950	Sel1 repeat-containing protein	-3.17	7.89E-08
CLSPO_c10710	*chiD*	chitinase D	-5.06	7.59E-29
CLSPO_c10800	CLSPO_c10800	radical SAM domain-containing protein	-1.63	1.61E-06
CLSPO_c11080	CLSPO_c11080	HD domain-containing protein	-1.23	1.90E-02
CLSPO_c11260	*yycG2*	sensor histidine kinase YycG	-1.07	1.67E-04
CLSPO_c11350	CLSPO_c11350	putative CoA-substrate-specific enzyme activase	-1.60	1.22E-03

Table S13: List of genes from the σ^E regulon (continuation).

Gene ID	Gene name	Gene function	Log_2FC	P-value
CLSPO_c11360	*opuD*	glycine betaine transporter OpuD	-1.08	2.51E-06
CLSPO_c11930	CLSPO_c11930	hypothetical protein UPF0061	-1.68	3.28E-05
CLSPO_c11940	CLSPO_c11940	amine oxidase	-2.39	8.75E-08
CLSPO_c12830	*ytrH*	sporulation protein YtrH	-4.71	9.84E-25
CLSPO_c12840	CLSPO_c12840	hypothetical protein	-4.27	5.08E-08
CLSPO_c12850	CLSPO_c12850	transcriptional regulator, TetR family protein	-1.73	6.09E-05
CLSPO_c13120	CLSPO_c13120	GGDEF domain-containing protein	-1.45	1.29E-06
CLSPO_c13130	CLSPO_c13130	STE like transcription factor domain-containing protein	-1.39	1.61E-11
CLSPO_c13140	CLSPO_c13140	group 2 family glycosyl transferase	-1.55	1.18E-10
CLSPO_c13150	CLSPO_c13150	cellulose synthase domain-containing protein	-1.29	2.32E-10
CLSPO_c13160	CLSPO_c13160	hypothetical protein	-1.42	7.72E-06
CLSPO_c13330	CLSPO_c13330	type IV leader peptidase family protein	-1.39	9.08E-03
CLSPO_c13900	*artM*	arginine transport ATP-binding protein ArtM	-1.61	4.20E-03
CLSPO_c14640	CLSPO_c14640	stage V sporulation protein B	-7.03	1.82E-11
CLSPO_c14810	*grdH*	betaine reductase complex component B subunit beta	-1.64	4.13E-16
CLSPO_c14820	*grdT*	glycine betaine transporter GrdT	-2.01	2.58E-13
CLSPO_c15090	*pyrP*	uracil permease PyrP	-1.45	1.45E-02
CLSPO_c15170	CLSPO_c15170	patatin-like phospholipase family protein	-2.22	6.77E-08
CLSPO_c15360	CLSPO_c15360	aldo/keto reductase family oxidoreductase	-3.50	1.89E-15
CLSPO_c15740	*kamA*	L-lysine 2,3-aminomutase KamA	-1.14	6.80E-03
CLSPO_c15770	CLSPO_c15770	calcium-transporting ATPase	-1.61	3.40E-06
CLSPO_c16680	CLSPO_c16680	hypothetical protein	-3.17	2.75E-11
CLSPO_c16710	CLSPO_c16710	hypothetical protein	-2.54	1.07E-08
CLSPO_c17040	*thrC*	threonine synthase ThrC	-1.82	2.90E-02
CLSPO_c17510	CLSPO_c17510	transcriptional regulator, TetR family	-1.98	1.51E-02
CLSPO_c17620	CLSPO_c17620	hypothetical protein	-6.32	2.30E-07
CLSPO_c17880	*miaA*	tRNA dimethylallyltransferase MiaA	-1.89	1.00E-09

Table S13: List of genes from the σ^E regulon (continuation).

Gene ID	Gene name	Gene function	Log_2FC	P-value
CLSPO_c17890	*mutL*	DNA mismatch repair protein MutL	-1.24	1.54E-06
CLSPO_c17900	*mutS1*	DNA mismatch repair protein MutS	-1.31	1.19E-04
CLSPO_c17950	CLSPO_c17950	acetyltransferase	-1.37	3.68E-04
CLSPO_c17960	CLSPO_c17960	putative spore coat protein	-3.60	8.19E-17
CLSPO_c18000	CLSPO_c18000	chorismate mutase	-1.01	3.85E-02
CLSPO_c18240	*hypE*	hydrogenase expression/formation protein HypE	-3.98	4.24E-17
CLSPO_c18250	*hypD*	hydrogenase isoenzymes formation protein HypD	-4.53	1.03E-17
CLSPO_c18260	*hypF*	carbamoyltransferase HypF	-4.66	8.72E-25
CLSPO_c18270	CLSPO_c18270	hydrogenase expression/synthesis hypA family	-6.02	3.75E-18
CLSPO_c18280	CLSPO_c18280	heme/steroid binding domain-containing protein	-6.04	1.03E-21
CLSPO_c18290	CLSPO_c18290	hydrogenase maturation protease	-7.23	9.46E-09
CLSPO_c18300	CLSPO_c18300	periplasmic [NiFeSe] hydrogenase large subunit	-4.59	7.84E-22
CLSPO_c18310	CLSPO_c18310	periplasmic [NiFeSe] hydrogenase small subunit	-5.49	4.26E-16
CLSPO_c18490	CLSPO_c18490	sporulation protein YtfJ	-2.17	6.03E-06
CLSPO_c18600	*spoIIM*	stage II sporulation protein M	-2.13	3.67E-02
CLSPO_c18630	CLSPO_c18630	UDP-N-acetylglucosamine 2-epimerase	-2.92	2.65E-14
CLSPO_c18810	*spoIIIAH*	stage III sporulation protein AH	-6.04	5.07E-39
CLSPO_c18820	*spoIIIAG*	stage III sporulation protein AG	-4.61	5.09E-17
CLSPO_c18830	*spoIIIAF*	stage III sporulation protein AF	-4.66	6.28E-14
CLSPO_c18840	*spoIIIAE*	stage III sporulation protein AE	-4.45	2.69E-17
CLSPO_c18850	*spoIIIAD*	stage III sporulation protein AD	-5.31	2.52E-26
CLSPO_c18860	*spoIIIAC*	stage III sporulation protein AC	-4.60	3.44E-09
CLSPO_c18870	*spoIIIAB*	stage III sporulation protein AB	-5.03	7.03E-13
CLSPO_c18880	*spoIIIAA*	stage III sporulation protein AA	-5.46	6.85E-27
CLSPO_c18970	CLSPO_c18970	hypothetical protein	-2.58	4.93E-03
CLSPO_c18990	CLSPO_c18990	hypothetical protein	-2.31	3.09E-03
CLSPO_c19000	CLSPO_c19000	putative type IV pilin	-2.85	3.68E-05

Table S13: List of genes from the σ^E regulon (continuation).

Gene ID	Gene name	Gene function	Log_2FC	P-value
CLSPO_c19010	*xcpS*	type II secretion system protein F	-1.88	4.00E-02
CLSPO_c19020	*epsE*	type II secretion system protein E	-2.69	5.52E-10
CLSPO_c19070	CLSPO_c19070	aminopeptidase N	-5.03	1.60E-10
CLSPO_c19800	CLSPO_c19800	spore germination protein	-2.28	2.05E-03
CLSPO_c19810	CLSPO_c19810	Ger(X)C family germination protein	-1.79	4.09E-02
CLSPO_c20090	CLSPO_c20090	transposase	-2.29	3.05E-03
CLSPO_c21360	CLSPO_c21360	multidrug resistance efflux transporter	-5.34	7.95E-19
CLSPO_c22000	CLSPO_c22000	M24 family metallopeptidase	-7.70	4.45E-11
CLSPO_c22010	CLSPO_c22010	M24 family metallopeptidase	-7.63	5.88E-11
CLSPO_c22020	*dtpA*	dipeptide and tripeptide permease A	-6.28	4.75E-58
CLSPO_c22180	CLSPO_c22180	hypothetical protein DUF2935	-8.19	1.53E-25
CLSPO_c22240	*lcdC*	activator of lactoyl-CoA dehydratase LcdC	-4.26	6.53E-30
CLSPO_c22250	CLSPO_c22250	hypothetical protein	-4.04	1.25E-17
CLSPO_c22260	*etfA1*	electron transfer flavoprotein subunit alpha	-5.25	3.08E-46
CLSPO_c22270	*etfB1*	electron transfer flavoprotein subunit beta	-5.37	1.27E-47
CLSPO_c22280	*bcd1*	acyl-CoA dehydrogenase, short-chain specific	-5.60	1.05E-45
CLSPO_c22290	CLSPO_c22290	benzoyl-CoA reductase/2-hydroxyglutaryl-CoA dehydratase subunit, BcrC/BadD/HgdB	-6.01	1.13E-59
CLSPO_c22300	CLSPO_c22300	benzoyl-CoA reductase/2-hydroxyglutaryl-CoA dehydratase subunit, BcrC/BadD/HgdB	-6.64	1.36E-62
CLSPO_c22310	CLSPO_c22310	putative acyl-CoA transferases/carnitine dehydratase	-6.40	2.08E-49
CLSPO_c22320	CLSPO_c22320	ferredoxin	-3.68	2.15E-21
CLSPO_c22490	CLSPO_c22490	protease family protein	-1.68	1.87E-07
CLSPO_c22580	CLSPO_c22580	alpha/beta hydrolase	-2.17	1.96E-02
CLSPO_c22630	*hypA1*	hydrogenase nickel incorporation protein HypA	-9.05	5.36E-09
CLSPO_c22640	*hypB1*	hydrogenase nickel incorporation protein HypB	-5.42	3.80E-23
CLSPO_c23340	CLSPO_c23340	hypothetical protein	-5.83	6.86E-06

Table S13: List of genes from the σ^E regulon (continuation).

Gene ID	Gene name	Gene function	Log_2FC	P-value
CLSPO_c23360	CLSPO_c23360	putative lipoprotein	-2.02	1.13E-06
CLSPO_c23370	*msrB*	peptide methionine sulfoxide reductase MsrB	-1.71	2.37E-03
CLSPO_c23380	*sodF*	putative superoxide dismutase	-3.01	3.15E-08
CLSPO_c23710	*mocA*	molybdenum cofactor cytidylyltransferase MocA	-1.87	1.42E-02
CLSPO_c23750	CLSPO_c23750	xanthine dehydrogenase family protein, molybdopterin-binding subunit	-1.84	1.15E-04
CLSPO_c23760	CLSPO_c23760	hypothetical protein	-1.64	4.81E-02
CLSPO_c23770	CLSPO_c23770	hypothetical protein	-1.51	5.34E-04
CLSPO_c24960	CLSPO_c24960	hypothetical protein	-8.08	6.51E-09
CLSPO_c25090	*dapG*	aspartokinase 1	-3.24	4.12E-14
CLSPO_c25100	CLSPO_c25100	PRC-barrel domain-containing protein	-4.04	5.79E-14
CLSPO_c25370	*comM*	competence protein ComM	-1.41	4.40E-03
CLSPO_c25860	CLSPO_c25860	nucleoside recognition domain-containing protein	-1.89	3.49E-08
CLSPO_c26120	*spoIVA*	stage IV sporulation protein A	-6.38	2.92E-46
CLSPO_c26180	*pstB2*	phosphate ABC transporter ATP-binding protein PstB	-1.13	3.69E-02
CLSPO_c26190	*pstA2*	phosphate ABC transporter, permease protein PstA	-1.41	1.75E-03
CLSPO_c26210	*pstS2*	phosphate ABC transporter substrate-binding protein PstS	-1.72	8.04E-05
CLSPO_c26260	CLSPO_c26260	PRC-barrel domain-containing protein	-6.88	4.62E-22
CLSPO_c26270	*sigG*	RNA polymerase sigma-G factor SigG	-1.84	3.33E-21
CLSPO_c26360	*sigK*	RNA polymerase sigma-28 factor SigK	-2.96	4.22E-15
CLSPO_c26680	CLSPO_c26680	hypothetical protein	-1.14	5.28E-06
CLSPO_c26690	CLSPO_c26690	hypothetical protein	-1.08	1.48E-05
CLSPO_c26720	CLSPO_c26720	putative lipoprotein	-8.78	1.52E-07
CLSPO_c27110	CLSPO_c27110	glucose / sorbosone dehydrogenase	-3.64	9.44E-15
CLSPO_c27120	CLSPO_c27120	putative calcium-translocating P-type ATPase, PMCA-type	-4.95	1.89E-15
CLSPO_c27140	CLSPO_c27140	hypothetical protein	-7.20	1.34E-53
CLSPO_c27530	CLSPO_c27530	lysophospholipase L1	-1.03	3.83E-02

Table S13: List of genes from the σ^E regulon (continuation).

Gene ID	Gene name	Gene function	Log_2FC	P-value
CLSPO_c27560	*tarA*	putative N-acetylmannosaminyltransferase TarA	-1.13	2.07E-02
CLSPO_c27620	CLSPO_c27620	hypothetical protein	-1.89	7.42E-03
CLSPO_c28390	CLSPO_c28390	NHL repeat-containing protein	-1.64	3.36E-06
CLSPO_c28440	CLSPO_c28440	glycerol-3-phosphate responsive antiterminator	-1.34	9.52E-04
CLSPO_c28450	*gyrA2*	DNA gyrase subunit A	-1.12	2.13E-04
CLSPO_c28680	CLSPO_c28680	membrane-spanning protein	-2.66	8.64E-09
CLSPO_c28700	CLSPO_c28700	spore germination protein, GerAC	-1.28	5.11E-05
CLSPO_c28710	CLSPO_c28710	spore germination protein	-1.52	2.08E-07
CLSPO_c28720	*gerKA3*	spore germination protein KA	-1.54	4.46E-08
CLSPO_c29070	CLSPO_c29070	chitinase A-like protein	-1.10	1.99E-04
CLSPO_c29180	*norV1*	anaerobic nitric oxide reductase flavorubredoxin NorV	-2.16	8.11E-09
CLSPO_c29190	*bcd2*	acyl-CoA dehydrogenase, short-chain specific	-2.59	2.05E-09
CLSPO_c29340	*ribH*	6,7-dimethyl-8-ribityllumazine synthase RibH	-3.64	8.49E-63
CLSPO_c29350	*ribBA*	riboflavin biosynthesis protein RibBA	-3.44	6.97E-44
CLSPO_c29380	*pepD3*	cytosol non-specific dipeptidase PepD	-1.05	4.06E-05
CLSPO_c29430	CLSPO_c29430	spore germination protein	-1.24	8.50E-03
CLSPO_c29900	*ribB*	3,4-dihydroxy-2-butanone 4-phosphate synthase RibB	-4.52	4.60E-67
CLSPO_c29920	CLSPO_c29920	cyclic nucleotide-binding protein	-2.33	8.41E-09
CLSPO_c29980	CLSPO_c29980	TPR domain-containing protein	-4.43	4.23E-21
CLSPO_c30080	CLSPO_c30080	hypothetical protein	-1.37	8.38E-03
CLSPO_c30240	CLSPO_c30240	Zn-ribbon protein	-1.38	4.44E-03
CLSPO_c30310	CLSPO_c30310	spore coat protein-like protein	-7.12	8.43E-06
CLSPO_c30420	*yqfC*	sporulation protein YqfC	-4.39	1.11E-04
CLSPO_c30560	CLSPO_c30560	stage II sporulation protein	-5.23	6.17E-36
CLSPO_c30640	*algI*	putative poly(beta-D-mannuronate) O-acetylase AlgI	-1.84	3.02E-02
CLSPO_c30730	*lytR2*	transcriptional regulator LytR	-1.01	5.44E-03
CLSPO_c30750	*nadD*	putative nicotinate-nucleotide adenylyltransferase NadD	-1.07	4.98E-02

Table S13: List of genes from the σ^E regulon (continuation).

Gene ID	Gene name	Gene function	Log_2FC	P-value
CLSPO_c30770	*obg*	GTPase Obg	-1.00	5.81E-03
CLSPO_c30850	CLSPO_c30850	putative metalloprotease	-3.68	9.24E-08
CLSPO_c30860	CLSPO_c30860	Peptidase, family M23/M37	-5.38	2.96E-23
CLSPO_c30980	CLSPO_c30980	hypothetical protein	-2.44	9.33E-13
CLSPO_c31440	CLSPO_c31440	hypothetical protein	-8.07	1.29E-06
CLSPO_c31450	CLSPO_c31450	hypothetical protein	-2.95	5.04E-03
CLSPO_c31600	CLSPO_c31600	hypothetical protein, TIGR04086 family	-2.55	2.70E-04
CLSPO_c31710	CLSPO_c31710	nucleotidyltransferase domain-containing protein	-1.24	2.21E-04
CLSPO_c31760	*yunB*	sporulation protein YunB	-1.56	5.51E-09
CLSPO_c32160	*gadB2*	glutamate decarboxylase GadB	-3.80	1.26E-19
CLSPO_c32230	CLSPO_c32230	hypothetical protein	-7.38	6.50E-25
CLSPO_c32500	CLSPO_c32500	small acid-soluble spore protein beta	-1.38	4.27E-02
CLSPO_c32510	CLSPO_c32510	oligopeptide/dipeptide ABC transporter, ATP-binding protein	-1.24	5.34E-04
CLSPO_c32530	*dppC2*	dipeptide transport system permease protein DppC	-1.14	9.89E-03
CLSPO_c32540	*oppB*	oligopeptide transport system permease protein OppB	-3.27	4.96E-10
CLSPO_c32550	*oppA1*	oligopeptide-binding protein OppA	-4.46	1.59E-70
CLSPO_c32630	CLSPO_c32630	polysaccharide deacetylase family protein	-4.39	4.81E-08
CLSPO_c32670	*folC*	folylpolyglutamate synthase FolC	-1.16	2.17E-02
CLSPO_c32760	CLSPO_c32760	tetratricopeptide repeat protein	-1.46	4.52E-02
CLSPO_c32820	CLSPO_c32820	ErfK/YbiS/YcfS/YnhG family protein	-4.17	2.63E-26
CLSPO_c32970	CLSPO_c32970	putative R-2-hydroxyglutaryl-CoA dehydratase subunit	-1.36	2.91E-02
CLSPO_c32980	*hgdC*	activator of (R)-2-hydroxyglutaryl-CoA dehydratase HgdC	-1.15	1.97E-02
CLSPO_c33000	*etfA2*	electron transfer flavoprotein subunit alpha	-1.93	2.97E-03
CLSPO_c33010	*etfB2*	electron transfer flavoprotein subunit beta	-1.84	4.13E-03
CLSPO_c33020	*bcd3*	acyl-CoA dehydrogenase, short-chain specific	-1.75	6.93E-03
CLSPO_c33030	*thlA*	acetyl-CoA acetyltransferase ThlA	-1.45	3.69E-02
CLSPO_c33050	*crt*	3-hydroxybutyryl-CoA dehydratase Crt	-1.72	1.51E-02

Table S13: List of genes from the σ^E regulon (continuation).

Gene ID	Gene name	Gene function	Log_2FC	P-value
CLSPO_c33130	CLSPO_c33130	hypothetical protein	-3.91	7.87E-03
CLSPO_c33140	CLSPO_c33140	glycerophosphoryl diester phosphodiesterase family protein	-5.93	5.19E-16
CLSPO_c33240	CLSPO_c33240	hypothetical protein	-4.67	8.77E-18
CLSPO_c33290	CLSPO_c33290	protease family protein	-1.07	1.44E-03
CLSPO_c33320	*lon2*	Lon protease	-1.63	1.51E-03
CLSPO_c33380	*pyrE*	orotate phosphoribosyltransferase PyrE	-1.49	1.08E-03
CLSPO_c33390	*pyrD*	dihydroorotate dehydrogenase B (NAD(+)), catalytic subunit	-1.30	1.95E-03
CLSPO_c33400	*pyrK*	dihydroorotate dehydrogenase B (NAD(+)), electron transfer subunit	-1.54	9.34E-03
CLSPO_c33410	*pyrF*	orotidine 5'-phosphate decarboxylase PyrF	-1.60	1.75E-04
CLSPO_c33420	*pyrC*	dihydroorotase PyrC	-1.52	9.08E-04
CLSPO_c33430	*pyrI*	aspartate carbamoyltransferase regulatory chain	-1.84	2.03E-03
CLSPO_c33440	*pyrB*	aspartate carbamoyltransferase PyrB	-1.47	3.50E-02
CLSPO_c33840	CLSPO_c33840	hypothetical protein	-7.16	3.07E-25
CLSPO_c34030	*ldhD*	D-lactate dehydrogenase LdhD	-6.28	7.58E-92
CLSPO_c34300	*cstA2*	carbon starvation protein CstA	-1.26	3.79E-03
CLSPO_c34390	CLSPO_c34390	putative FAD-dependent dehydrogenase	-1.06	4.82E-04
CLSPO_c35170	CLSPO_c35170	hypothetical protein	-2.09	8.66E-04
CLSPO_c35190	CLSPO_c35190	hypothetical protein	-2.40	2.44E-06
CLSPO_c35290	*pfkA*	6-phosphofructokinase PfkA	-1.30	9.43E-05
CLSPO_c35540	*alr3*	alanine racemase 1	-1.03	1.92E-06
CLSPO_c35550	CLSPO_c35550	hypothetical protein	-1.06	1.73E-03
CLSPO_c36000	*ptsI*	phosphoenolpyruvate-protein phosphotransferase PtsI	-1.08	2.58E-02
CLSPO_c36050	CLSPO_c36050	N-acetylmuramoyl-L-alanine amidase	-3.57	5.41E-25
CLSPO_c36970	CLSPO_c36970	hypothetical protein	-1.24	4.20E-02
CLSPO_c37140	*yabQ*	spore cortex biosynthesis protein YabQ	-4.63	5.96E-08
CLSPO_c37150	*yabP*	sporulation protein YabP	-4.79	4.02E-23
CLSPO_c37190	*spoVB2*	stage V sporulation protein B	-2.39	6.04E-05

Table S13: List of genes from the σ^E regulon (continuation).

Gene ID	Gene name	Gene function	Log_2FC	P-value
CLSPO_c37390	CLSPO_c37390	putative phage head-tail adaptor	-1.33	2.39E-02
CLSPO_c37400	CLSPO_c37400	hypothetical protein	-6.99	1.33E-30
CLSPO_c37410	CLSPO_c37410	radical SAM domain-containing protein	-2.70	9.42E-10
CLSPO_c38300	CLSPO_c38300	putative sporulation protein	-1.32	2.97E-02

RNA-Seq: List of genes from the σ^G regulon

Table S14: List of genes from the σ^G regulon. All genes have a $log_2FC < -1$ and adjusted p-value ≤ 0.05.

Gene ID	Gene name	Gene function	Log_2FC	P-value
CLSPO_c00020	*dnaN*	DNA polymerase III subunit beta	-1.07	2.78E-02
CLSPO_c00790	CLSPO_c00790	amidohydrolase-like protein	-3.48	1.53E-10
CLSPO_c01150	CLSPO_c01150	ferredoxin	-2.26	2.68E-03
CLSPO_c01160	CLSPO_c01160	hypothetical protein	-3.70	5.71E-05
CLSPO_c01240	CLSPO_c01240	putative zinc-dependent hydrolase	-2.84	1.20E-04
CLSPO_c01440	*cphB*	cyanophycinase CphB	-2.65	4.44E-02
CLSPO_c01450	*cphA*	cyanophycin synthetase CphA	-2.04	1.89E-02
CLSPO_c01470	CLSPO_c01470	hypothetical protein	-6.29	1.61E-03
CLSPO_c01480	*gerKA4*	spore germination protein KA	-3.05	1.65E-09
CLSPO_c01490	CLSPO_c01490	spore germination protein	-2.23	7.85E-04
CLSPO_c01500	CLSPO_c01500	germination protein GerC family	-2.43	3.44E-05
CLSPO_c01510	*spoIIR*	stage II sporulation protein R	-1.89	2.70E-02
CLSPO_c01530	*ypeB*	germination protein YpeB	-3.74	4.23E-06
CLSPO_c02190	CLSPO_c02190	rhodanese-like sulfurtransferase	-3.85	4.05E-10
CLSPO_c02270	CLSPO_c02270	hypothetical protein DUF4363	-2.76	8.41E-03
CLSPO_c02280	CLSPO_c02280	hypothetical protein	-3.43	5.33E-08

Table S14: List of genes from the σ^G regulon (continuation).

Gene ID	Gene name	Gene function	Log_2FC	P-value
CLSPO_c02370	*sleB*	spore cortex-lytic enzyme	-4.90	5.13E-10
CLSPO_c02970	CLSPO_c02970	germination protein, Ger(X)C family	-2.29	5.89E-05
CLSPO_c02980	CLSPO_c02980	spore germination protein	-2.81	3.27E-11
CLSPO_c02990	*gerKA1*	spore germination protein KA	-2.32	4.24E-04
CLSPO_c03780	CLSPO_c03780	hypothetical protein DUF3189	-1.38	2.72E-03
CLSPO_c04240	CLSPO_c04240	sporulation lipoprotein, YhcN/YlaJ family	-5.74	1.96E-23
CLSPO_c04600	CLSPO_c04600	hypothetical protein	-3.92	4.60E-08
CLSPO_c04790	*artQ1*	arginine transport system permease protein ArtQ	-1.16	1.97E-02
CLSPO_c05100	CLSPO_c05100	hypothetical protein DUF1540	-9.02	1.30E-32
CLSPO_c05270	CLSPO_c05270	ABC-type multidrug transport system, permease component	-1.92	3.28E-02
CLSPO_c05380	CLSPO_c05380	hypothetical protein DUF3892	-2.77	1.37E-05
CLSPO_c05510	CLSPO_c05510	hypothetical protein	-5.79	4.51E-06
CLSPO_c05650	CLSPO_c05650	amino acid permease family protein	-3.09	4.92E-12
CLSPO_c05670	CLSPO_c05670	metallo-beta-lactamase family protein	-2.79	3.12E-06
CLSPO_c05800	CLSPO_c05800	hypothetical protein	-1.95	1.37E-02
CLSPO_c06220	CLSPO_c06220	transcriptional regulator	-1.33	1.70E-02
CLSPO_c06450	*vanS4*	sensor protein VanS	-1.12	2.59E-02
CLSPO_c06570	CLSPO_c06570	ferredoxin	-4.81	9.07E-14
CLSPO_c06920	CLSPO_c06920	hypothetical protein	-2.69	1.81E-02
CLSPO_c08070	*sspH1*	small, acid-soluble spore protein H 1	-6.00	2.82E-18
CLSPO_c08250	CLSPO_c08250	hypothetical protein	-3.02	3.97E-05
CLSPO_c08260	CLSPO_c08260	coat F domain-containing protein	-3.08	1.79E-08
CLSPO_c08360	CLSPO_c08360	hypothetical protein	-6.02	6.41E-06
CLSPO_c08390	CLSPO_c08390	spore germination protein	-2.29	3.81E-03
CLSPO_c08750	CLSPO_c08750	hypothetical protein	-1.12	1.20E-02
CLSPO_c09020	CLSPO_c09020	hypothetical protein	-1.12	6.38E-03
CLSPO_c09770	*dhaD1*	glycerol dehydrogenase DhaD	-1.84	3.36E-02

Table S14: List of genes from the σ^G regulon (continuation).

Gene ID	Gene name	Gene function	Log_2FC	P-value
CLSPO_c10660	CLSPO_c10660	NAD-dependent epimerase/dehydratase	-2.74	5.40E-04
CLSPO_c10720	CLSPO_c10720	thioredoxin domain-containing protein	-2.87	4.03E-11
CLSPO_c11150	CLSPO_c11150	hypothetical protein DUF2971	-1.67	3.34E-02
CLSPO_c11240	CLSPO_c11240	putative heavy metal-binding protein	-3.95	1.52E-09
CLSPO_c11340	*norV3*	anaerobic nitric oxide reductase flavorubredoxin NorV	-1.90	1.48E-03
CLSPO_c11540	CLSPO_c11540	hypothetical protein	-6.23	1.91E-08
CLSPO_c11750	CLSPO_c11750	hypothetical protein	-3.29	8.63E-03
CLSPO_c11900	CLSPO_c11900	hypothetical protein DUF3787	-2.04	6.05E-03
CLSPO_c11910	*nifJ1*	pyruvate-flavodoxin oxidoreductase NifJ	-2.48	1.05E-08
CLSPO_c12530	CLSPO_c12530	hypothetical protein	-5.30	5.13E-10
CLSPO_c12810	CLSPO_c12810	hypothetical protein	-3.76	1.27E-11
CLSPO_c12820	CLSPO_c12820	hypothetical protein DUF4363	-3.58	4.76E-07
CLSPO_c12890	*ytfJ*	sporulation protein YtfJ	-6.23	9.93E-14
CLSPO_c12900	CLSPO_c12900	membrane associated protein	-6.14	2.19E-03
CLSPO_c13320	CLSPO_c13320	hypothetical protein	-4.44	1.38E-04
CLSPO_c13830	CLSPO_c13830	hypothetical protein DUF4396	-2.57	1.46E-02
CLSPO_c13840	CLSPO_c13840	putative cell-surface protein	-2.00	1.49E-02
CLSPO_c13850	CLSPO_c13850	hypothetical protein	-2.25	2.14E-02
CLSPO_c13860	CLSPO_c13860	hypothetical protein DUF1657	-4.38	1.57E-03
CLSPO_c14060	CLSPO_c14060	ABC-type antimicrobial peptide transport system, ATPase component	-2.53	2.08E-02
CLSPO_c14490	CLSPO_c14490	hypothetical protein	-3.04	2.44E-02
CLSPO_c15070	CLSPO_c15070	pseudouridine synthase	-2.04	6.23E-05
CLSPO_c15180	CLSPO_c15180	hypothetical protein	-2.58	1.04E-02
CLSPO_c15330	CLSPO_c15330	FAD dependent oxidoreductase	-4.81	2.43E-25
CLSPO_c15520	CLSPO_c15520	hypothetical protein DUF326	-1.97	3.91E-05
CLSPO_c15650	CLSPO_c15650	hypothetical protein	-2.52	8.22E-06

Table S14: List of genes from the σ^G regulon (continuation).

Gene ID	Gene name	Gene function	Log_2FC	P-value
CLSPO_c15660	CLSPO_c15660	hypothetical protein	-3.37	1.15E-04
CLSPO_c15670	CLSPO_c15670	hypothetical protein	-2.40	4.42E-05
CLSPO_c15680	CLSPO_c15680	hypothetical protein DUF1657	-4.31	1.94E-04
CLSPO_c15690	CLSPO_c15690	putative manganese catalase	-3.88	2.66E-05
CLSPO_c15820	CLSPO_c15820	FtsX-like permease family	-1.55	2.21E-02
CLSPO_c15920	*spoVAF*	stage V sporulation protein AF	-1.91	9.46E-05
CLSPO_c16000	*queC*	7-cyano-7-deazaguanine synthase QueC	-1.78	1.09E-02
CLSPO_c16010	CLSPO_c16010	AraC family transcription regulator	-1.55	3.76E-02
CLSPO_c16520	CLSPO_c16520	hypothetical protein DUF2383	-3.37	3.27E-11
CLSPO_c17260	CLSPO_c17260	hypothetical protein	-1.48	2.99E-02
CLSPO_c17320	CLSPO_c17320	short chain dehydrogenase/reductase family oxidoreductase	-3.91	3.14E-12
CLSPO_c17600	*safA*	spore coat assembly protein SafA	-5.98	5.05E-22
CLSPO_c17640	CLSPO_c17640	cold shock protein	-2.29	1.40E-03
CLSPO_c17800	CLSPO_c17800	small, acid-soluble spore protein alpha	-10.17	1.32E-72
CLSPO_c17810	CLSPO_c17810	small, acid-soluble spore protein beta	-8.00	6.74E-33
CLSPO_c18470	*dacB1*	D-alanyl-D-alanine carboxypeptidase DacB	-2.82	5.85E-03
CLSPO_c18520	*dacF2*	D-alanyl-D-alanine carboxypeptidase DacF	-2.83	2.79E-09
CLSPO_c18660	*spoIVB*	SpoIVB peptidase	-3.37	3.97E-05
CLSPO_c19770	*gerKA2*	spore germination protein KA	-1.73	2.27E-03
CLSPO_c19780	*gerAB*	spore germination protein AB	-2.67	6.54E-04
CLSPO_c19790	*gerKA5*	spore germination protein KA	-2.14	1.52E-02
CLSPO_c19820	CLSPO_c19820	spore germination protein	-2.26	6.40E-05
CLSPO_c20620	CLSPO_c20620	hypothetical protein	-4.25	5.58E-18
CLSPO_c22170	*nlhH*	carboxylesterase NlhH	-2.88	9.54E-05
CLSPO_c22220	CLSPO_c22220	hypothetical protein	-2.36	4.04E-04

Table S14: List of genes from the σ^G regulon (continuation).

Gene ID	Gene name	Gene function	Log_2FC	P-value
CLSPO_c22330	CLSPO_c22330	benzoyl-CoA reductase/2-hydroxyglutaryl-CoA dehydratase subunit, BcrC/BadD/HgdB	-1.77	5.00E-03
CLSPO_c22340	CLSPO_c22340	activator of 2-hydroxyglutaryl-CoA dehydratase	-1.87	1.51E-02
CLSPO_c22400	CLSPO_c22400	linear amide C-N hydrolase, choloylglycine hydrolase family protein	-2.29	1.71E-07
CLSPO_c22520	*hopD*	effector protein HopD2	-1.78	4.08E-03
CLSPO_c22620	*cooS2*	carbon monoxide dehydrogenase CooS	-3.32	1.27E-09
CLSPO_c22780	CLSPO_c22780	short chain dehydrogenase/reductase family oxidoreductase	-3.30	1.18E-04
CLSPO_c22860	CLSPO_c22860	hypothetical protein	-4.02	2.13E-03
CLSPO_c23470	CLSPO_c23470	exonuclease family protein	-1.94	2.00E-02
CLSPO_c23720	CLSPO_c23720	putative membrane protein	-2.59	3.00E-03
CLSPO_c23730	CLSPO_c23730	xanthine dehydrogenase	-2.32	5.35E-03
CLSPO_c23740	CLSPO_c23740	xanthine dehydrogenase family protein, FAD-binding subunit	-2.09	2.36E-02
CLSPO_c25080	*tepA*	translocation-enhancing protein TepA	-1.97	9.45E-03
CLSPO_c25640	*prdC2*	D-proline reductase electron transfer protein PrdC	-1.61	1.32E-02
CLSPO_c26490	CLSPO_c26490	putative permease	-1.37	4.67E-02
CLSPO_c27520	CLSPO_c27520	hypothetical protein	-3.99	3.95E-13
CLSPO_c27570	CLSPO_c27570	heparinase II/III-like protein	-1.48	1.49E-02
CLSPO_c27580	*capK*	protein CapK	-1.14	4.04E-02
CLSPO_c27610	CLSPO_c27610	hypothetical protein	-2.16	3.35E-02
CLSPO_c28320	CLSPO_c28320	hypothetical protein	-3.90	8.85E-13
CLSPO_c28760	CLSPO_c28760	polysaccharide deacetylase family protein	-3.74	1.52E-09
CLSPO_c28900	CLSPO_c28900	putative amidase domain-containing protein	-4.81	5.75E-19
CLSPO_c28910	CLSPO_c28910	D-alanyl-D-alanine carboxypeptidase	-4.42	4.85E-10
CLSPO_c29440	*gerKA6*	spore germination protein KA	-3.51	1.53E-05
CLSPO_c29450	CLSPO_c29450	spore germination protein	-3.46	2.63E-04
CLSPO_c29800	CLSPO_c29800	HlyC/CorC family protein	-3.13	6.40E-05

Table S14: List of genes from the σ^G regulon (continuation).

Gene ID	Gene name	Gene function	Log_2FC	P-value
CLSPO_c29960	CLSPO_c29960	phage infection protein Pip	-1.53	1.02E-02
CLSPO_c30060	CLSPO_c30060	hypothetical protein	-2.27	7.58E-03
CLSPO_c30550	CLSPO_c30550	hypothetical protein	-2.59	4.93E-02
CLSPO_c30570	*gpr*	germination protease Gpr	-1.40	6.78E-03
CLSPO_c30630	CLSPO_c30630	alginate O-acetyltransferase	-2.38	7.47E-04
CLSPO_c31420	CLSPO_c31420	small, acid-soluble spore protein, alpha/beta type	-8.03	2.12E-25
CLSPO_c31430	CLSPO_c31430	hypothetical protein	-2.90	6.60E-04
CLSPO_c31460	CLSPO_c31460	hypothetical protein	-3.03	2.84E-02
CLSPO_c31780	CLSPO_c31780	SpoVA family protein	-1.87	5.29E-04
CLSPO_c31790	*spoVAD*	stage V sporulation protein AD	-2.38	1.19E-04
CLSPO_c31800	*spoVAC*	stage V sporulation protein AC	-3.65	4.96E-08
CLSPO_c31860	CLSPO_c31860	polysaccharide deacetylase family protein	-4.39	2.15E-05
CLSPO_c31870	CLSPO_c31870	nitroreductase	-1.78	3.80E-02
CLSPO_c32390	*ytxC*	putative sporulation protein YtxC	-3.65	8.61E-06
CLSPO_c33090	CLSPO_c33090	HDIG domain-containing protein	-1.76	2.51E-04
CLSPO_c33230	CLSPO_c33230	extracellular solute-binding protein	-3.70	1.20E-07
CLSPO_c33300	*engB*	putative GTP-binding protein EngB	-1.15	2.84E-02
CLSPO_c33850	CLSPO_c33850	hypothetical protein	-2.84	9.70E-03
CLSPO_c34570	CLSPO_c34570	hypothetical protein	-3.01	3.39E-02
CLSPO_c34580	*ndvB*	protein NdvB	-1.94	1.20E-04
CLSPO_c35180	CLSPO_c35180	hypothetical protein	-1.72	2.27E-02
CLSPO_c35680	CLSPO_c35680	coat F domain-containing protein	-4.25	6.78E-06
CLSPO_c35940	CLSPO_c35940	cell wall hydrolase family protein	-3.40	2.83E-03
CLSPO_c36790	CLSPO_c36790	hypothetical protein	-2.75	8.57E-03
CLSPO_c37200	*spoVT*	stage V sporulation protein T	-4.80	2.73E-14
CLSPO_c37320	CLSPO_c37320	hypothetical protein	-3.17	1.20E-02
CLSPO_c37340	*hydA*	periplasmic [Fe] hydrogenase large subunit	-1.47	2.71E-03

Table S14: List of genes from the σ^G regulon (continuation).

Gene ID	Gene name	Gene function	Log_2FC	P-value
CLSPO_c37650	*hydE*	iron-only hydrogenase maturation protein HydE	-1.47	6.75E-03
CLSPO_c37660	*hydG*	iron-only hydrogenase maturation protein HydG	-1.50	1.13E-02
CLSPO_c37670	*hydF*	iron-only hydrogenase maturation protein HydF	-1.66	1.25E-02
CLSPO_c37730	CLSPO_c37730	hypothetical protein	-1.88	1.53E-03
CLSPO_c37910	CLSPO_c37910	putative peptidase	-2.14	3.36E-05
CLSPO_c38310	CLSPO_c38310	hypothetical protein	-3.11	1.03E-02

RNA-Seq: List of genes from the σ^K regulon

Table S15: List of genes from the σ^K regulon. All genes have a $log_2FC < -1$ and adjusted p-value ≤ 0.05.

Gene ID	Gene name	Gene function	Log_2FC	P-value
CLSPO_c00760	CLSPO_c00760	LD-carboxypeptidase family protein	-1.28	1.64E-02
CLSPO_c00800	CLSPO_c00800	inhibitor of sigma-G Gin	-2.08	2.89E-03
CLSPO_c01180	CLSPO_c01180	putative Mg2+ transporter-C, MgtC family protein	-1.00	3.97E-02
CLSPO_c01350	CLSPO_c01350	CotS family spore coat protein	-4.15	6.40E-12
CLSPO_c01380	*cotI*	spore coat protein I	-1.71	1.14E-02
CLSPO_c01390	CLSPO_c01390	glycosyl transferase, group 1 family protein	-6.57	7.02E-87
CLSPO_c01410	CLSPO_c01410	peptidase family protein	-2.68	4.59E-05
CLSPO_c01880	CLSPO_c01880	hypothetical protein DUF1779	-2.30	1.66E-10
CLSPO_c01890	*murA2*	UDP-N-acetylglucosamine 1-carboxyvinyltransferase MurA	-2.62	6.11E-15
CLSPO_c02290	CLSPO_c02290	putative threonyl-tRNA synthetase/uridine kinase	-1.88	4.22E-03
CLSPO_c02300	CLSPO_c02300	hypothetical protein	-5.80	1.99E-34
CLSPO_c02320	CLSPO_c02320	CBS domain-containing protein	-1.63	1.69E-02
CLSPO_c02680	CLSPO_c02680	hypothetical protein	-2.51	6.09E-04
CLSPO_c02730	CLSPO_c02730	mannose-6-phosphate isomerase	-6.30	5.12E-56

Table S15: List of genes from the σ^K regulon (continuation).

Gene ID	Gene name	Gene function	Log_2FC	P-value
CLSPO_c02920	CLSPO_c02920	putative oxidoreductase, FAD-binding protein	-2.74	2.35E-09
CLSPO_c03040	*alr1*	alanine racemase 1	-4.15	3.36E-19
CLSPO_c03790	CLSPO_c03790	hypothetical protein DUF346	-9.43	3.53E-93
CLSPO_c03920	*csd1*	putative cysteine desulfurase Csd	-7.71	1.19E-106
CLSPO_c04650	CLSPO_c04650	hypothetical protein	-9.23	5.84E-91
CLSPO_c05150	CLSPO_c05150	hypothetical protein	-3.20	1.18E-08
CLSPO_c05570	CLSPO_c05570	hypothetical protein	-4.96	4.98E-32
CLSPO_c05580	CLSPO_c05580	hypothetical protein	-1.22	4.55E-03
CLSPO_c05590	*sodB*	superoxide dismutase	-3.47	8.88E-25
CLSPO_c05950	*potA1*	spermidine/putrescine import ATP-binding protein PotA	-1.12	5.82E-03
CLSPO_c06000	CLSPO_c06000	rubrerythrin	-5.49	5.42E-43
CLSPO_c06660	*dppD*	dipeptide transport ATP-binding protein DppD	-2.33	1.84E-04
CLSPO_c06700	CLSPO_c06700	hypothetical protein	-2.40	6.73E-07
CLSPO_c07830	CLSPO_c07830	phosphoglycerate mutase family protein	-3.53	3.14E-07
CLSPO_c07980	CLSPO_c07980	H-type lectin domain protein	-7.45	1.64E-26
CLSPO_c08170	*feoA1*	ferrous iron transport protein A	-6.33	4.20E-40
CLSPO_c08180	*feoB2*	ferrous iron transport protein B	-4.94	1.67E-16
CLSPO_c08190	*feoB3*	ferrous iron transport protein B	-3.64	2.78E-13
CLSPO_c08540	CLSPO_c08540	hypothetical protein	-6.75	1.65E-49
CLSPO_c08550	CLSPO_c08550	hypothetical protein	-7.45	2.14E-65
CLSPO_c08560	CLSPO_c08560	hypothetical protein	-1.69	8.64E-05
CLSPO_c08590	*mraY1*	phospho-N-acetylmuramoyl-pentapeptide-transferase MraY	-2.56	2.21E-07
CLSPO_c08700	*cotA*	spore coat protein A	-4.83	6.30E-27
CLSPO_c08780	CLSPO_c08780	ErfK/YbiS/YcfS/YnhG family protein	-6.78	2.27E-42
CLSPO_c08790	CLSPO_c08790	hypothetical protein	-3.22	9.35E-09
CLSPO_c09700	CLSPO_c09700	ABC-2 family transporter protein	-1.82	2.32E-04

Table S15: List of genes from the σ^K regulon (continuation).

Gene ID	Gene name	Gene function	Log_2FC	P-value
CLSPO_c09960	CLSPO_c09960	rubrerythrin	-1.43	1.53E-02
CLSPO_c10010	CLSPO_c10010	hypothetical protein DUF3298	-4.22	4.01E-13
CLSPO_c10600	CLSPO_c10600	hypothetical protein	-6.74	6.47E-77
CLSPO_c10610	CLSPO_c10610	hypothetical protein	-7.04	1.57E-54
CLSPO_c10620	CLSPO_c10620	hypothetical protein	-3.94	3.68E-31
CLSPO_c10630	CLSPO_c10630	hypothetical protein	-2.34	3.04E-07
CLSPO_c10640	CLSPO_c10640	TPR repeat-containing protein	-4.04	1.18E-08
CLSPO_c10650	CLSPO_c10650	hypothetical protein	-6.95	1.18E-70
CLSPO_c10670	CLSPO_c10670	NAD-dependent epimerase/dehydratase	-2.70	1.90E-03
CLSPO_c10680	CLSPO_c10680	glycosyl transferase family protein	-3.78	5.53E-08
CLSPO_c10690	CLSPO_c10690	carboxypeptidase regulatory-like domain	-4.19	2.67E-24
CLSPO_c10700	CLSPO_c10700	hypothetical protein	-2.59	1.94E-05
CLSPO_c10740	CLSPO_c10740	hypothetical protein	-2.60	1.45E-12
CLSPO_c11270	CLSPO_c11270	hypothetical protein	-1.36	8.93E-03
CLSPO_c12250	CLSPO_c12250	hypothetical protein	-9.23	3.99E-60
CLSPO_c12260	CLSPO_c12260	hypothetical protein DUF3794	-7.04	5.08E-74
CLSPO_c12270	CLSPO_c12270	hypothetical protein	-3.70	1.12E-03
CLSPO_c12280	CLSPO_c12280	putative glycosyl hydrolase	-3.68	2.01E-19
CLSPO_c12330	CLSPO_c12330	hypothetical protein	-1.23	2.26E-02
CLSPO_c12340	CLSPO_c12340	hypothetical protein	-1.36	3.05E-02
CLSPO_c12350	CLSPO_c12350	putative nucleotidase	-1.44	1.55E-03
CLSPO_c12360	CLSPO_c12360	dipeptidase family protein	-2.22	7.36E-13
CLSPO_c12440	*cotJA*	spore coat associated protein CotJA	-7.28	2.13E-111
CLSPO_c12450	*cotJB*	spore coat peptide assembly protein CotJB	-7.00	2.39E-112
CLSPO_c12460	*cotJC*	spore coat protein CotJC	-6.96	1.28E-100
CLSPO_c12520	*sspH2*	small, acid-soluble spore protein H 2	-1.53	1.27E-02
CLSPO_c12800	CLSPO_c12800	cation efflux family protein	-1.22	1.39E-02

Table S15: List of genes from the σ^K regulon (continuation).

Gene ID	Gene name	Gene function	Log_2FC	P-value
CLSPO_c13490	CLSPO_c13490	rubrerythrin	-2.39	6.56E-08
CLSPO_c13500	CLSPO_c13500	hypothetical protein	-3.44	2.57E-14
CLSPO_c13510	CLSPO_c13510	putative transcriptional regulator	-1.94	4.48E-10
CLSPO_c13920	CLSPO_c13920	methyl-accepting chemotaxis protein	-2.29	2.06E-04
CLSPO_c14040	CLSPO_c14040	GGDEF domain-containing protein	-1.25	3.66E-03
CLSPO_c14170	CLSPO_c14170	gamma-D-glutamyl-L-diamino acid endopeptidase 1	-3.94	1.33E-22
CLSPO_c14400	CLSPO_c14400	hypothetical protein DUF3298	-3.18	1.26E-11
CLSPO_c14410	CLSPO_c14410	hypothetical protein DUF2804	-3.15	1.39E-14
CLSPO_c15530	CLSPO_c15530	putative methyl-accepting chemotaxis protein	-1.19	2.04E-02
CLSPO_c16030	CLSPO_c16030	caax amino protease family protein	-1.27	4.63E-02
CLSPO_c16610	*ytaF*	putative sporulation protein YtaF	-7.51	1.10E-39
CLSPO_c16630	CLSPO_c16630	hypothetical protein	-4.15	1.29E-07
CLSPO_c16640	CLSPO_c16640	hypothetical protein	-6.52	3.12E-53
CLSPO_c16730	CLSPO_c16730	sigma-70 family RNA polymerase sigma factor	-1.82	9.43E-03
CLSPO_c16740	CLSPO_c16740	YvrJ protein family	-1.23	7.30E-03
CLSPO_c17560	CLSPO_c17560	hypothetical protein	-1.24	1.23E-02
CLSPO_c17570	CLSPO_c17570	putative amidase domain-containing protein	-4.01	7.55E-27
CLSPO_c18640	CLSPO_c18640	YmaF family protein	-4.41	1.25E-21
CLSPO_c19500	CLSPO_c19500	alpha/beta superfamily hydrolase	-3.46	1.75E-13
CLSPO_c19830	CLSPO_c19830	hypothetical protein	-2.93	1.75E-12
CLSPO_c19840	CLSPO_c19840	hypothetical protein	-2.67	8.54E-04
CLSPO_c19850	CLSPO_c19850	hypothetical protein	-3.93	1.30E-27
CLSPO_c19890	CLSPO_c19890	TPR repeat-containing protein	-3.12	2.15E-08
CLSPO_c21170	CLSPO_c21170	hypothetical protein	-1.97	9.69E-09
CLSPO_c21410	CLSPO_c21410	hypothetical protein	-7.34	4.21E-55
CLSPO_c22160	CLSPO_c22160	hypothetical protein	-7.48	8.13E-51
CLSPO_c22390	CLSPO_c22390	putative oxidoreductase	-1.00	3.22E-02

Table S15: List of genes from the σ^K regulon (continuation).

Gene ID	Gene name	Gene function	Log_2FC	P-value
CLSPO_c22840	CLSPO_c22840	N-acetylmuramoyl-L-alanine amidase	-2.63	1.85E-08
CLSPO_c23150	CLSPO_c23150	hypothetical protein	-4.88	1.63E-29
CLSPO_c23260	CLSPO_c23260	hypothetical protein DUF3862	-5.06	1.13E-34
CLSPO_c23290	CLSPO_c23290	hypothetical protein	-6.22	4.72E-40
CLSPO_c23320	CLSPO_c23320	flavodoxin oxidoreductase	-8.05	5.99E-86
CLSPO_c23330	CLSPO_c23330	Spo0E like sporulation regulatory protein	-3.89	1.89E-06
CLSPO_c23350	CLSPO_c23350	phage holin protein	-3.60	1.49E-18
CLSPO_c23910	CLSPO_c23910	Spo0E like sporulation regulatory protein	-1.19	3.41E-02
CLSPO_c24170	CLSPO_c24170	tRNA uridine 5-carboxymethylaminomethyl modification enzyme GidA	-7.16	2.98E-59
CLSPO_c24700	CLSPO_c24700	hypothetical protein	-2.92	6.31E-08
CLSPO_c24900	CLSPO_c24900	hypothetical protein	-5.87	8.02E-21
CLSPO_c24940	CLSPO_c24940	tRNA uridine 5-carboxymethylaminomethyl modification enzyme GidA	-4.69	4.72E-37
CLSPO_c25280	*cdsA*	phosphatidate cytidylyltransferase CdsA	-1.28	3.62E-02
CLSPO_c25790	CLSPO_c25790	putative permease	-2.70	3.83E-07
CLSPO_c25800	CLSPO_c25800	hypothetical protein	-3.02	2.05E-07
CLSPO_c25940	CLSPO_c25940	hypothetical protein	-4.37	2.36E-21
CLSPO_c26130	CLSPO_c26130	collagen triple helix repeat protein	-1.53	1.80E-02
CLSPO_c26620	CLSPO_c26620	hypothetical protein DUF4397	-8.16	3.59E-98
CLSPO_c28470	CLSPO_c28470	hypothetical protein	-2.99	4.38E-13
CLSPO_c28560	*mepA3*	multidrug export protein MepA	-1.57	2.16E-02
CLSPO_c29930	CLSPO_c29930	cell wall hydrolase	-1.53	9.68E-04
CLSPO_c29940	CLSPO_c29940	hypothetical protein	-1.61	5.27E-04
CLSPO_c32100	*pgcA1*	phosphoglucomutase PgcA	-2.55	3.72E-06
CLSPO_c32220	CLSPO_c32220	hypothetical protein	-6.58	5.82E-67
CLSPO_c32240	CLSPO_c32240	hypothetical protein	-5.21	3.72E-43

Table S15: List of genes from the σ^K regulon (continuation).

Gene ID	Gene name	Gene function	Log_2FC	P-value
CLSPO_c32250	CLSPO_c32250	beta 1,4 glucosyltransferase	-3.10	8.89E-14
CLSPO_c32260	CLSPO_c32260	collagen triple helix repeat-containing protein	-4.47	4.81E-25
CLSPO_c32270	CLSPO_c32270	collagen triple helix repeat-containing protein	-3.82	1.80E-17
CLSPO_c32280	CLSPO_c32280	exosporium protein	-7.12	1.76E-54
CLSPO_c32650	CLSPO_c32650	hypothetical protein	-5.96	5.17E-43
CLSPO_c32720	CLSPO_c32720	putative ABC transporter, substrate-binding protein	-6.53	2.61E-21
CLSPO_c32740	CLSPO_c32740	ABC transporter, ATP-binding protein	-5.96	3.83E-29
CLSPO_c32750	CLSPO_c32750	ABC transporter, permease protein	-4.37	4.72E-43
CLSPO_c32950	CLSPO_c32950	iron-sulfur cluster-binding protein	-1.48	1.19E-02
CLSPO_c33100	CLSPO_c33100	hypothetical protein	-1.53	7.79E-03
CLSPO_c33250	CLSPO_c33250	putative transcriptional regulator	-1.78	1.35E-02
CLSPO_c33260	CLSPO_c33260	hypothetical protein	-6.79	4.43E-37
CLSPO_c33270	CLSPO_c33270	hypothetical protein	-9.21	7.64E-77
CLSPO_c33280	CLSPO_c33280	hypothetical protein	-9.65	2.58E-94
CLSPO_c33980	CLSPO_c33980	hypothetical protein	-1.73	1.22E-02
CLSPO_c34630	CLSPO_c34630	hypothetical protein	-2.26	3.27E-04
CLSPO_c34790	CLSPO_c34790	insect kinin peptide	-7.04	6.24E-49
CLSPO_c35260	CLSPO_c35260	hypothetical protein DUF1540	-1.22	3.68E-02
CLSPO_c35710	CLSPO_c35710	superoxide dismutase	-7.03	3.11E-44
CLSPO_c37360	CLSPO_c37360	hypothetical protein	-3.15	2.89E-13
CLSPO_c37750	*aprX*	serine protease AprX	-2.30	3.63E-11